10236

Volume 60 • Number 6 • December 2010

Ocean Dynamics

Theoretical Computational and Observational Oceanography

Special Issue
Multi-scale modelling of coastal, shelf and global ocean dynamics

Editors:
Eric Deleersnijder
Vincent Legat
Pierre F. J. Lermusiaux

Springer

Ocean Dynamics

Volume 60 • Number 6 • December 2010

EDITORIAL

ORIGINAL PAPERS

Instructions for authors
Please go to the journal's website: www.springer.com/10236

Forthcoming articles in print
Online and citable with the DOI (Digital Object Identifier) can be found at springerlink.com

Indexed in Science Citation Index

Ocean Dynamics (2010) 60:1357–1359
DOI 10.1007/s10236-010-0363-6

EDITORIAL

Multi-scale modelling of coastal, shelf and global ocean dynamics

Eric Deleersnijder · Vincent Legat · Pierre F. J. Lermusiaux

Received: 12 November 2010 / Accepted: 12 November 2010 / Published online: 27 November 2010

1 Introduction

Methods for widening the range of resolved scales (i.e. performing multi-scale simulations) in ocean sciences and engineering are developing rapidly, now allowing multi-scale ocean dynamics studies. Having recourse to grid nesting has been and still is a popular method for increasing marine models' resolution when and where needed and for easily allowing the use of different dynamics at different resolution. However, this is not the only way to achieve this goal. Various techniques for modifying locally the grid resolution or dealing with complex-geometry domains are available. For instance, composite, structured grids and unstructured meshes offer an almost infinite geometrical flexibility.

This special issue focuses on multi-scale modelling of coastal, shelf and global ocean dynamics, including the development of new methodologies and schemes and their applications to ocean process studies. Several articles focus on numerical aspects of unstructured mesh space discretisation. Danilov (2010) shows that the noise developing on triangular meshes on which the location of the variables is inspired by Arakawa's C-grid is the largest for regimes close to geostrophic balance. The noise can be reduced by specific operators but cannot be entirely suppressed, "making the triangular C-grid a suboptimal choice for large-scale ocean modelling". Then, the companion articles of Blaise et al. (2010) and Comblen et al. (2010) describe the space and time discretisation of a three-dimensional, baroclinic, finite element model based on the discontinuous Galerkin (DG) technique. This is a significant step forward in the field of finite element ocean modelling, though this model cannot yet be regarded as suitable for tackling realistic applications. Ueckermann and Lermusiaux (2010) also consider DG finite element techniques, focusing on biological–physical dynamics in regions with complex bathymetric features. They compare low- to high-order discretisations, both in time and space, for regimes in which biology dominates, advection dominates or terms are balanced. They find that higher-order schemes on relatively coarse grids generally perform better than low-order schemes on fine grids. Kleptsova et al. (2010) assess various advection schemes for z-coordinate, three-dimensional models in which flooding and drying is taken into account. In this study, the ability to conserve momentum is regarded as the main criterion for selecting a suitable method. On the other hand, Maβmann (2010)

Responsible Editor: Jörg-Olaf Wolff

E. Deleersnijder · V. Legat
Institute of Mechanics, Materials and Civil Engineering (iMMC),
Université catholique de Louvain (UCL),
4 Avenue G. Lemaître,
1348 Louvain-la-Neuve, Belgium

E. Deleersnijder (✉)
Earth and Life Institute (ELI), G. Lemaître Centre for Earth and Climate Research (TECLIM),
Université catholique de Louvain (UCL),
2 Chemin du Cyclotron,
1348 Louvain-la-Neuve, Belgium
e-mail: eric.deleersnijder@uclouvain.be

V. Legat
G. Lemaître Centre for Earth and Climate Research (TECLIM),
Université catholique de Louvain,
2 Chemin du Cyclotron,
1348 Louvain-la-Neuve, Belgium

P. F. J. Lermusiaux
Department of Mechanical Engineering,
Massachusetts Institute of Technology,
Room 5-207B, 77 Massachusetts Avenue,
Cambridge, MA 02139-4307, USA

assesses automatic differentiation for obtaining the adjoint of an unstructured mesh, tidal model of the European continental shelf.

Two articles deal with grid nesting. Nash and Hartnett (2010) introduce a flooding and drying method that can be used in structured, nested grid systems. This can be regarded as an alternative to flooding and drying techniques that are being developed for unstructured mesh models (e.g. Kärnä et al. 2010). Then, Haley and Lermusiaux (2010) derive conservative time-dependent structured finite volume discretisations and implicit two-way embedded schemes for primitive equations with the intent to resolve tidal-to-mesoscale processes over large multi-resolution telescoping domains with complex geometries including shallow seas with strong tides, steep shelf breaks and deep ocean interactions. The authors present realistic simulations with data assimilation in three regions with diverse dynamics and show that their developments enhance the predictive capability, leading to better match with ocean data.

Various multi-scale, realistic simulations are presented. Using a finite element ice model and a slab ocean as in Lietaer et al. (2008), Terwisscha van Scheltinga et al. (2010) model the Canadian Arctic Archipelago, focusing on the pathways for freshwater and sea-ice transport from the Arctic Ocean to the Labrador Sea and the Atlantic Ocean. The unstructured mesh can represent the complex geometry and narrow straits at high resolution and allows improving transports of water masses and sea ice. Walters et al. (2010) have recourse to an unstructured mesh model to study tides and current in Greater Cook Strait (New Zealand). They identify the mechanisms causing residual currents. By means of the unstructured mesh Finite Volume Coastal Ocean Model (FVCOM), Wang et al. (2010) study the hydrodynamics of the Bohai Sea. Xu et al. (2010) simulate coastal and urban inundation due to storm surges along US East and Gulf Coasts. A sensitivity analysis reveals the importance of precise topographic data and the need for a bottom drag coefficient accounting for the presence of mangroves. Finally, Yang and Khangaonkar (2010) resort to FVCOM to simulate the three-dimensional circulation of Puget Sound, a large complex estuary system in the Pacific Northwest coastal ocean, including variable forcing from tides, the atmosphere and river inflows. Comparisons of model estimates with measurements for tidal elevation, velocity, temperature and salinity are deemed to be promising, from larger-scale circulation features to near-shore tide flats.

This special issue suggests that numerical techniques for multi-scale space discretisation are progressively becoming mature. One direction for future progress lies in the improvement of time discretisation methods for the new generation models, so that they can successfully compete with finite difference, structured mesh models based on (almost) constant resolution grids that have been developed and used over the past 40 years (e.g. Griffies et al. 2009).

Acknowledgements The guest editors are indebted to the authors and anonymous reviewers for their contributions to this special issue. Eric Deleersnijder is a research associate with the Belgian Fund for Scientific Research (F.R.S.-FNRS). The contributions of Eric Deleersnijder and Vincent Legat to this special issue were achieved in the framework of the project "Taking up the challenges of multi-scale marine modelling", which is funded by the Communauté Française de Belgique under contract ARC 10/15-028 and the Interuniversity Attraction Pole Timothy (www.climate.be/timothy). The latter is funded by BELSPO (www.belspo.be) under contract IAP6.13. Pierre Lermusiaux is very grateful to the Office of Naval Research for support under grants N00014-07-1-1061 and N00014-08-1-1097 to the Massachusetts Institute of Technology. Some of the articles of this special issue derive from presentations made at the *8th International Workshop on Unstructured Mesh Numerical Modelling of Coastal, Shelf and Ocean Flows* (16–18 September 2009, Louvain-la-Neuve, Belgium, www.uclouvain.be/umm2009), for which the F.R.S.-FNRS and the contract ARC 04/09-316 (Communauté Française de Belgique) provided financial support.

References

Blaise S, Comblen R, Legat V, Remacle J-F, Deleersnijder E, Lambrechts J (2010) A discontinuous finite element baroclinic marine model on unstructured prismatic meshes. Part I: space discretization. Ocean Dynamics (this issue)

Comblen R, Blaise S, Legat V, Remacle J-F, Deleersnijder E, Lambrechts J (2010) A discontinuous finite element baroclinic marine model on unstructured prismatic meshes. Part II: implicit/explicit time discretization. Ocean Dynamics (this issue)

Danilov S (2010) On utility of triangular C-grid type discretization for numerical modeling of large-scale ocean. Ocean Dynamics (this issue)

Griffies SM, Adcroft AJ, Banks H, Böning CW, Chassignet EP, Danabasoglu G, Danilov S, Deleersnijder E, Drange H, England MH, Fox-Kemper B, Gerdes R, Gnanadesikan A, Greatbatch RJ, Hallberg RW, Hanert E, Harrison MJ, Legg SA, Little CM, Madec G, Marsland S, Nikurashin M, Pirani A, Simmons HL, Schröter J, Samuels BL, Treguier AM, Toggweiler JR, Tsujino H, Vallis GK, White L (2009) Problems and prospects in large-scale ocean circulation models. In: Hall J, Harrison DE, Stammer D (eds) Proceedings of the OceanObs'09 conference: sustained ocean observations and information for society. ESA Publication WPP-306. ESA, Venice, pp 21–25

Haley PJ Jr, Lermusiaux PFJ (2010) Multiscale two-way embedding schemes for free-surface primitive-equations in the "Multidisciplinary simulation, estimation and assimilation system". Ocean Dynamics (this issue)

Kärnä T, de Brye B, Gourgue O, Lambrechts J, Comblen R, Legat V, Deleersnijder E (2010) A fully implicit wetting–drying method for DG-FEM shallow water models, with an application to the Scheldt Estuary. Computer Methods in Applied Mechanics and Engineering. doi:10.1016/j.cma.2010.07.001

Kleptsova O, Stelling GS, Pietrzak JD (2010) An accurate momentum advection scheme for a *z*-level coordinate model. Ocean Dynamics (this issue)

Lietaer O, Fichefet T, Legat V (2008) The effects of resolving the Canadian Arctic Archipelago in a finite element sea ice model. Ocean Model 24:140–152

Maβmann S (2010) Sensitivities of an adjoint, unstructured mesh, tidal model on the European Continental Shelf. Ocean Dynamics (this issue)

Nash S, Hartnett M (2010) Nested circulation modelling of inter-tidal zones: details of a nesting approach incorporating moving boundaries. Ocean Dynamics (this issue)

Terwisscha van Scheltinga AD, Myers PG, Pietrzak JD (2010) A finite element sea ice model of the Canadian Arctic Archipelago. Ocean Dynamics (this issue)

Ueckermann MP, Lermusiaux PFJ (2010) High order schemes for 2D unsteady biogeochemical ocean models. Ocean Dynamics (this issue)

Walters RA, Gillibrand PA, Bell R.G, Lane EM (2010) A study of tides and currents in Cook Strait, New Zealand. Ocean Dynamics (this issue)

Wang J, Shen Y, Guo Y (2010) Seasonal circulation and influence factors of the Bohai Sea: a numerical study based on Lagrangian particle tracking method. Ocean Dynamics (this issue)

Xu H, Zhang K, Shen J, Li Y (2010) Storm surge simulation along the U.S. East and Gulf Coasts using a multi-scale numerical model approach. Ocean Dynamics (this issue)

Yang Z, Khangaonkar T (2010) Multi-scale modeling of Puget Sound using an unstructured-grid coastal ocean model: from tide flats to estuaries and coastal waters. Ocean Dynamics (this issue)

Ocean Dynamics (2010) 60:1361–1369
DOI 10.1007/s10236-010-0339-6

On utility of triangular C-grid type discretization for numerical modeling of large-scale ocean flows

Sergey Danilov

Received: 15 April 2010 / Accepted: 7 September 2010 / Published online: 24 September 2010

Abstract Ocean circulation models based on triangular C-grid discretization are frequently employed to simulate coastal ocean dynamics on unstructured meshes. It is shown that on time and space scales dominated by slow geostrophic dynamics, this discretization tends to exhibit checkerboard noise in the field of horizontal velocity divergence and vertical velocity, respectively. The noise is linked to the geometry of triangular C-grid and is amplified in regimes that are close to geostrophic balance through the particular structure of the Coriolis operator. It can be partly suppressed in some cases but remains a problem in a general case and makes the triangular C-grid a suboptimal choice for large-scale ocean modeling.

Keywords Unstructured mesh ocean modeling · Triangular C-grid discretization

1 Introduction

A C-grid type of triangular discretization, frequently referred to as a mixed finite volume–finite difference approach, offers a numerically efficient choice for modeling oceanic flows on unstructured meshes. It is very close to the finite-element discretization employing the Raviart–Thomas low-order element RT0 and emerges as a particular option of mass matrix lumping in the latter case; for comparison and discussion, see Walters et al. (2009) and references therein. There are several models exploiting C-grid type of discretization (see, for example, Casulli and Walters 2000; Fringer et al. 2006) which were conceived for simulating coastal dynamics and are successfully exploited with that purpose. Judged by its numerical efficiency, the triangular C-grid (or lumped RT0) implementations outperform most of other methods suitable for unstructured meshes. This suggests, as first proposed by Stuhne and Peltier (2006), to apply it for simulating the large-scale ocean circulation, a task where numerical efficiency is of paramount importance because of vast ocean volumes and long temporal scales. The question, however, is whether such a discretization provides a good framework in this case. Compared to coastal applications, the large-scale flows more closely follow the geostrophy while long integration times impose their own constraints on the accuracy of mass, heat, and salt balances.

The goal of this note is to discuss a particular problem emerging for the C-grid/RT0 triangular discretization in geostrophically dominated regimes. It easily occurs in (but is not limited to) wind-driven flows and is manifested as a checkerboard pattern in the field of horizontal velocity divergence (or vertical velocity). Its origin is rooted in the geometry of placing variables which affects the properties of discrete divergence, gradient, and Coriolis operators while its strength is sensitive to many other factors, including the representation of subgrid-scale dissipation operators and mesh non-uniformity.

Le Roux et al. (2007) point to a numerical noise of RT0 finite-element pair manifested through a checkerboard pattern in the elevation field in the case of two-dimensional linearized shallow water flow

Responsible Editor: Pierre Lermusiaux

S. Danilov (✉)
Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany
e-mail: Sergey.Danilov@awi.de

driven by steady sources and sinks and approaching the geostrophic equilibrium. In their case, the mode emerges when the Rossby radius of deformation is not resolved. It is attributed to the particular behavior of dispersion curves and the presence of a spurious numerical wave mode. In fact there is coupling between the physical and spurious mode that deforms the physical dispersion. While the barotropic Rossby radius is large and commonly well resolved in real-world applications, this finding has far going implications for the baroclinic modes because even the first baroclinic Rossby radius is seldom resolved in large-scale simulations.

Gassmann (2010) suggest an insightful analysis of the geometrical roots of the divergence noise on triangular C-grid and proves that it is linked to the inability of the gradient operator to maintain a consistent representation of a vector which is in turn linked to the large size of scalar space (element circumcenters). The problem is alleviated on hexagonal C-grids which thus present a much better departing point for numerical ocean modeling.

Below it is shown that the divergence mode is almost always present in solutions obtained on unstructured triangular C-grids and contaminates not only transient but also stationary solutions. Its manifestations are affected by the structure of the discretized Coriolis operator which has a large null space on uniform triangular meshes (see Hanert 2004; Rostand et al. 2008). In the regimes approaching the geostrophic balance inverting this operator (in the presence of viscosity) for the velocity may increase the noise projecting on the divergence mode. Although the rank deficiency of Coriolis operator disappears on general unstructured meshes, the geostrophic velocity still preserves projection on the divergence noise mode.

We will be dealing with the C-grid case. The code used in numerical examples[1] in Section 4 combines two approaches, by Stuhne and Peltier (2006) and by Fringer et al. (2006). It follows Stuhne and Peltier (2006) in keeping spherical geometry but uses pressure (elevation) correction method as in Fringer et al. (2006). Similarly to Fringer et al. (2006), the second-order Adams–Bashforth method is used to estimate the right-hand sides of momentum and tracer equations. The elevation is fully implicit, as is common in large-scale ocean applications. Clearly, the approaches listed here all derive from the earlier one by Casulli and Walters (2000).

The plan of this note is as follows: First the origin of numerical mode will be explained using Fourier expansion (Section 2) and elementary setup of wind-driven flow in a homogeneous fluid (Section 3). Its manifestations are further illustrated in Section 4 using a Munk gyre setup and both homogeneous and stratified fluids. The last section concludes.

[1]The code can be supplied on request.

2 Wave modes on a uniform triangular mesh

This section follows the analysis in Le Roux et al. (2007) and is inspired by the analysis of Gassmann (2010). It serves to explain the origin of the mode for linearized flat-bottom shallow water equations in an elementary case of infinite mesh composed of equilateral triangles in planar geometry. In this case, one may consider a single Fourier component $e^{i(-\omega t+kx+ly)}$, where k and l are wavenumbers and ω the frequency. Let $(-d/2, 0)$, $(d/2, 0)$, and $(0, h)$ be vertices of a basic triangle pointing northward and $(-d/2, 0)$, $(d/2, 0)$, and $(0, -h)$ of that pointing southward ($h = \sqrt{3}d/2$). All other triangles are obtained by translations of these two. The normals associated with edges are outer ones for the n-triangles. There are three sets of normal velocities, $\overline{u_\alpha}$, $\overline{u_\beta}$, and $\overline{u_\gamma}$, defined at centers of similarly oriented edges, and two sets of elevations, $\overline{\eta_n}$ and $\overline{\eta_s}$, defined at circumcenters of respective triangles (for an n-triangle α, β, γ label its bottom, left, and right sides, respectively). One seeks a solution of discretized linear shallow water equations

$$\mathrm{I}\partial_t u_n + \mathrm{C}u_n + g\mathrm{G}\eta = 0, \quad \mathrm{I}\partial_t \eta + H\mathrm{D}u_n = 0.$$

by substituting

$$\left(\overline{u_\alpha}, \overline{u_\beta}, \overline{u_\gamma}, \overline{\eta_n}, \overline{\eta_s}\right) = \left(u_\alpha, u_\beta, u_\gamma, \eta_n, \eta_s\right) e^{i(-\omega t+kx+ly)},$$

where $u_\alpha, u_\beta, u_\gamma, \eta_n$, and η_s denote the amplitudes. In the equations above, $u_n = (\overline{u_\alpha}, \overline{u_\beta}, \overline{u_\gamma})^T$, $\eta = (\overline{\eta}_n, \overline{\eta}_s)^T$, I, C, G, and D are the identity, Coriolis, gradient, and divergence operators, respectively; H the uniform fluid depth; and g the acceleration due to gravity. This leads to a system of five linear equations for the amplitudes. Equating its determinant to zero gives the dispersion relation. It has a single root $\omega = 0$ corresponding to the geostrophic mode leaving the fourth-order equation for a coupled set of two physical and two spurious modes.

The system matrix can readily be composed of discrete operators given further. Its eigenvalues and eigenvectors can be analyzed numerically (it yields to analytical analysis too).

The discrete divergence for the two basic triangles can be written as

$$D_n = (2/h)\left(u_\alpha e^{-ilh/3} + u_\beta e^{-ikd/4+ilh/6} + u_\gamma e^{ikd/4+ilh/6}\right),$$

and

$$D_s = -(2/h)\left(u_\alpha e^{ilh/3} + u_\beta e^{ikd/4 - ilh/6} + u_\gamma e^{-ikd/4 - ilh/6}\right),$$

the matrix of Coriolis operator (acts on $(u_\alpha, u_\beta, u_\gamma)^T$) is written as

$$\mathrm{C} = \frac{f}{\sqrt{3}} \begin{pmatrix} 0 & -a_1 & a_2 \\ a_1 & 0 & -a_3 \\ -a_2 & a_3 & 0 \end{pmatrix},$$

where f is the (constant) Coriolis parameter, $a_1 = \cos(-kd/4 + lh/2)$, $a_2 = \cos(kd/4 + lh/2)$, and $a_3 = \cos(kd/2)$. Finally, the discrete elevation gradient is

$$G_\alpha = (3/2h)\left(-\eta_n e^{ilh/3} + \eta_s e^{-ilh/3}\right),$$

$$G_\beta = (3/2h)\left(-\eta_n e^{ikd/4 - ilh/6} + \eta_s e^{-ikd/4 + ilh/6}\right),$$

and

$$G_\gamma = (3/2h)\left(-\eta_n e^{-ikd/4 - ilh/6} + \eta_s e^{ikd/4 + ilh/6}\right).$$

Up to a numerical factor, the operators of divergence D and gradient G are transpose of each other.

Now, the argument of Gassmann (2010) is that the representation of velocity with three components is ambiguous because a physical vector satisfies the consistency constraint $u_\alpha + u_\beta + u_\gamma = 0$, which is not necessarily maintained by discretization. Indeed, the components of elevation gradient given above will fail to maintain it if $\eta_n \neq \eta_s$ in the limit of small wavenumbers, and dynamics may allow such states. Similarly, by expanding exponents in expressions for D_n and D_s, one sees that any error in the consistency projects into the checkerboard pattern in divergence. This is a general argument; its manifestations depend on dynamics.

As applied to the linear shallow water equations, it can be seen, by computing the eigenvectors, that of five modes the two spurious modes do correspond to checkerboard patterns in both elevation ($\eta_n \approx -\eta_s$) and divergence ($D_n \approx -D_s$). The physical wave modes have $\eta_n \approx \eta_s$ for small wave numbers. For $|kd|, |lh| \sim 1$ or larger, even physical modes show some difference in η_n and η_s, which becomes more pronounced if the Rossby radius $\lambda = \sqrt{gH}/f$ is not resolved (the "physical mode" can be described by different branches of dispersion equation depending on parameters). This behavior is in full analogy with the RT0 case (the system matrix is given in Le Roux et al. (2007)).

In the context of linear shallow water equations considered above, the presence of divergence noise is set by initial conditions, and if they are sufficiently smooth and the Rossby radius is resolved, the spurious mode is not necessarily excited with a significant amplitude. This explains why the triangular C-grid is reported to be a good performer in applications dominated by barotropic tidal dynamics.

The geostrophic mode does not show any checkerboard pattern in the two-dimensional case because it corresponds to zero frequency and hence zero divergence. The constraint that the horizontal divergence be zero is no longer imposed in three dimensions and we shall see that strong divergence noise evolves in quasi-geostrophic states. This is the main reason explaining why the noise affects three-dimensional flows.

Noteworthy, the null space of Coriolis operator is spanned by the vector $(a_3, a_2, a_1)^T$ which is approximately $(1, 1, 1)^T$ for small wavenumbers. This implies that any deviation of elevation (pressure) gradient from the constraint will be amplified by inverting the Coriolis operator (friction or viscosity are needed to regularize it) in order to obtain the geostrophically balanced velocity.

We do not carry this analysis further because it cannot be generalized to unstructured meshes. Instead we will consider simple flow configurations this time dealing with the geostrophic mode in three dimensions.

3 Zonally reentrant flow driven by wind

Consider a zonal flow driven by wind in a channel of depth H which is zonally reentrant in the x-direction and bounded by rigid walls at $y = L$ at the north and $y = 0$ at the south. Let the density be uniform and equal to some reference value ρ_0. Assume the flow be driven by the force $\tau(y)/h$ per unit mass applied in the x-direction in the fluid layer of thickness h (τ is due to wind, it is spread over the upper model layer in order to avoid using the vertical viscosity). For reference, we present first the solution for the continuous case. The steady linear equations of motion take the form

$$\mathbf{f} \times \mathbf{u} + g\nabla\eta = \theta(z + h)\tau(y)/h + \mathrm{V}_h\mathbf{u}, \tag{1}$$

$$\partial_z w + \nabla \times \mathbf{u} = 0. \tag{2}$$

Here $\mathbf{u} = (u, v)$ is the horizontal velocity, η the elevation, $\nabla = (\partial_x, \partial_y)$ the horizontal gradient operator, $\mathbf{f} = f\mathbf{k}$ the Coriolis parameter with $\mathbf{k}$ the unit vertical vector, V_h the horizontal viscosity (dissipation) operator, and θ is the Heaviside theta function. Integrating the continuity equation in the vertical direction from the bottom $z = -H$ to surface $z = 0$, one obtains in the linear approximation

$$\nabla \times \mathbf{U} = 0, \quad \mathbf{U} = (U, V) = \int_{-H}^{0} \mathbf{u} dz. \tag{3}$$

To further simplify the solution assume that $\mathrm{V}_h = -r$, where r^{-1} sets the time scale of dissipation. Integrating Eq. 1 over the depth and making use of Eq. 3, one finds

$$U = \tau/r$$

and then from the meridional component of vertically integrated momentum equation

$$gH\partial_y\eta = -fU - rV = -f\tau/r.$$

Notice that the divergence of vertically integrated velocity is identically zero. Notice also that the zonal momentum balance is fully ageostrophic and that one needs friction or viscosity to ensure the existence of stationary solution.

Consider any level below $z = -h$. The momentum equations there reduce to

$$-vf = -ru, \quad uf + g\partial_y\eta = -rv$$

yielding

$$u = \tau(rH)^{-1}\left(1 + r^2/f^2\right)^{-1},$$

and

$$v = \tau(fH)^{-1}\left(1 + r^2/f^2\right)^{-1}.$$

For the midlatitude ocean, $r \ll f$, so that the meridional velocity is much smaller than the zonal one and is virtually independent of friction. On the contrary, the amplitude of zonal, geostrophic component is controlled by the friction. Importantly, τ should be zero at $y = 0, L$ to ensure impermeability in the case $\mathrm{V}_h = -r$ (infinitely thin boundary layers will be required otherwise). The limitation is lifted if true viscous operator is used. In contrast to the vertically integrated transport, the divergence of horizontal velocity, $\partial_y v$, is not zero.

Let's rewrite the equations in the form relevant to the C-grid to learn how this flow configuration is represented with this discretization. Let u_n denote the vector of normal velocities at mid-edges. The notation for the vectors of elevation and vertical velocity at circumcenters will be shared with the continuous case. The stationary equations can then be written as

$$\mathrm{C}u_n + g\mathrm{G}\eta = (\tau/h)\theta(z+h) + \mathrm{V}_h u_n, \quad \mathrm{D}u_n + \mathrm{Z}w = 0. \tag{4}$$

Here V_h and Z stand for discrete operators of viscosity and vertical derivative, respectively. The Coriolis operator on C-grids is nonlocal, and one has to invert global operators to solve Eq. 4. Integrating Eq. 4 vertically, one gets $(\mathrm{C} - \mathrm{V}_h)U_n + g\mathrm{HG}\eta = \tau$ and $\mathrm{D}U_n = 0$, where U_n denotes the vertically integrated u_n and H is the diagonal matrix of depth at edges in the general case of variable bottom topography. Expressing

$$U_n = (\mathrm{C} - \mathrm{V}_h)^{-1}(\tau - g\mathrm{HG}\eta)$$

and requiring its divergence be zero results in the equation on η

$$g\mathrm{D}(\mathrm{C} - \mathrm{V}_h)^{-1}\mathrm{HG}\eta = \mathrm{D}(\mathrm{C} - \mathrm{V}_h)^{-1}\tau.$$

Inverting the operator $\mathrm{D}(\mathrm{C} - \mathrm{V}_h)^{-1}\mathrm{HG}$ gives η. Below the surface layer $u_n = -g(\mathrm{C} - \mathrm{V}_h)^{-1}\mathrm{G}\eta$. Its horizontal divergence is then

$$\mathrm{D}u_n = -g\mathrm{D}(\mathrm{C} - \mathrm{V}_h)^{-1}\mathrm{G}\eta = -\mathrm{D}(\mathrm{C} - \mathrm{V}_h)^{-1}\tau/H, \tag{5}$$

where the last equality is valid only for the flat bottom case.

We construct the operators (matrices) mentioned above and compute velocity divergence for a mesh shown schematically in Fig. 1. The mesh is cyclic in the zonal direction. Position of boundary nodes (including those at the cyclic boundary) are fixed, but the positions of internal nodes may be selected arbitrarily except for the requirement that circumcenters stay inside the

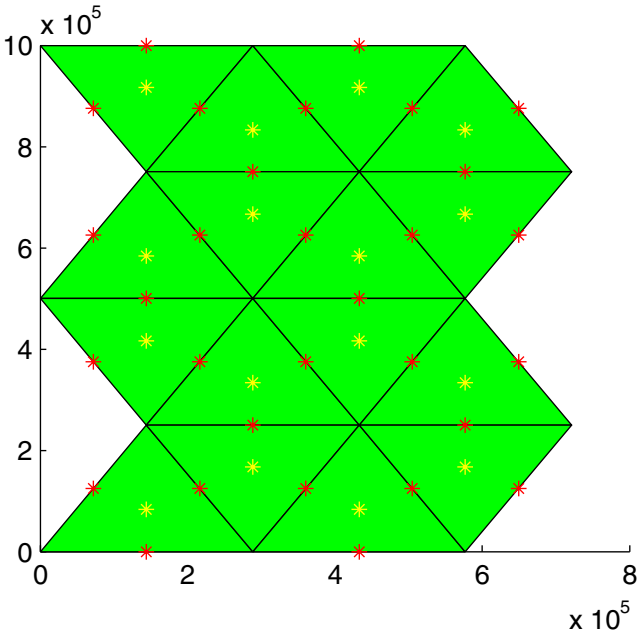

Fig. 1 Schematics of the mesh fragment used for matrix inversion. The case $N = 4$ is shown, and $N = 19$ is used in computations. The elevation is at circumcenters (*yellow stars*), and normal velocities are at mid-edges (*red stars*). The mesh is cyclic (eastern edges are the cyclic image of the western ones). To obtain distorted meshes, the internal nodes are displaced. The distance is in meters

triangles. The number of triangles in the meridional direction can be arbitrary; increasing it will increase the resolution. If $4N$ is the number of triangles (the case $N = 4$ is shown in Fig. 1), the dimension of u_n in any layer is $6N - 2$ (the northern and southern edges do not contribute because of impermeability, and eastern and western edges coincide because of cyclicity). The number of layers in vertical does not matter as soon as it exceeds one because all operators necessary for computations are two-dimensional. Assembling operators G and D is straightforward. Assembling C is implemented through constructing operators returning elemental velocities, turning them and returning back to edges as in Stuhne and Peltier (2006). This is a convenient way of implementing the reconstruction by Perot (2000). We will use notation of Stuhne and Peltier (2006). Let indices c and e enumerate triangles (cells) and edges, respectively. One introduces the function $\delta_{e,c}$ which is zero if triangle c does not contain edge e, and ± 1 if it does, with the positive sign if the normal at e points to c. The elemental vectors are obtained from their edge components as

$$A_c \mathbf{u}_c = \sum_e \delta_{e,c} u_e l_e (\mathbf{x}_c - \mathbf{x}_e),$$

and their normal components are obtained from vectors as

$$d_e u_e = \sum_c \delta_{e,c} \mathbf{u}_c (\mathbf{x}_c - \mathbf{x}_e).$$

Here the following notation is used: l_e is the length of the edge e, d_e is the distance between the centers of triangles sharing e, $\mathbf{x}_c$ and $\mathbf{x}_e$ are vectors drawn to the circumcenter and mid-edge respectively, and A_c is the area of triangle.

Construction of viscous operators of the form used in Stuhne and Peltier (2006) and Fringer et al. (2006) follows the same ideology as with the Coriolis operator but requires additionally assembling augmented gradient and divergence operators of dimensions $(6N + 2, 4N)$ and $(4N, 6N + 2)$, respectively, to take into account the boundary conditions of no slip. We will only mention a particular problem of this operator; the illustrations in this section are made with a substitute "viscous" operator $-r\mathrm{I}$. The case $N = 19$ is considered further.

Figure 2 displays the divergence of horizontal velocity below the surface layer in solutions obtained for $\tau = \tau_0 \sin(\pi y/L)$ ($\tau_0 = 0.0001$ m^2/s^2, $L = 1{,}000$ km,

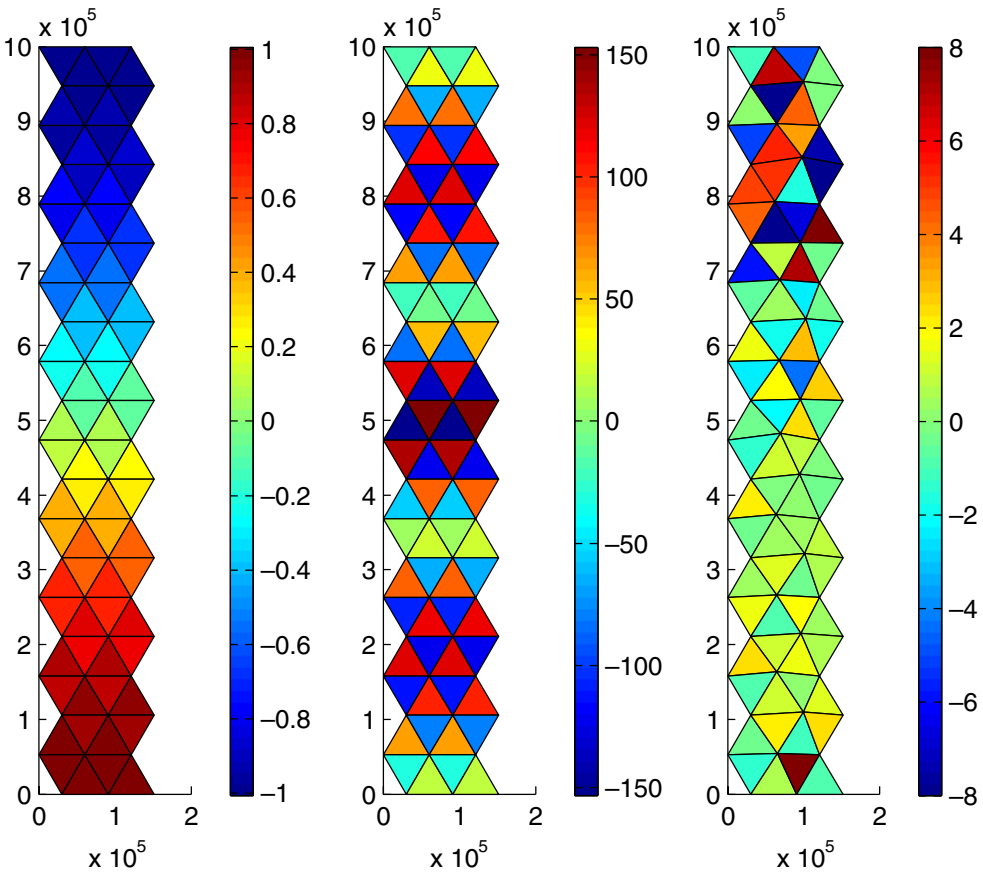

Fig. 2 Divergence of horizontal velocity field excited by a sine wind stress below the surface Ekman layer for the case of linear horizontal friction. The bottom is flat and mesh is uniform (*left*), the depth is varying by 10% on uniform mesh (*middle*), and the bottom is flat, but mesh is distorted by 20% (*right*). The divergence is normalized with $\pi \tau_0/(LfH)$; the distance is in meters

and $r^{-1} = 10$ days), i.e., in the absence of boundary layers at the northern and southern walls. The left panel corresponds to the flat bottom ($H = 1{,}000$ m) and uniform mesh. The bottom topography varies around the mean depth as a sine function with 10% amplitude in the case shown in the middle panel. Finally, the right panel corresponds to the flat bottom but distorted mesh (its inner nodes are randomly displaced within 20% of uniform grid spacing). The divergence is normalized by the analytical solution amplitude ($\pi\tau_0/(LfH)$; the mean depth is used when topography varies). Solutions for the elevation are smooth and close to the analytical result in the flat bottom cases; they are not shown. The matrix $\mathrm{D}(\mathrm{C} - \mathrm{V}_h)^{-1}\mathrm{HG}$ is singular because there is freedom up to a constant in η; it is fixed by letting the elevation be zero at a particular point. The velocity field is then computed as $u_n = -g(\mathrm{C} - \mathrm{V}_h)^{-1}\mathrm{G}\eta$, and its divergence is obtained by further applying D.

Consider the case of flat bottom and uniform mesh first. If the wind stress is discretized by projecting the full wind stress vector on α, β, and γ directions (the consistency constraint is satisfied), the noise in the velocity divergence $\mathrm{D}u_n$ can only be linked to the inversion of $\mathrm{C} - \mathrm{V}_h$ according to the second equality in Eq. 5. The operator C is singular on uniform meshes, but τ selected here is sufficiently smooth and has a negligible projection on its null space. As a consequence, small regularization provided by friction ($r \ll f$) is sufficient to recover the analytical solution, as shown in the left panel of Fig. 2.

If the bottom topography is varying, only the first equality in Eq. 5 is valid. Our topography is gentle and in this case $\mathrm{G}\eta$ gives a vector that satisfies the consistency constraint up to a very small error. To prove this, we computed $\mathrm{DG}\eta$ (not shown), which turns to be very close to $-\pi^2\eta/L^2$ (except for small wall effects) and does not exhibit noticeable noise pattern. However, the solution for velocity divergence $\mathrm{D}(\mathrm{C} - \mathrm{V}_h)^{-1}\mathrm{G}\eta$ in the middle panel of Fig. 2 shows only the noise. It seems plausible to argue that the small deviation from consistency in $\mathrm{G}\eta$ is amplified by the action of $(\mathrm{C} - \mathrm{V}_h)^{-1}$ leading to a much stronger effect.

When the mesh is distorted and bottom is flat, as in the case shown in the right panel of Fig. 2, the second of Eq. 5 is once again applicable. Although C^{-1} exists in this case, the action of $(\mathrm{C} - \mathrm{V}_h)^{-1}$ fails to maintain the consistency of an originally consistent vector leading to the divergence noise. It is worth of mentioning that, compared to the previous case, $\mathrm{G}\eta$ in this case shows much stronger deviations and $\mathrm{DG}\eta$ contains well-pronounced noise. The noise in velocity divergence is, however, smaller, presumably because of the existence of C^{-1}.

We briefly describe what happens for a more general forcing. If the wind stress supports nonzero ageostrophic transport at the channel walls, the boundary layers are forming, and true viscosity operator is required to maintain them.

Unexpectedly, replacing friction with the Laplacian viscosity creates a problem: The velocity divergence becomes extremely noisy even for the sine wind stress profile. We found that the elevation in this case has a correct shape, but a wrong amplitude (a factor 1.5 smaller). The velocity divergence shows a checkerboard pattern and, for $A_h = 10^3$ m^{2}s^{-1}, is approximately two orders of magnitude larger than that of the true solution.

Apparently the regularization of Coriolis operator with viscosity works unsatisfactorily. The difficulty arises because the conservative discretization of viscosity operator used in Stuhne and Peltier (2006) and Fringer et al. (2006) does not provide the approximation to the true operator at the locations of velocity. It is a straightforward exercise to show that on uniform meshes for $\mathbf{u} = (u(y), 0)$, one gets $(3/2)\partial_{yy}u(y)$ while for $\mathbf{u} = (0, v(y))$ the result is $(1/2)\partial_{yy}v(y)$ on non-zonal edges of equilateral triangles. The factor 3/2 explains the reduced amplitude of the elevation. The dependence on the flow direction leads to an error that has a strong projection on the velocity divergence.

For equilateral or isosceles triangles, the correct performance of viscosity operator can be recovered if velocity vectors at circumcenters are reconstructed in a more accurate way. The procedure is as follows: First compute the tangent component of velocity at edges to get full velocity vectors there. This can be done by averaging full vectors of standard reconstruction to the edges. The edge reconstruction of full velocities is of the second order while the original, elemental reconstruction is only the first order even on equilateral triangles. Next, interpolate full edge velocities to circumcenters and proceed as Stuhne and Peltier (2006) and Fringer et al. (2006) to estimate the Laplacian. The generalization to nonuniform meshes remains an open question. Inversions performed with the modified Laplacian on uniform meshes look similar to the cases already considered but for checkerboarding in boundary layers if the latter are not resolved.

The examples presented above were selected to illustrate that specific mechanisms leading to the divergence noise may vary. In particular, the emergence of divergence noise in this configuration is independent of whether the Rossby radius is resolved (it influences only the amplitude of elevation), and the noise amplitude is affected by characteristics of discrete Coriolis and viscosity operators.

4 Munk gyre flows

When the beta effect, nonlinearity, baroclinicity, thermal forcing, and boundary effects are included, it is highly improbable that the divergence noise will become less significant. The Munk gyre flow generated by the wind stress

$$\tau_x = -\tau_0 \cos(\pi y/L)$$

in a closed basin offers a more complex example. The theory of Munk gyre flow can be found in many textbooks (see, for example, Pedlosky 1996) and is not repeated here.

The Munk gyre is simulated in a hexagonal domain (chosen to simplify the generation of orthogonal mesh). The mesh contains 21 unevenly spaced z-levels. The meridional size of the domain is slightly smaller than 9° and the largest size in the zonal direction is about 18°. Other parameters are as follows: The horizontal and vertical viscosities are 1,000 and 0.01 m^2/s, $\tau_0\rho_0$ =0.2 N/m^2 (ρ_0 is the reference density), time step Δt =20 min, the resolution is 1/3°, and fluid depth is 4,000 m. The geometry is spherical, so that mesh triangles are not equilateral and the Coriolis parameter varies with latitude. The viscous operator is implemented as in Stuhne and Peltier (2006) and Fringer et al. (2006). The horizontal viscosity coefficient is intentionally large to ensure that the western Munk boundary layer is well resolved.

The upper panel of Fig.3 shows the vertical velocity w at the depth of 100 m after only 2 months of integration in a setup with fluid of uniform density. The momentum advection is switched off to remove any influence of w on the circulation. Although the gyre circulation is itself smooth (and not shown for this reason), the divergence of the horizontal velocity field is patchy as visualized by the pattern of w. Its amplitude is two orders of magnitude higher than the true vertical velocity, similarly as in the simple flow of previous section.

The attempt to use the modified Laplace operator of previous section (the triangles are approximately isosceles here) shows only a limited success in this case. Although the noise does not fill the entire domain over years of integration, it nevertheless propagates too far from the walls, and the solution for w remains unsatisfactory.

Admittedly, the lack of stratification exaggerates the problem because the vertical velocity in a stratified fluid should be constrained so that motion follows isopycnals. However, the checkerboard pattern in w is well apparent even then. To illustrate it, the full

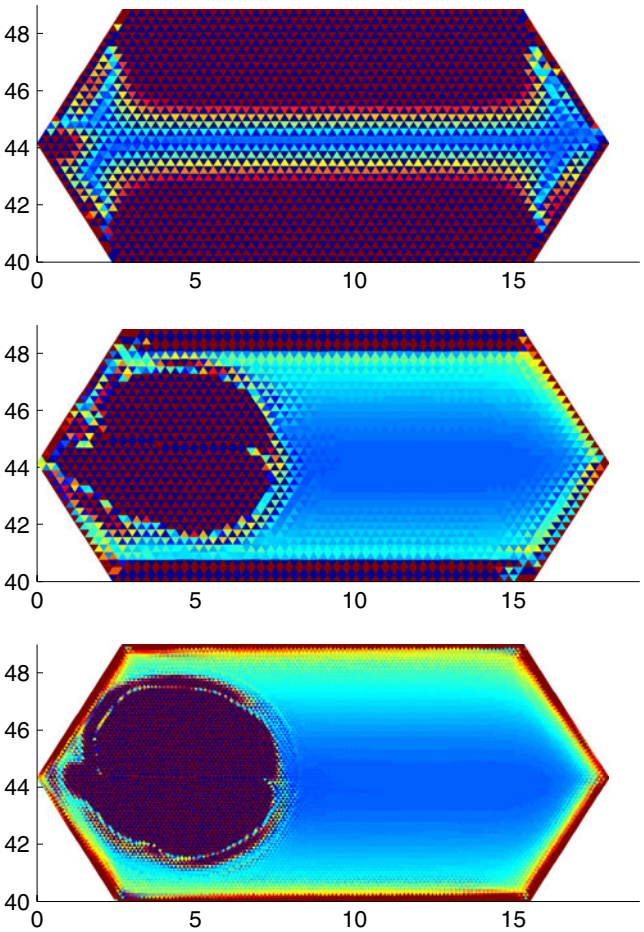

Fig. 3 Vertical velocity field in the Munk gyre flow at 100 m depth after 2 months of simulations. The *colorbar* is between $\pm 10^{-5}$ m/s, the amplitudes over the noisy part are from one to three orders of magnitude larger. The *axes* correspond to longitude and latitude. Shown are the case of homogeneous fluid without momentum advection (*top*), stratified fluid obeying full dynamics on standard (1/3°, *middle*) and refined (1/8°, *bottom*) meshes. Stratification damps the mode, but not in boundary layers and over convection site

system of primitive equations is now integrated. The stratification is due to temperature which drops linearly from 20° at the surface to 12.5° at the bottom (4,000 m). The salinity is kept fixed at 35 psu. Additionally to the wind forcing, the relaxation to prescribed temperature is applied at the surface. This temperature field contains a localized cold spot to the east of the western boundary current (the spot is Gaussian in shape with anomaly of −1° and is centered at approximately 4.5° E and 44.4° N; convection would penetrate to the depth of about 500 m if the fluid was quiescent).

The middle plot in Fig. 3 shows, when compared to the upper one, that although the amplitude of the mode is essentially reduced by the stratification, it remains

large over the convection site and in the boundary layers. The vertical mixing parameterization is switched on in the model so that vertical mixing coefficient over the convection site may contain mesh-scale fluctuations. One can hypothesize that they trigger the noise in w there. The first baroclinic Rossby radius is about 25 km which is comparable to the mesh resolution.

The simulation of the stratified case was repeated on a refined mesh with the resolution of 1/8°, keeping the same horizontal viscosity of 1,000 m^2s^{-1} which is in this case more than an order of magnitude higher than the standard usage. In this case, the representation of boundary layers is improving (bottom panel in Fig. 3), but the mode is still preserved within the convection site and in its vicinity. The first internal Rossby radius is well resolved on this mesh; thus, resolution alone is insufficient to suppress the mode.

The noise in vertical velocity is not immediately destroying the integration, and Fig. 4 serves to illustrate that one still obtains smooth horizontal velocity and temperature fields (top panel) despite the noise in the field of vertical velocity (bottom panel) after 2 years of integration. However, compared to the earlier time (Fig. 3, middle panel), the noise in w increases and finally leads to instability.

In order to learn how the non-uniformity of mesh influences the noise, we repeated the computations of Fig. 4 on a distorted mesh obtained by randomly displacing the inner nodes within 15% of grid spacing. The result is very similar to that shown in the bottom panel of Fig. 4 apart from a slightly larger area where strong checkerboard pattern is visible.

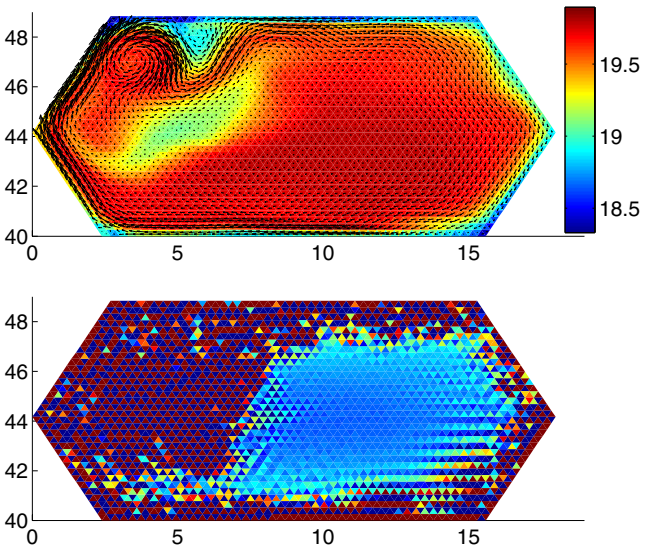

Fig. 4 Temperature, horizontal velocity (*top*), and vertical velocity (*bottom*) at 100 m depth in the Munk gyre flow after 2 years of simulations. The noise in w is increasing compared to the middle panel of Fig. 4. The *colorbar* for temperature is in degrees Celsius, and for vertical velocity, between $\pm 10^{-5}$ m/s. The *coordinate axes* are the longitude and latitude

Some reduction in the amplitude of spurious divergence can be obtained by a brute force approach—through adding second viscosity $\nabla(A\nabla\cdot\mathbf{u})$ to the rhs of the momentum equation. The magnitude of second viscosity coefficient A should be taken at minimum which still does not allow the numerical mode to develop. It turns, however, out that relatively high values are required for a noticeable effect. Values as high as $A = 500$ m^2/s suppress the noise considerably on 1/8° mesh, but they are by far too high to recommend this method.

5 Conclusions

Among discretization methods suited for unstructured meshes, the C-grid type of discretization ranks high in terms of numerical efficiency. However, as follows from the consideration above, it is not necessarily an optimal choice to model the large-scale ocean circulation because of its tendency to generate noise in the divergence of horizontal velocity. The examples considered above are related to quasi-stationary flows that are close to the geostrophic balance. However, the noise may appear in transient regimes as well if internal Rossby radii are not resolved. The stabilization in the form of second viscosity partly removes this noise, and on uniform meshes one may try divergence averaging (see Gassmann 2010), but they both fail to ensure a fully satisfactory solution in a general case.

The origin of noise has geometrical roots and is linked to inability of triangular C-grid to maintain consistent vector representation Gassmann (2010). The examples considered above illustrate how manifestation of divergence noise are affected by the properties of Coriolis and viscosity operators or mesh inhomogeneity.

The presence of stratification suppresses the mode in the deep ocean, but it is still present in the boundary layers and over locations where the fluid is diabatic (like convection sites or mixed layers). The existence of noise in regions with baroclinic dynamics was already mentioned by Stuhne and Peltier (2009).

When dynamics are effectively barotropic, the spurious modes are not necessarily excited provided proper initialization. This is seemingly the case in coastal applications, which explains the obvious success of approaches by Casulli and Walters (2000) or Fringer et al. (2006) there. For large-scale flows which approach

the geostrophic balance, the divergence noise will be present in a general case, as demonstrated by examples and analysis here. This prevents triangular C-grids from being recommended for modeling large-scale ocean circulation.

Acknowledgements I am indebted to E. Hanert and A. Gassmann for their critical remarks on the earlier version of this work which led to substantial changes. The discussion with D. Le Roux was very helpful too.

References

Casulli V, Walters RA (2000) An unstructured grid, three-dimensional model based on the shallow water equations. Int J Numer Methods Fluids 32:331–348

Fringer OB, Gerritsen M, Street RL (2006) An unstructured-grid, finite-volume, nonhydrostatic, parallel coastal ocean simulator. Ocean Model 14:139–173

Gassmann A (2010) C-grid discretization of the SWE in planar trivariate coordinate system. J Comput Phys (in press)

Hanert E (2004) Toward a finite element ocean circulation model. Ph.D. thesis, Université Catolique de Louvain, Belgium. Available online at http://sites.uclouvain.be/slim/assets/files/hanert_phd.pdf

Le Roux DY, Rostand V, Pouliot B (2007) Analysis of numerically induced oscillations in 2D finite-element shallow water models. Part I: inertia-gravity waves. SIAM J Sci Comput 29:331–360

Pedlosky J (1996) Ocean circulation theory. Springer, Berlin

Perot B (2000) Conservation properties of unstructured staggered mesh schemes. J Comput Phys 159:58–89

Rostand V, Le Roux DY, Carey G (2008) Kernel analysis of the discretized finite difference and finite element shallow-water models. SIAM J Sci Comput 31:531–556

Stuhne GR, Peltier WR (2006) A robust unstructured grid discretization for 3-dimensional hydrostatic flows in spherical geometry: a new numerical structure for ocean general circulation modeling. J Comput Phys 213:704–729

Stuhne GR, Peltier WR (2009) An unstructured C-grid based method for 3D global ocean dynamics: free-surface formulations and tidal test cases. Ocean Model 28:97–105

Walters RA, Hanert E, Pietrzak J, Le Roux DY (2009) Comparison of unstructured, staggered grid methods for the shallow water equations. Ocean Model 28:106–117

Ocean Dynamics (2010) 60:1371–1393
DOI 10.1007/s10236-010-0358-3

A discontinuous finite element baroclinic marine model on unstructured prismatic meshes

Part I: space discretization

Sébastien Blaise · Richard Comblen · Vincent Legat · Jean-François Remacle · Eric Deleersnijder · Jonathan Lambrechts

Received: 6 July 2010 / Accepted: 27 October 2010 / Published online: 26 November 2010

Abstract We describe the space discretization of a three-dimensional baroclinic finite element model, based upon a discontinuous Galerkin method, while the companion paper (Comblen et al. 2010a) describes the discretization in time. We solve the hydrostatic Boussinesq equations governing marine flows on a mesh made up of triangles extruded from the surface toward the seabed to obtain prismatic three-dimensional elements. Diffusion is implemented using the symmetric interior penalty method. The tracer equation is consistent with the continuity equation. A Lax–Friedrichs flux is used to take into account internal wave propagation. By way of illustration, a flow exhibiting internal waves in the lee of an isolated seamount on the sphere is simulated. This enables us to show the advantages of using an unstructured mesh, where the resolution is higher in areas where the flow varies rapidly in space, the mesh being coarser far from the region of interest. The solution exhibits the expected wave structure. Linear and quadratic shape functions are used, and the extension to higher-order discretization is straightforward.

Responsible Editor: Pierre Lermusiaux

S. Blaise · R. Comblen · V. Legat (✉) · J.-F. Remacle · E. Deleersnijder · J. Lambrechts
Institute of Mechanics, Materials and Civil Engineering (IMMC), Université catholique de Louvain, Euler, 4 Avenue G. Lemaître, 1348 Louvain-la-Neuve, Belgium
e-mail: vincent.legat@uclouvain.be

S. Blaise · R. Comblen · V. Legat · J.-F. Remacle · E. Deleersnijder · J. Lambrechts
Georges Lemaître Centre for Earth and Climate Research (TECLIM), Université catholique de Louvain, Euler, 4 Avenue G. Lemaître, 1348 Louvain-la-Neuve, Belgium

S. Blaise · R. Comblen · V. Legat · J.-F. Remacle · E. Deleersnijder · J. Lambrechts
Centre for Systems Engineering and Applied Mechanics (CESAME), Université catholique de Louvain, Euler, 4 Avenue G. Lemaître, 1348 Louvain-la-Neuve, Belgium

E. Deleersnijder
Earth and Life Institute (ELI), Université catholique de Louvain, Euler, 4 Avenue G. Lemaître, 1348 Louvain-la-Neuve, Belgium

S. Blaise
Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, 1850 Table Mesa Drive, Boulder, CO 80305, USA

1 Introduction

Ocean models have reached a high level of complexity, and eddy resolving simulations are now much more affordable than in the past. However, mainstream models still fit into the same framework as the pioneering model published by Bryan (1969). This approach, which uses conservative finite differences on structured grids, approximates the coastlines as staircases and prevents flexible implementation of variable resolution. Yet, during the last 40 years, numerical methods have dramatically evolved. It is now time for ocean modeling to benefit from all those advances by developing ocean models using state-of-the-art numerical methods on unstructured grids (Griffies et al. 2009).

Unstructured grid methods are mainly of two kinds: finite volumes and finite elements. In short, finite volumes were first developed for problems predominantly hyperbolic (i.e., dominated by waves or advective transport), while finite element methods were first developed for problems dominated by elliptic terms. The two research communities have evolved toward solutions that manage to treat efficiently both hyperbolic and elliptic problems. Unstructured grid marine modeling is an active area of research for coastal applications (Deleersnijder and Lermusiaux 2008). Indeed, the coastlines must be accurately represented, as they have a much stronger influence at the regional scale than at the global scale. Finite volume methods are now widely used, and models like FVCOM (Finite Volume Community Ocean Model) (Chen et al. 2003) have a large community of users. Many other groups are developing finite volume codes for ocean, coastal, and estuarine areas, such as Fringer et al. (2006), Ham et al. (2005), Stuhne and Peltier (2006), and Casulli and Walters (2000). Nonhydrostatic finite element methods are found in Labeur and Wells (2009) for small-scale problems. For large-scale ocean modeling, continuous finite element methods are used in FEOM (Finite Element Ocean Model) (Wang et al. 2008a, b; Timmermann et al. 2009), and Imperial College Ocean Model (Ford et al. 2004a) relies on mesh adaptivity to capture the multiscale aspects of the flow (Piggott et al. 2008).

In the realm of finite difference methods, Arakawa's C grid allows for a stable and relatively noise-free discretization of the shallow water equations and is now very popular for ocean modeling (Arakawa and Lamb 1977; Griffies et al. 2000). However, the search for an equivalent optimal finite element pair for the shallow water equations is still an open issue. Le Roux et al. (1998) gave the first review of available choices. More recent mathematical and numerical analysis of finite element pairs for gravity and Rossby waves are provided in Le Roux et al. (2007, 2008), Rostand and Le Roux (2008), and Rostand et al. (2008). Hanert et al. (2005) proposed to use the P_1^{NC}–P_1 pair, following Hua and Thomasset (1984). It appears that the P_1^{NC}–P_1 pair is a stable discretization, but its rate of convergence is suboptimal on unstructured grids (Bernard et al. 2008b). Following the same philosophy, the P_1^{DG}–P_2 pair was proposed by Cotter et al. (2009a). Such an element exhibits both stability and optimal rates of convergence for the Stokes problem and the wave equation (Cotter et al. 2009b). A review of the numerical properties of those pairs stabilized by interface terms can be found in Comblen et al. (2010b).

This paper focuses on the development of a marine model, called Second-generation Louvain-la-Neuve Ice–ocean Model (SLIM[1]) that should be able to deal with problems ranging from local and regional scales to global scales. In this model, we choose equal-order discontinuous interpolations for the elevation and velocity fields, as it allows us an efficient and easier implementation. Such an equal-order mixed formulation is stable as the interface stabilizing terms allows us to circumvent the Ladyzhenskaya–Babŭska–Brezzi conditions and to take advantage of the inherent good numerical properties of the discontinuous Galerkin (DG) methods for advection dominated processes. It also allows us to decouple horizontal and vertical dynamics, thanks to the block-diagonal nature of the corresponding mass matrix. DG methods can be viewed as a kind of hybrid between finite elements and finite volumes. They enjoy most the strengths of both schemes while avoiding most of their weaknesses: Advection schemes take into account the characteristic structure of the equations, as for finite volume methods, and the polynomial interpolation used inside each element allows for a high-order representation of the solution. Moreover, no degree of freedom is shared between two geometric entities, and this high level of locality considerably simplifies the implementation of the method. Finally, the mass matrix is block diagonal, and for explicit computations, no linear solver is needed. We also observe a growing interest for the discontinuous Galerkin methods in coastal and estuarine modeling (Aizinger and Dawson 2002; Dawson and Aizinger 2005; Kubatko et al. 2006; Aizinger and Dawson 2007; Bernard et al. 2008a; Blaise et al. 2010). For atmosphere modeling, the high-order capabilities of this scheme are really attractive (Nair et al. 2005; Giraldo 2006), and the increasing use of DG follows the trend to replace spectral transform methods with local ones.

Herein, we provide the detailed description of the spatial discretization used in our discontinuous Galerkin finite element marine model SLIM, as well as an illustrative example of the ability of the model to represent complex baroclinic flows. Section 2 describes the partial differential equations considered. Sections 3 and 4 provide the numerical tools needed to derive an efficient stable and accurate discrete formulation. Section 5 details the discrete discontinuous formulation. Finally, in Section 6, we study the internal waves generated in the lee of an isolated seamount as computed with our model. In a companion paper, the time integration procedure will be discussed.

[1] http://www.climate.be/slim

2 Governing equations

Large-scale ocean models usually solve the hydrostatic Boussinesq equations for the ocean. As a result of the hydrostatic approximation, the vertical momentum equation is reduced to a balance between the pressure gradient force and the weight of the fluid. The conservation of mass degenerates into volume conservation, and the density variations are taken into account in the buoyancy term only. This set of equations is defined on a moving domain, as the sea-surface evolves according to the flow.

Using the notations given in Table 1, the governing equations read:

– Horizontal momentum equation:

$$\frac{\partial \boldsymbol{u}}{\partial t}+\nabla_h\cdot(\boldsymbol{u}\boldsymbol{u})+\frac{\partial(w\boldsymbol{u})}{\partial z}+f\mathbf{e}_z\wedge\boldsymbol{u}+\frac{1}{\rho_0}\nabla_h p+g\nabla_h\eta = \nabla_h\cdot(\nu_h\nabla_h\boldsymbol{u})+\frac{\partial}{\partial z}\left(\nu_v\frac{\partial \boldsymbol{u}}{\partial z}\right), \quad (1)$$

Table 1 Notations for the governing equations of the three-dimensional baroclinic marine model

Coordinates and spatial operators	
x, y	Horizontal coordinates
z	Vertical coordinate, pointing upwards with its origin at the sea surface at rest
∇_h	Horizontal gradient operator
$\boldsymbol{e}_z$	Upward unit normal
$\wedge$	Cross product symbol
Material parameters or functions	
g	Gravitational acceleration
ρ_0	Reference density
f	Coriolis parameter
h	Depth at rest
ν_h	Horizontal turbulent viscosity parameter
ν_t	Vertical turbulent viscosity parameter
κ_h	Horizontal turbulent diffusivity parameter
κ_t	Vertical turbulent diffusivity parameter
Variables	
$\boldsymbol{u}$	Horizontal three-dimensional velocity vector
w	Vertical three-dimensional velocity vector
$\boldsymbol{u}^\eta$	Surface horizontal three-dimensional velocity vector
w^η	Surface vertical three-dimensional velocity vector
$\boldsymbol{u}^{-h}$	Bottom horizontal three-dimensional velocity vector
w^{-h}	Bottom vertical three-dimensional velocity vector
η	Sea surface elevation
p	Baroclinic pressure
$\boldsymbol{p}$	Baroclinic pressure gradient
c	Three-dimensional tracer, can be S or T
S	Salinity
T	Temperature

– Vertical momentum equation:

$$\frac{\partial p}{\partial z}=-g\rho(T,S) \qquad \text{with} \qquad \rho=\rho_0+\rho'(T,S). \quad (2)$$

– Continuity equation:

$$\nabla_h\cdot\boldsymbol{u}+\frac{\partial w}{\partial z}=0, \quad (3)$$

– Free-surface equation:

$$\frac{\partial \eta}{\partial t}+\boldsymbol{u}^\eta\cdot\nabla_h\eta-w^\eta=0. \quad (4)$$

– Tracer equation:

$$\frac{\partial c}{\partial t}+\nabla_h\cdot(\boldsymbol{u}c)+\frac{\partial(wc)}{\partial z}=\nabla_h\cdot(\kappa_h\nabla_h c)+\frac{\partial}{\partial z}\left(\kappa_v\frac{\partial c}{\partial z}\right), \quad (5)$$

This set of equations defines the mathematical three-dimensional baroclinic marine model and must be solved simultaneously with the suitable initial and boundary conditions. However, no models solve the primitive equations simultaneously. In practice, dynamic and thermodynamic equations are always decoupled.

In order to build a numerical discrete spatial scheme, it is usual to associate the unknown field with a given equation. The horizontal velocity $\boldsymbol{u}(x, y, z, t)$ is obtained from the horizontal momentum equation (Eq. 1), while the vertical velocity $w(x, y, z, t)$ is deduced from the continuity equation (Eq. 3). The three-dimensional tracers $c(x, y, z, t)$, which can be T or S, are associated with the tracer equation (Eq. 5). The density deviation $\rho'(x, y, z, t)$ is then deduced from temperature and salinity using an appropriate equation of state $\rho'(T, S)$. As we only need the pressure gradient and not the pressure itself, we follow Wang et al. (2008b) and only calculate the pressure gradient $\boldsymbol{p}(x, y, z, t) = \nabla_h p(x, y, z, t)$ from

$$\frac{\partial \boldsymbol{p}}{\partial z}=-g\nabla_h\rho'(T,S), \quad (6)$$

which is the horizontal gradient of equation (Eq. 2), as ρ_0 is constant. Such an approach allows us to partly circumvent the numerical inaccuracies observed in the calculation of the baroclinic pressure gradient term in the momentum equation with a transformed vertical coordinate (e.g., sigma coordinates; Deleersnijder and Beckers 1992; Haney 1991). Finally, the sea elevation $\eta(x, y, t)$ can be deduced from a modified form of the free-surface equation (Eq. 4) which specifies the associated impermeability condition at the sea surface.

Integrating the continuity equation (Eq. 3) along the vertical direction yields

$$w^{\eta} - w^{-h} + \int_{-h}^{\eta} \nabla_h \cdot \boldsymbol{u} dz = 0,$$

which may be transformed to:

$$\underbrace{\frac{\partial \eta}{\partial t} + \boldsymbol{u}^{\eta} \cdot \nabla_h \eta}_{w^{\eta}} - \underbrace{\boldsymbol{u}^{-h} \cdot \nabla_h (-h)}_{w^{-h}} + \int_{-h}^{\eta} \nabla_h \cdot \boldsymbol{u} dz = 0,$$

where we substitute the vertical velocities by using both associated impermeability conditions at the sea surface and the sea bottom. Applying the Leibniz integral rule leads to

$$\frac{\partial \eta}{\partial t} + \nabla_h \cdot \int_{-h}^{\eta} \boldsymbol{u} dz = 0. \tag{7}$$

The integral free-surface equation (Eq. 7) exactly corresponds to the mass conservation of the shallow water equations. This prompts us to use this form rather than the local form. Mode-splitting procedures, where three-dimensional baroclinic and two-dimensional barotropic modes are time-stepped separately, are often resorted to. Therefore, it is very useful to have the same free-surface equation for the three-dimensional and two-dimensional formulations. The three-dimensional equations are designed so that the discretely depth-integrated equations are as close as possible to an accurate discretization of the shallow water equations. To achieve this, most ocean models rely on the integral form of the free-surface equation for the baroclinic three-dimensional model.

Finally, a very common approximation consists in substituting Eq. 7 by the *linear free surface equation* defined by :

$$\frac{\partial \eta}{\partial t} + \nabla_h \cdot \int_{-h}^{0} \boldsymbol{u} dz = 0., \tag{8}$$

This equation simply corresponds to the mass conservation equation of the linear shallow water problem and describes the surface height of the flow with a linear free surface approximation. Such an approximation cannot be used in coastal flows but is very convenient in large-scale applications.

3 Geometrical numerical tools

Before deriving the discrete formulation, we first present the geometrical tools that are valid for all finite element (continuous or discontinuous) discretizations.

- The computational domain evolves with time, and it is required to take into account the evolution of the domain in the discrete model. In Section 3.1, the standard ALE technique (arbitrary Lagrangian–Eulerian) implemented in the model is described.
- Moreover, the computational domain lies on a sphere. In Section 3.2, we recall the algorithm that renders the model able to operate on any manifold including a sphere or a planar surface.

3.1 Arbitrary Lagrangian–Eulerian methods

As the variation of the sea surface elevation modifies the domain of integration, the position of the nodes at the sea surface will move in the vertical direction as prescribed by the elevation field η. Moving the free surface nodes without changing the position of interior nodes will lead to unacceptable element distortions along the sea surface. Then, we must propagate the motion of the boundary nodes into the domain by means of a *moving mesh algorithm*. Its purpose is to avoid mesh distortion due to the sea surface motion and to maintain the original element density in the deformed mesh. In the model, the computational domain is stretched uniformly in the vertical direction. If we denote z^* the original vertical coordinate of the nodes in the initial reference fixed domain Ω^* (the computational domain with the sea surface at rest), the vertical position $z(x, y, z^*, t)$ and the vertical velocity $w_z(x, y, z^*, t)$ of the nodes in the moving domain $\Omega(t)$ are prescribed by

$$z(x, y, z^*, t) = z^* + \eta(x, y, t) \frac{h(x, y) + z^*}{h(x, y)}, \tag{9}$$

$$\underbrace{\frac{\partial z}{\partial t}(x, y, z^*, t)}_{w_z} = \frac{\partial \eta}{\partial t}(x, y, t) \frac{h(x, y) + z^*}{h(x, y)}. \tag{10}$$

The mesh velocity w_z is relative to the motion of the mesh that is stretched uniformly only to maintain the original aspect ratio in the deformed mesh. Such a velocity is fully independent of the Lagrangian velocity of the fluid particle. This velocity can be chosen in order to maintain the mesh quality and, in this sense, can be viewed as arbitrary with respect to the motion of the fluid. This is why such an approach is usually called an arbitrary Lagrangian–Eulerian method. Dealing with a moving domain requires the modification of the material derivative of a field c

$$\frac{Dc}{Dt} = \frac{\partial c}{\partial t} + \boldsymbol{u} \cdot \nabla_h c + (w - w_z) \frac{\partial c}{\partial z}. \tag{11}$$

Therefore, the tracer equation (Eq. 5) has to be modified to take into account the moving mesh algorithm

$$\frac{\partial c}{\partial t} + \nabla_h \cdot (\boldsymbol{u}c) + \frac{\partial (wc - w_z c)}{\partial z} + c\frac{\partial w_z}{\partial z} = \nabla_h \cdot (\kappa_h \nabla_h c) + \frac{\partial}{\partial z}\left(\kappa_v \frac{\partial c}{\partial z}\right), \tag{12}$$

where the mesh velocity is subtracted from the vertical fluid velocity in the vertical advection term. The second additional term can be viewed as a correction to the volume modification introduced by the displacement of the moving mesh. The vertical derivative of the mesh velocity can be directly deduced from Eq. 10

$$\frac{\partial w_z}{\partial z} = \frac{\partial \eta}{\partial t}\frac{1}{h}. \tag{13}$$

A similar adaptation must be applied for the horizontal momentum equation (Eq. 1). In other words, all new terms involving the vertical velocity in Eq. 12 have to be included for all material derivatives appearing in discrete equations which are calculated on a moving domain. For sake of brevity, we will not add these terms in the later sections, even if they are really included in the model.

3.2 Dealing with flows on the sphere

The model operates on arbitrarily shaped surfaces, including the sphere or plane surfaces, following Comblen et al. (2009). The basic idea of the procedure is that each local geometrical entity supporting vectorial degrees of freedom has its own Cartesian coordinate system. There are also coordinate systems associated with each vector test and shape functions for the horizontal velocity field. To supply a vectorial degree of freedom from a frame of reference to another, we only need to build local rotation operators.

The global linear system of discrete equations is then formulated in terms of the vector degrees of freedom expressed in their own frame of reference. To build and assemble the local matrix corresponding to an element, we first fetch all the needed vectorial degrees of freedom into the coordinate system of this element, and then we compute the local matrix or vector. We then apply rotation matrices to this matrix so that its lines and columns are expressed in the frame of reference of the corresponding test and shape functions, respectively. The transformation of the local linear system can be expressed in terms of ${}_x\mathbf{P}_\xi$ and ${}_\xi\mathbf{P}_x$, the rotation matrices from and to the frame of reference in which the integration is performed, respectively (Comblen et al. 2009).

All the transfer matrices operators are computed once at the initialization of the algorithm. The local system of discrete equations in the basis of the element can be written as:

$${}_\xi\mathbf{A}_\xi \; {}_\xi\mathbf{U} = {}_\xi\mathbf{B}$$

where ${}_\xi\mathbf{A}_\xi$ and ${}_\xi\mathbf{B}$ are the local discrete matrix and right hand in the basis of the element. The local vector of unknowns is denoted ${}_\xi\mathbf{U}$ in this local basis. Once this system is computed, the system can be rewritten with velocities and test functions expressed in the basis of the degrees of freedom:

$$\begin{array}{lll} {}_\xi\mathbf{A}_\xi \; {}_\xi\mathbf{U} & = & {}_\xi\mathbf{B} \\ \downarrow & & \\ {}_\xi\mathbf{A}_\xi \; \overbrace{{}_\xi\mathbf{P}_x \; {}_x\mathbf{U}} & = & {}_\xi\mathbf{B} \\ \underbrace{{}_x\mathbf{P}_\xi \; {}_\xi\mathbf{A}_\xi \; {}_\xi\mathbf{P}_x} \; {}_x\mathbf{U} & = & \underbrace{{}_x\mathbf{P}_\xi \; {}_\xi\mathbf{B}} \\ \downarrow & & \downarrow \\ {}_x\mathbf{A}_x \; {}_x\mathbf{U} & = & {}_x\mathbf{B}. \end{array}$$

Similarly, to assemble local matrices and vectors corresponding to the integral over an interface between two elements with different coordinate systems, we first fetch all the information in the frame of reference of the interface, then compute the integral, and fetch back the lines and columns of the matrices in the frame of reference of the corresponding test and shape functions. All the curvature treatment is embedded in the rotation matrices, and the discrete equations are expressed exactly as if the domain was planar.

With such a procedure, it is possible to solve the equations on the sphere, circumventing completely any possible singularity problem. For notational convenience, the full discrete formulation will be presented within a Cartesian framework, but it is important to note that the model is fully implemented to operate on the sphere.

4 Discontinuous Galerkin methods

Now, let us introduce the finite element mesh and the discrete discontinuous approximations of the field variables of the model (η, $\boldsymbol{u}$, w, c, ρ', $\boldsymbol{p}$). The three-dimensional mesh is made up of prismatic elements, as illustrated in Fig. 1, and is obtained from the extrusion of triangular two-dimensional elements. The vertical length scale is typically much smaller than the

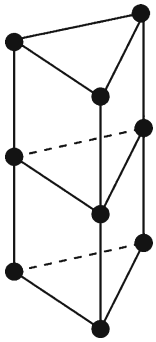

Fig. 1 Sketch of prismatic elements. The vertical length scale is typically much smaller than the horizontal length scale, i.e., the prisms are thin

Table 2 Summary of the finite element spaces used for each field. Triangular linear elements are noted P_1 while vertical linear elements are noted L_1. The superscript DG stands for Discontinuous Galerkin

Field		Finite element space
Free surface elevation	η	P_1^{DG}
Horizontal three-dimensional velocity vector	$\boldsymbol{u}$	$P_1^{\mathrm{DG}} \times L_1^{\mathrm{DG}}$
Vertical three-dimensional velocity	w	$P_1^{\mathrm{DG}} \times L_1^{\mathrm{DG}}$
Three-dimensional tracer	c	$P_1^{\mathrm{DG}} \times L_1^{\mathrm{DG}}$
Density deviation	ρ'	$P_1^{\mathrm{DG}} \times L_1$
Baroclinic pressure gradient	$\boldsymbol{p}$	$P_1^{\mathrm{DG}} \times L_1$

horizontal length scale. In other words, the prisms are thin. We choose prismatic elements to obtain a mesh unstructured in the horizontal direction and structured in the vertical direction. On the sphere, these columns of prisms are obtained by extruding the surface triangles in the direction normal to these triangles. As the extrusion is parallel, the prisms have the same width at the sea surface and at the sea bottom. This alignment of the elements along the vertical axis allows natural treatment of the continuity equation (Eq. 3) and the pressure gradient equation (Eq. 6) that can be integrated along the vertical direction.

The three-dimensional fields ($\boldsymbol{u}$, w, c, ρ', $\boldsymbol{p}$) are discretized on the mesh of prisms. The two-dimensional elevation field η is discretized onto the two-dimensional mesh of triangles. The shape functions for $\boldsymbol{u}$, w, c are obtained as the tensorial product of the linear discontinuous triangle P_1^{DG} and the linear discontinuous one-dimensional element L_1^{DG}. The shape functions of the density deviation and the baroclinic pressure gradient are $P_1^{\mathrm{DG}} \times L_1$. The use of different discrete vertical spaces for ρ' and the baroclinic pressure gradient $\boldsymbol{p}$ can be viewed as an appropriate way to average the vertical variation of the tracers in the calculation of the baroclinic pressure gradient. The summary of the finite element spaces is given in Table 2. For the procedure to simulate flow on the sphere, each column of prisms defines the basic geometrical entity to assemble. It has its own coordinate system, as the degrees of freedom for a discontinuous approximation are all associated with elements and not with interfaces or vertices. Finally, we will describe in details the two most important aspects of the spatial discretization :

– Choosing the way to compute the discrete values at the inter-element interfaces is the critical ingredient to obtain a stable and accurate discrete formulation in the framework of the DG methods. The discrete fields are dual-valued at the inter-element interfaces. For the advective fluxes at these interfaces, the values of the variables are obtained on Riemann solvers applied to the hyperbolic terms of the model. Details about the Riemann solvers are given in Section 4.1.
– Incorporating the diffusion operators inside a DG formulation also requires special care. We use the Symmetric Interior Penalty Galerkin (SIPG) technique to accommodate diffusion operators. Moreover, the mathematical formulation exhibits anisotropic diffusion and the algorithm is adapted by computing the interior penalty coefficients on a virtual stretched geometry. The methodology used is summarized in Section 4.2.

4.1 Riemann solvers

A two-dimensional set of barotropic discrete equations can be obtained by the vertical depth integration (or the algebraic stacking of the resulting lines and columns of the global system) of the three-dimensional set of baroclinic discrete equations. The basic idea of this procedure is to define the lateral interface in the discrete three-dimensional baroclinic equations in such a way that the corresponding two-dimensional discretization by depth integration is a robust stable formulation. In particular, the use of the integral free-surface equation (Eq. 7) and the selection of the discrete spaces should lead to a stable and accurate corresponding two-dimensional discrete problem. Here, the resulting corresponding discrete problem is close to the discretization of the shallow water equations with P_1^{DG} shape functions for the two-dimensional velocities and elevation. A robust formulation can be derived for this problem, following Comblen et al. (2010b).

The key ingredient of a stable and accurate DG formulation is the choice of the definition for a common value for the variables along the interfaces. It is necessary to define adequately these common values with a Riemann solver. For nonlinear problems, it can be quite complicated to compute the exact Riemann solver, and

approximate Riemann solvers are usually resorted to. For the shallow water equations, approximate Riemann solvers are deduced from the conservative form of the equations (LeVeque 2002; Toro 1997; Comblen et al. 2010b).

For the linear shallow water equations, the exact solver yields the following interfacial values:

$$\boldsymbol{u}^{\text{riemann}} \cdot \mathbf{n} = \{\boldsymbol{u}\} \cdot \mathbf{n} + \sqrt{\frac{g}{h}}[\eta], \tag{14}$$

$$\eta^{\text{riemann}} = \{\eta\} + \sqrt{\frac{h}{g}}[\boldsymbol{u}] \cdot \mathbf{n}, \tag{15}$$

where { } and [ ] denote the mean and the jump operators, respectively. The jump is defined by $[a] = (a_L - a_R)/2$ and the mean by $\{a\} = (a_L + a_R)/2$ where a_L and a_R are the left and right values of the field a. The vector $\mathbf{n}$ is the rightward normal corresponding to the jump operator. The linear shallow water equations are simple wave equations, and those interfacial values correspond only to the terms generating the surface gravity waves and are not valid for the non-linear two-dimensional barotropic problem. However, as oceanic flows usually exhibit small Froude numbers, the linear Riemann solver can be viewed as a very good approximation of the nonlinear solver.

The three-dimensional equations allow several hyperbolic phenomena to take place. Surface gravity waves are the fastest phenomena, with phase speed $\sqrt{gh}$. The second fastest phenomena are the internal gravity waves. Their propagation speed depends on the stratification. It can be as large as a few meters per second. As the set of the three-dimensional baroclinic marine flow equations cannot be cast into a conservative form, it is not possible to deduce an approximate Riemann solver such as the Roe solver by simply linearizing the problem. Therefore, we selected a Lax–Friedrichs flux. This flux is commonly used due to its simplicity. The flux is simply defined as the sum of the mean flux and the jump of the variables multiplied by an upper bound γ on the phase speed of the fastest internal wave (the maximum eigenvalue of the Jacobian matrix):

$$\text{flux}^{\text{lax–friedrichs}} = \{\text{flux}\} + \gamma \left[\text{variable}\right] \tag{16}$$

In this work, we use the Riemann solver of the linear shallow-water equations (Eqs. 14 and 15) for the terms corresponding to surface gravity waves (i.e., elevation gradient in the two-dimensional, depth-averaged momentum equation and velocity divergence in the continuity equation (Comblen et al. 2010a)). In the three-dimensional problem, i.e., in the momentum equation and in the active tracers equations (typically temperature and salinity), we add to the mean flux the jump of the variables multiplied by an upper bound on the second fastest wave, which is the sum of the fastest internal wave phase speed and the advection velocity. Determining the phase speed of the fastest internal wave is not easy for a complicated stratification profile.

In our examples, we simply use values of γ deduced from some numerical experiments, by a trial and error procedure. The selection of the coefficient γ is a key ingredient of the numerical technique. In fact, in our computations, we select an ad hoc coefficient that is selected with some physical intuition. However, it is not a robust procedure and this can prevent the model from being used if there are no general solutions for it. In other words, this problem is not fully solved but the importance of Riemann solvers in shallow water models were discussed in Comblen et al. (2010b).

4.2 Symmetric interior penalty Galerkin methods

In the realm of discontinuous Galerkin methods, various discretizations of the Laplacian operator are reviewed in, e.g., Arnold et al. (2002). Two of them are especially popular: the interior penalty methods (Riviere 2008) and the local DG method (Cockburn and Shu 1998).

To accurately handle the diffusion terms, we use the SIPG technique. Basically, the weak form of the Laplace equation $\nabla^2 c = 0$ can be obtained by multiplying this equation with a test function and by integrating this product on the whole domain. Integrating by parts and choosing the mean values at the interfaces yields:

$$\sum_{e=1}^{N_e} \Big[\ll \hat{c}\,\{\nabla c\} \cdot \boldsymbol{n} \gg_e - < \nabla \hat{c} \cdot \nabla c >_e \Big] = 0, \tag{17}$$

where $<>_e$ and $\ll\gg_e$ denote respectively the integral over the element Ω_e and its boundary. The number of elements is N_e. Choosing the mean values at the interface seems natural for an elliptic operator where the information propagates along all directions. However, such a simple and intuitive treatment of the Laplacian operator is incomplete. Indeed, the discrete solution is not unique, as at the boundary of each element, only Neumann boundary conditions are prescribed. In

order to partly complete the discrete formulation, the Incomplete Interior Penalty Galerkin method (IIPG) consists in adding a penalty term on the discontinuities of the field at the inter-element interfaces

$$\sum_{e=1}^{N_e}\Big[\ll \hat{c}\,\{\nabla c\}\cdot\boldsymbol{n}\gg_e+\sigma\ \ll \hat{c}\,[c]\gg_e-<\nabla\hat{c}\cdot\nabla c>_e\Big]=0, \tag{18}$$

where σ is a penalty parameter scaled in such a way that $\sigma\,[c]$ is a term similar to a gradient, at the interface level. In other words, $1/\sigma$ has to be an suitable length scale. It is shown in Riviere (2008) that this formulation provides optimal results when shape functions of odd polynomial order are used (i.e., P_1, P_3, ...) but lacks convergence for even polynomial orders. Further, the resulting discrete matrix is not symmetric, while the continuous operator is symmetric. The SIPG solves both of these issues (Riviere 2008), by adding a symmetrizing term to the IIPG formulation (Eq. 18).

$$\sum_{e=1}^{N_e}\Big[\ll \hat{c}\,\{\nabla c\}\cdot\boldsymbol{n}\gg_e+\ll\nabla\hat{c}\cdot\boldsymbol{n}\,[c]\gg_e +\sigma\ll\hat{c}\,[c]\gg_e-<\nabla\hat{c}\cdot\nabla c>_e\Big]=0. \tag{19}$$

There is a lower bound on σ that ensures optimal convergence. This bound must be as tight as possible, as the larger the value of σ, the worse the conditioning of the operator. Shahbazi (2005) suggests to use the following formula:

$$\sigma=\left[\frac{2(k+1)(k+3)}{3}\frac{A(\Gamma_k)}{V(\Omega_e,\Omega_f)}\right], \tag{20}$$

where k is the order of the interpolation, $A(\Gamma_k)$ is the area of the interface Γ_k between the two considered elements, and $V(\Omega_e,\Omega_f)$ is the mean volume of the two neighboring elements Ω_e and Ω_f.

Finally, the diffusion terms are split into a horizontal part and a vertical part, preventing the implementation of rotated diffusion tensors as described for instance by Redi (1982). Such rotated diffusion is not implemented in the current SLIM model. Distinct viscosity and diffusivity coefficients are chosen to represent the different effects of each of the many unresolved physical processes. Both the diffusion operator and the mesh can be anisotropic. A simple procedure consists in virtually stretching the mesh in the vertical direction so that we recover an isotropic diffusion in the deformed geometry. The mesh is not really modified, but the local interior penalty coefficients are chosen in such a way that they correspond to an isotropic diffusion on the modified mesh.

5 Discrete DG finite element formulations

For the sake of completeness, we provide here the full weak DG finite element formulations for each equation of the SLIM model (Eqs. 1, 3, 5, 6, and 7) using the numerical tools described in both previous sections. The discrete formulations can be then derived by replacing the test functions and the solution by the corresponding DG polynomial approximation.

5.1 Horizontal momentum equation

The discrete formulation of the horizontal momentum equation is obtained by multiplying Eq. 1 by a test function $\widehat{\boldsymbol{u}}$ and integrating over the whole domain Ω:

$$\begin{aligned}&<\widehat{\boldsymbol{u}}\cdot\frac{\partial\boldsymbol{u}}{\partial t}>+<\widehat{\boldsymbol{u}}\cdot(\nabla_h\cdot(\boldsymbol{u}\boldsymbol{u}))>+<\widehat{\boldsymbol{u}}\cdot\frac{\partial(w-w_z)\boldsymbol{u}}{\partial z}>\\&+<\widehat{\boldsymbol{u}}\cdot f\mathbf{e}_z\wedge\boldsymbol{u}>+<\widehat{\boldsymbol{u}}\cdot\frac{\boldsymbol{p}}{\rho_0}>+<\widehat{\boldsymbol{u}}\cdot g\nabla_h\eta>\\&-<\widehat{\boldsymbol{u}}\cdot(\nabla_h\cdot(\nu_h\nabla_h\boldsymbol{u}))>-<\widehat{\boldsymbol{u}}\cdot\frac{\partial}{\partial z}\left(\nu_v\frac{\partial\boldsymbol{u}}{\partial z}\right)>=0,\end{aligned} \tag{21}$$

where $<\ >$ denotes the volume integral over the domain Ω. In order to be able to introduce discontinuous approximations, this integral is split into N_e integrals on each element Ω_e.

$$\sum_{e=1}^{N_e}\Bigg[<\widehat{\boldsymbol{u}}\cdot\frac{\partial\boldsymbol{u}}{\partial t}>_e+\underbrace{<\widehat{\boldsymbol{u}}\cdot(\nabla_h\cdot(\boldsymbol{u}\boldsymbol{u}))>_e}_{\text{horizontal advection}}+\underbrace{\widehat{\boldsymbol{u}}\cdot\frac{\partial(w-w_z)\boldsymbol{u}}{\partial z}}_{\text{vertical advection}}+\underbrace{<\widehat{\boldsymbol{u}}\cdot f\mathbf{e}_z\wedge\boldsymbol{u}>_e}_{\text{Coriolis}}$$
$$+\underbrace{<\widehat{\boldsymbol{u}}\cdot\frac{\boldsymbol{p}}{\rho_0}>_e}_{\text{baroclinic pressure gradient}}+\underbrace{<\widehat{\boldsymbol{u}}\cdot g\nabla_h\eta>_e}_{\text{elevation gradient}}-\underbrace{<\widehat{\boldsymbol{u}}\cdot(\nabla_h\cdot(\nu_h\nabla_h\boldsymbol{u}))>_e}_{\text{horizontal diffusion}}-\underbrace{<\widehat{\boldsymbol{u}}\cdot\frac{\partial}{\partial z}\left(\nu_v\frac{\partial\boldsymbol{u}}{\partial z}\right)>_e}_{\text{vertical diffusion}}\Bigg]=0. \tag{22}$$

Apart from the baroclinic pressure terms, all terms containing spatial derivatives are integrated by parts on each element. Therefore, boundary fluxes appear along the interfaces between the elements. If we use discontinuous approximations, the key ingredient of the weak formulation is the way to define those fluxes as the variables are not uniquely defined on those interfaces. Each term of Eq. 22 is then derived as follows:

– Horizontal advection:

$$\sum_{e=1}^{N_e}\left[- < \nabla_h \widehat{\boldsymbol{u}} : \boldsymbol{u}\boldsymbol{u} >_e + \ll \widehat{\boldsymbol{u}} \cdot \{\boldsymbol{u}\}\{\boldsymbol{u}\} \cdot \boldsymbol{n}_h \gg_e + \ll \gamma\, [\boldsymbol{u}] \cdot \widehat{\boldsymbol{u}} \gg_e \right]$$

– Vertical advection:

$$\sum_{e=1}^{N_e}\left[- < \frac{\partial \widehat{\boldsymbol{u}}}{\partial z} \cdot (w - w_z)\boldsymbol{u} >_e + \ll \widehat{\boldsymbol{u}} \cdot (w - w_z)^{\text{down}}\, \boldsymbol{u}^{\text{upwind}}\, n_z \gg_e \right]$$

– Elevation gradient:

$$\sum_{e=1}^{N_e}\left[- < \nabla_h \cdot \widehat{\boldsymbol{u}}\, g\eta >_e + \ll g\eta^{\text{riemann}}\, \widehat{\boldsymbol{u}} \cdot \boldsymbol{n}_h \gg_e \right]$$

– Horizontal diffusion:

$$\sum_{e=1}^{N_e}\left[- < \nu_h\, (\nabla_h \widehat{\boldsymbol{u}}) : (\nabla_h \boldsymbol{u})^T >_e + \ll \nu_h \widehat{\boldsymbol{u}} \cdot \{\nabla_h \boldsymbol{u}\} \cdot \boldsymbol{n} \gg_e + \ll \nu_h \nabla_h \widehat{\boldsymbol{u}} \cdot \boldsymbol{n} \cdot [\boldsymbol{u}] \gg_e + \sigma \ll \nu_h \widehat{\boldsymbol{u}} \cdot [\boldsymbol{u}] \gg_e \right]$$

– Vertical diffusion:

$$\sum_{e=1}^{N_e}\left[- < \nu_v \frac{\partial \widehat{\boldsymbol{u}}}{\partial z} \cdot \frac{\partial \boldsymbol{u}}{\partial z} >_e + \ll \nu_v \widehat{\boldsymbol{u}} \cdot \left\{ \frac{\partial \boldsymbol{u}}{\partial z} \right\} n_z \gg_e + \ll \nu_v \frac{\partial \widehat{\boldsymbol{u}}}{\partial z} n_z \cdot [\boldsymbol{u}] \gg_e + \sigma \ll \nu_v \widehat{\boldsymbol{u}} \cdot [\boldsymbol{u}] \gg_e \right]$$

– Baroclinic pressure gradient and Coriolis:

$$\sum_{e=1}^{N_e}\left[< \widehat{\boldsymbol{u}} \cdot \frac{\boldsymbol{p}}{\rho_0} >_e + f \mathbf{e}_z \wedge \boldsymbol{u} \right]$$

where $\boldsymbol{n}_h = (n_x, n_y)$ and n_z are respectively the horizontal and the vertical components of the outgoing normal of the boundary of the element. In the interface term for vertical advection, we perform upwinding of the advected variable $\boldsymbol{u}$, and we use the vertical velocity of the prism below the interface, so that the discrete vertical advection term is as close as possible to the corresponding term in the continuity equation. As we use the Riemann solver associated to the non-conservative $P_1^{\text{DG}} - P_1^{\text{DG}}$ discrete formulation of the two-dimensional shallow water equations for the gravity waves, it is critical that the resulting two-dimensional discrete equations obtained by integrating the momentum equation along the vertical axis approximately degenerate to this discrete formulation of the two-dimensional shallow water equations. To achieve this, the test function $\widehat{\boldsymbol{u}}$ is now the shape function divided by the depth at rest. Finally, the Lax–Friedrichs flux for the internal waves requires the additional boundary term $\ll \gamma\, [\boldsymbol{u}] \cdot \widehat{\boldsymbol{u}} \gg_e$ where γ is an upper bound of the fastest propagation speed of a three-dimensional phenomena, namely the sum of the phase speed of the fastest internal wave and the advection velocity. The interface terms for horizontal and vertical diffusion are directly derived from the SIPG procedure described by Eq. 19.

5.2 Vertical momentum equation

As the discrete variable associated with the vertical momentum equation is the vector field $\boldsymbol{p}$ that stands for the numerically computed baroclinic pressure horizontal gradient, we discretize the gradient of the balance of the vertical momentum (Eq. 6) as follows:

$$\sum_{e=1}^{N_e}\left[< \widehat{\boldsymbol{p}}^{\text{up}} \cdot \frac{\partial \boldsymbol{p}}{\partial z} >_e \right] = \sum_{e=1}^{N_e}\left[- < \widehat{\boldsymbol{p}}^{\text{up}} \cdot (g \nabla_h \rho'(T, S)) >_e \right]. \tag{23}$$

To take into account that the integration is performed from top to bottom with a constant pressure at the sea surface (and therefore a vanishing pressure gradient), we use some fully upwinded test functions derived as the tensorial product between the usual P_1^{DG} triangle and the upwinded linear unidimensional element, whose values are 1 for the degree of freedom above the element and 0 for the degree of freedom below the element.

5.3 Continuity equation

The continuity equation can be viewed as a steady transport equation along the vertical direction where the divergence of the horizontal velocity acts as a

source term. The continuity equation is used to deduce the vertical velocity by integrating the horizontal velocity divergence from bottom to top. The discrete formulation of the continuity equation is obtained by multiplying Eq. 3 by a test function $\hat{w}$ and integrating over the whole domain Ω:

$$\sum_{e=1}^{N_e}\left[< \hat{w}\frac{\partial w}{\partial z} >_e + < \hat{w}\nabla_h \cdot \boldsymbol{u} >_e \right] = 0. \tag{24}$$

where the integral on the domain Ω is split into N_e integrals on Ω_e. Integrating by parts all terms containing spatial derivatives yields:

$$\sum_{e=1}^{N_e}\left[\ll \hat{w} w^{\text{down}} n_z \gg_e - < \frac{\partial \hat{w}}{\partial z} w >_e + \ll \hat{w}\boldsymbol{u}^{\text{riemann}} \cdot \boldsymbol{n}_h \gg_e - < \nabla_h \hat{w} \cdot \boldsymbol{u} >_e \right] = 0. \tag{25}$$

where we use w^{down} the value from the bottom element at the interfaces between layers of prisms, as the information goes from bottom to top in this pure transport equation. Moreover, a impermeability condition has to be prescribed at the sea bed, namely:

$$w^{-h} + \boldsymbol{u}^{-h} \cdot \nabla_h h = 0, \tag{26}$$

This boundary condition is weakly imposed by using $-\boldsymbol{u}^{-h} \cdot \nabla_h h$ for w^{down} in the first term of Eq. 25 at the sea bed. This only occurs on the first layer of prisms.

In the lateral interface, we use $\boldsymbol{u}^{\text{riemann}}$ because the discrete two-dimensional integral free surface equations will be obtained by aggregating the three-dimensional discrete continuity equations. In fact, the discrete procedure mimics the algebra performed to obtain the integral free-surface equation (Eq. 7) by integrating the equation of continuity (Eq. 3) and substituting the impermeability conditions at both the sea bed and the sea surface. The sea bed impermeability is already included in the discrete formulation of the continuity equation, and the sea-surface condition will be incorporated by the motion of the free surface. Finally, let us recall that when we deduce $\boldsymbol{u}^{\text{riemann}}$ and η^{riemann} with the exact Riemann solver of the linear shallow water equations, we use the fact that the depth integration of the momentum equation coupled with the free surface equation has to degenerate into a stable and an accurate $P_1^{\text{DG}} - P_1^{\text{DG}}$ discretization of the two-dimensional shallow water equations. Therefore, as the discrete free-surface equation will be obtained by aggregating the discrete continuity equation, it is mandatory to use $\boldsymbol{u}^{\text{riemann}}$ here, to have it in the resulting free-surface equation. As a last remark, it is also important to emphasize that the vertical velocity is not a prognostic field. It is a by-product used to deduce the vertical advection terms in the momentum and tracer equations. Elevation, velocities, and tracer are prognostic fields. An accurate DG discretization of our set of equation should be such that these fields are computed with an optimal accuracy, i.e., $p+1$ convergence rate if order p shape functions are used. It is *not* the case for vertical velocity. Vertical velocity is not smooth because it results from the integration of the divergence of the horizontal velocity. It behaves similarly to the volume term from an advection term integrated by parts: It is not smooth, but it does not impair the optimal convergence of the other fields. Indeed, at the discrete level, it is easily seen that, if the prisms are straight, the vertical velocity lives in a discrete space that is piecewise constant in the horizontal direction rather than linear. The smoothness of tracer and horizontal velocity field is recovered as usual in DG, using the interface term, acting as a penalty term.

5.4 Free-surface equation

Formally, the discrete formulation of the free-surface equation is obtained by multiplying Eq. 7 by a test function $\hat{\eta}$ and integrating over the two-dimensional basement of the three-dimensional computational domain Ω. This basement is paved of N_f triangles Δ_f that are the elements of the initial two-dimensional mesh that was extruded to produce the three-dimensional mesh of prisms. This discrete formulation reads:

$$\sum_{f=1}^{N_f}\Bigg[\ll \hat{\eta}\frac{\partial \eta}{\partial t} \gg_{\Delta f} + \underbrace{\ll \hat{\eta}\,\boldsymbol{u}^{\eta} \cdot \nabla_h \eta \gg_{\Delta f} + \ll \hat{\eta}\,\boldsymbol{u}^{-h} \cdot \nabla_h h \gg_{\Delta f} + \ll \hat{\eta}\int_{-h}^{\eta} \nabla_h \cdot \boldsymbol{u} dz \gg_{\Delta f}}_{\text{aggregated discrete continuity equations}} \Bigg] = 0. \tag{27}$$

where $\ll \gg_{\Delta f}$ denotes the integral over the triangle Δ_f of the two-dimensional mesh of N_f triangles.

The last two terms can be obtained by the aggregation of the discrete continuity equation (Eq. 25). It can be shown that the aggregation of unconstrained discrete continuity equation corresponds to the discrete form of the depth-integrated continuity equation :

$$\sum_{f=1}^{N_f}\left[\ll \hat{\eta} \int_{-h}^{\eta} \nabla_h \cdot \boldsymbol{u} dz \gg_{\Delta f} \right].$$

Moreover, imposing the impermeability at the sea bed on the bottom interface will add an additional contribution that corresponds to :

$$\sum_{f=1}^{N_f}\left[\ll \hat{\eta}\, \boldsymbol{u}^{-h}\cdot\nabla_h h \gg_{\Delta f}\right].$$

The opposite sign and the difference between the areas of integration are counterbalanced by the vertical component of the outgoing normal n_z. It is only necessary to add the missing term in order to obtain the full linear free surface equation.

At this stage, the *linear free surface approximation* consists in omitting this additional term or in substituting Eq. 27 by:

$$\sum_{f=1}^{N_f}\Bigg[\ll \hat{\eta}\frac{\partial \eta}{\partial t} \gg_{\Delta f} \underbrace{+\ll \hat{\eta}\, \boldsymbol{u}^{-h}\cdot\nabla_h h \gg_{\Delta f} + \ll \hat{\eta}\int_{-h}^{0}\nabla_h\cdot\boldsymbol{u}dz \gg_{\Delta f}}_{\text{aggregated discrete continuity equation}}\Bigg] = 0. \tag{28}$$

Again, such an approximation may be convenient in some large-scale application but must be avoided in coastal problems. Equation 28 can also be viewed as the discretization of Eq. 8 which is the mass conservation equation of the linear shallow water problem.

5.5 Tracer equation

As for the momentum equation, the weak formulation for the tracer equation can be written on each element as follows:

$$\sum_{e=1}^{N_e}\Bigg[< \hat{c}\frac{\partial c}{\partial t} >_e \underbrace{+ < \hat{c}\ (\nabla_h\cdot(\boldsymbol{u}c)) >_e}_{\text{horizontal advection}} \underbrace{+ < \hat{c}\frac{\partial (w-w_z)c}{\partial z} >_e}_{\text{vertical advection}} \underbrace{- < \hat{c}\ (\nabla_h\cdot(\kappa_h\nabla_h c)) >_e}_{\text{horizontal diffusion}} \underbrace{- < \hat{c}\ \frac{\partial}{\partial z}\left(\kappa_v\frac{\partial c}{\partial z}\right) >_e}_{\text{vertical diffusion}}\Bigg] = 0. \tag{29}$$

We integrate by parts the transport and diffusion terms and choose the suitable values for the interface terms. As for the momentum equation, we add the interface term $\ll \hat{c}\gamma\,[c] \gg_e$ that is deduced from the Lax–Friedrichs solver for internal waves. It is a very important term for the numerical properties of the model as internal waves are a phenomenon that couples momentum and tracers. Each term of Eq. 29 is then derived as follows:

– Horizontal advection:

$$\sum_{e=1}^{N_e}\Big[- < \nabla_h\hat{c}\cdot\boldsymbol{u}c >_e + \ll \hat{c}\,\{c\}\,\boldsymbol{u}^{\text{riemann}}\cdot\boldsymbol{n}_h \gg_e + \ll \hat{c}\gamma\,[c] \gg_e\Big]$$

– Vertical advection:

$$\sum_{e=1}^{N_e}\Big[- < \frac{\partial\hat{c}}{\partial z}(w-w_z)c >_e + \ll \left(\hat{c}\,(w-w_z)^{\text{down}}c^{\text{upwind}}\right)n_z \gg_e\Big]$$

– Horizontal diffusion:

$$\sum_{e=1}^{N_e}\Big[- < \kappa_h\left(\nabla_h\hat{c}\right)\cdot(\nabla_h c) >_e + \ll \kappa_h\hat{c}\,\{\nabla_h c\}\cdot\boldsymbol{n} \gg_e + \ll \kappa_h\nabla_h\hat{c}\cdot\boldsymbol{n}\cdot[c] \gg_e + \sigma \ll \kappa_h\hat{c}\,[c] \gg_e\Big]$$

– Vertical diffusion:

$$\sum_{e=1}^{N_e}\Big[- < \kappa_v\frac{\partial\hat{c}}{\partial z}\frac{\partial c}{\partial z} >_e + \ll \kappa_v\hat{c}\left\{\frac{\partial c}{\partial z}\right\}n_z \gg_e + \ll \kappa_v\frac{\partial\hat{c}}{\partial z}n_z\,[c] \gg_e + \sigma \ll \kappa_v\hat{c}\,[c] \gg_e\Big]$$

To ensure consistency, it is mandatory that the discrete advection term degenerates to the continuity equation when a unit tracer concentration is considered (White et al. 2008). Therefore, the same discrete space must be used for both c and w. Moreover, we must use the same velocity approximations in the interface terms $\boldsymbol{u}^{\text{riemann}}$ for the horizontal advection and w^{down} for the vertical advection. But the interface value for the tracer concentration c at the interface is not constrained by consistency considerations: Upwind or centered values can be used. The additional term from the Lax–Friedrichs solver does not impair consistency as it is exactly nil for a constant tracer. The bottom boundary conditions must also be compatible, this being ensured by suppressing the boundary terms for advection at the sea bottom.

5.6 Validation and mesh refinement analysis

As a first numerical check, we consider a simple gravity waves problem. The domain is an rectangular cuboid ocean of $[0, L] \times [0, L] \times [0, H]$. The length and depth are respectively 10^3 and 1 km. An analytical solution can be derived for the initial elevation given by:

$$\eta(x, y, t = 0) = \sin^2\left(\frac{x\pi}{L}\right) \sin^2\left(\frac{y\pi}{L}\right).$$

The maximal value reached at the center of the square is 1 m. Even if this problem is two-dimensional, it can be a quite efficient test for a three-dimensional if we use several layers with nonconstant depths as illustrated in Fig. 2. If the fluxes through horizontal and vertical faces are not accurately computed, the behavior will be strongly affected and the theoretical rates of convergence cannot be achieved. We deliberately select highly unstructured pattern in order to check that the code converges for both regular and irregular grids. The same convergence test is also obtained for uniformly refined graded mesh. In Fig. 3, we observe the theoretical quadratic rate of convergence for a time step of 250 s and four meshes with following characteristic lengths $h = 200, 100, 50$, and 25 km. The time step was selected

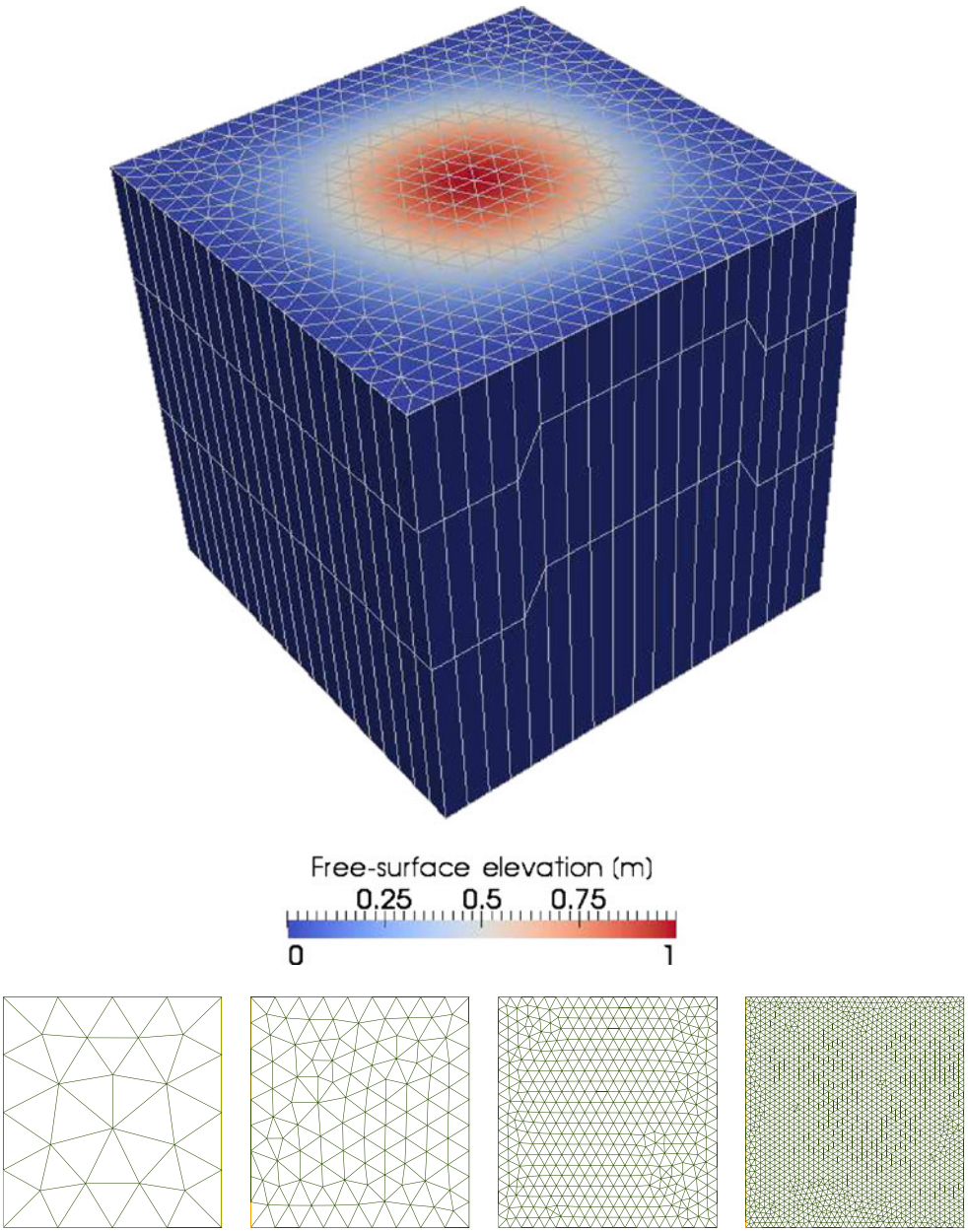

Fig. 2 Gravity waves for a cuboid ocean of $1{,}000 \times 1{,}000 \times 1$ km: mesh refinement analysis with four meshes exhibiting three layers of different depths. Initial elevation (*top*) and two-dimensional traces of the used meshes (*bottom*). The characteristic lengths are $h = 200, 100, 50$, and 25 km, respectively

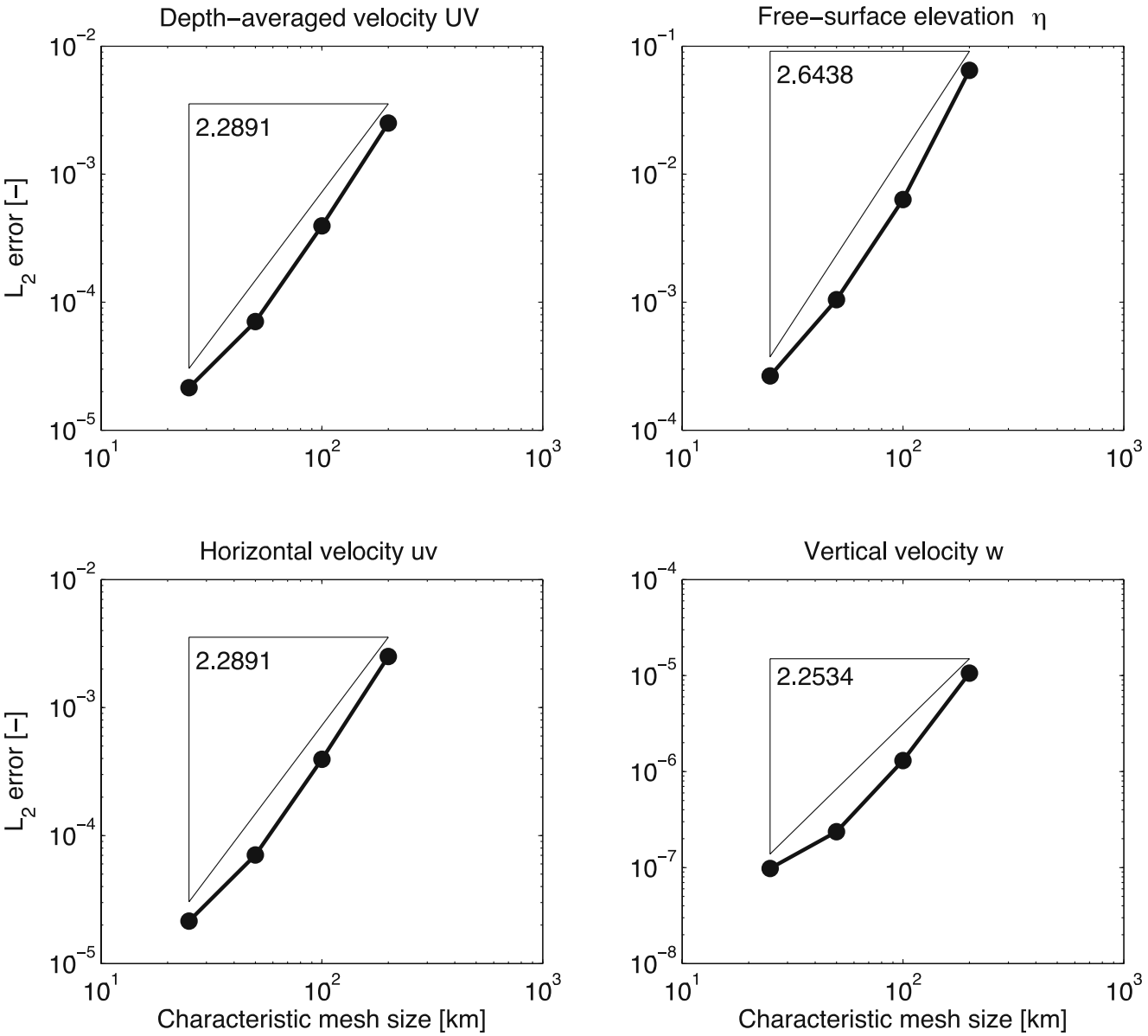

Fig. 3 Gravity waves test case: mesh refinement analysis

in such a way that the temporal error is negligible with respect to the spatial error.

6 Numerical results

We simulate the internal waves in the lee of a moderately tall seamount. The simulation of such a complex flow can be considered as a relevant test case. Complicated phenomena can be observed in the wake of mountains, such as internal wave structures and vortex streets (Chapman and Haidvogel 1992, 1993; Ford et al. 2004b). Such a problem was simulated with three-dimensional baroclinic finite difference (Huppert and Bryan 1976; Chapman and Haidvogel 1992, 1993), finite volume (Adcroft et al. 1997), and finite element models (Ford et al. 2004b; Wang et al. 2008a, b). If the seamount is small enough, a complicated structure of standing internal waves can develop in the wake of the seamount. Chapman and Haidvogel (1993) provide a detailed numerical study of internal lee waves trapped over isolated Gaussian-shaped seamounts. Such a test-case is also used by Ford et al. (2004b) to assess the qualities and drawbacks of their model. With our three-dimensional baroclinic marine model SLIM, we simulate the internal lee waves for a seamount whose height is 30% of the total depth.

The setup of the problem can be summarized as follows: A Gaussian seamount is located at a latitude of 45° N with a bathymetry given by:

$$1-\frac{h(x,y)}{H}=\delta\exp\left(\frac{\left(x-\frac{R}{2}\right)^2+\left(y-\frac{R}{2}\right)^2+\left(z-\frac{R}{\sqrt{2}}\right)^2}{-L^2}\right), \tag{30}$$

where $H=4.5$ km is the maximum depth, $\delta=0.3$ is the relative height of the seamount, $R=6{,}372$ km is Earth radius, and $L=25$ km is the length scale of the seamount. The coordinates x, y, and z are relative to the global Cartesian frame of reference. The flow simulation is initiated with a global zonal geostrophic equilibrium ignoring the seamount. In other words, the initial guess of the calculation is selected as in the testcase #5 of Williamson et al. (1992) where the velocity field only exhibits a nonvanishing east component u_e. In this testcase, the elevation and velocity fields are shown in Fig. 4 and are respectively given by

$$\frac{\eta}{U^2/g}=-\frac{z^2}{R^2}\left(1+\sqrt{2}\,\frac{R\Omega}{U}\right), \tag{31}$$

$$\frac{u_e}{U}=\sqrt{\frac{x^2+y^2}{R^2}}, \tag{32}$$

where $U=0.516$ m s^{-1} is the velocity scale at a latitude of 45° N, $\Omega=7.292\times10^{-5}$ s^{-1} is Earth rotation rate, and $g=9.81$ m s^{-2} is the gravitational acceleration. We consider the density deviation ρ' as the unique tracer of the model, and the initial value of the density deviation is a linear function of the vertical coordinate, with vanishing mean. The derivative of the density with respect to the vertical coordinate is given by $\partial\rho/\partial z=-3.43\times10^{-5}$ kg m^{-4}, and the reference density is selected as $\rho_0=1{,}025$ kg m^{-3}. The turbulent viscosities and diffusivities are given by: $\nu_h=\kappa_h=12.9$ m^2 s^{-1}, $\nu_v=0.0001$ m^2 s^{-1}, and $\kappa_v=0$. With those parameters,

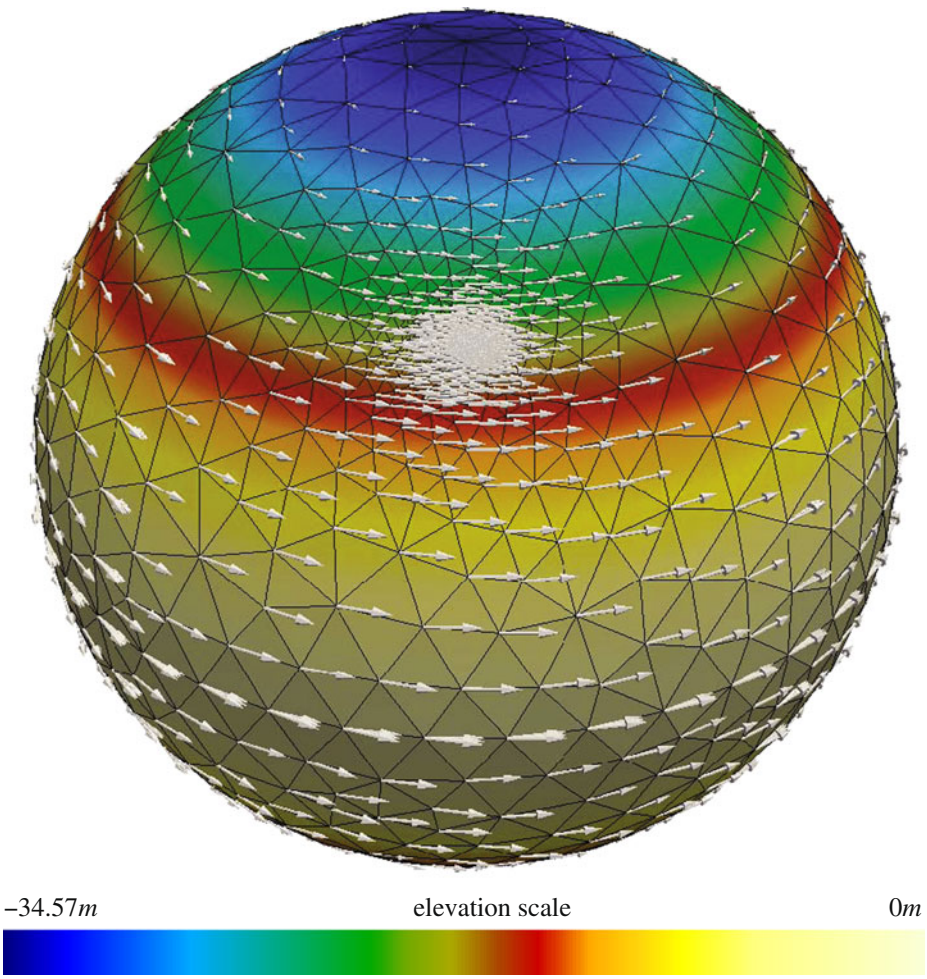

Fig. 4 Initial condition for sea-surface elevation and velocities. The mesh is refined in the lee of the seamount

the flow is characterized by the same four dimensionless numbers as that of the Section 3d of Ford et al. (2004b). These dimensionless numbers are defined as follows:

$$\text{Seamount ratio} \quad \delta = 0.3$$
$$\text{Rossby number} \quad Ro = \frac{U}{fL} = 0.2$$
$$\text{Reynolds number} \quad Re = \frac{UL}{\nu_h} = 1{,}000$$
$$\text{Burger number} \quad Bu = \frac{NH}{fL} = \sqrt{\frac{-g}{\rho_0}\frac{\partial\rho}{\partial z}}\frac{H}{fL} = 1$$

where N is the Brunt–Väisälä frequency.

The computational domain is the whole sphere, as we can take advantage of a highly variable mesh density. It allows us to avoid open boundary conditions, while previous publications use rectangular domains, with imposed inflow, sponge layer as outflow condition, and lateral walls (Chapman and Haidvogel 1992). Figure 5 shows a close-up view of the mesh and the bathymetry near the seamount. The mesh resolution is refined in the lee of the seamount to capture accurately the flow structure. Indeed, we know a priori that the structures generated in the lee of the seamount are deviated to the right, due to the mean transverse flow generated between the two vortices that are generated. In this zone, we refined the mesh so that the element size is sufficiently small compared to the wave length of the generated internal waves. The edge length in this refined zone is 2 km. This mesh is made up of 23,562 triangles extruded into 25 σ layers. Basically, it can also be viewed as a trial and error procedure where a preliminary calculation allows us to a fine tuning of the mesh refinement. Obviously, the automatic adaptive refinement procedure will be a more general approach.

The only numerical parameter that has to be selected in the three-dimensional baroclinic model is the jump penalty parameter γ in the Lax–Friedrichs solver. For this problem, we select $\gamma = 4$ m s^{-1}. This parameter must be an upper bound of the phase speed of the fastest wave. From the linear theory, we know that with the prescribed stratification, the maximum phase speed of an internal wave is about $c = 1$ m s^{-1}, so that the fastest three-dimensional phenomenon propagates at $c + U \approx 1.5$ m s^{-1}. For discontinuous linear elements combined with the second-order explicit Runge–Kutta time-stepper (Chevaugeon et al. 2007) used in this simulation, the relevant CFL conditions leads us to select a time step of 30 s.

The two-dimensional mesh on the sea bottom and the time evolution for the isosurfaces of the density perturbations are shown in Fig. 6. The density

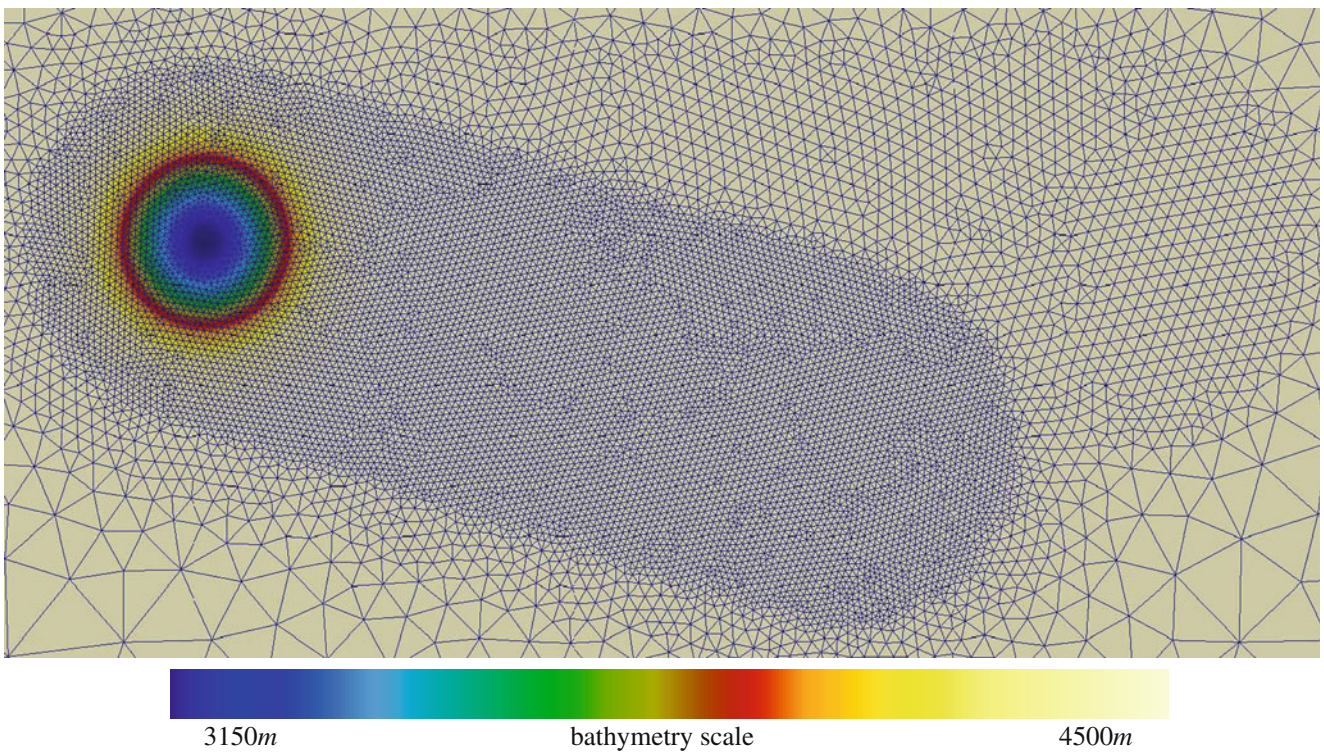

Fig. 5 Close-up view on the mesh and the bathymetry around the seamount. The mesh is refined in the lee of the seamount

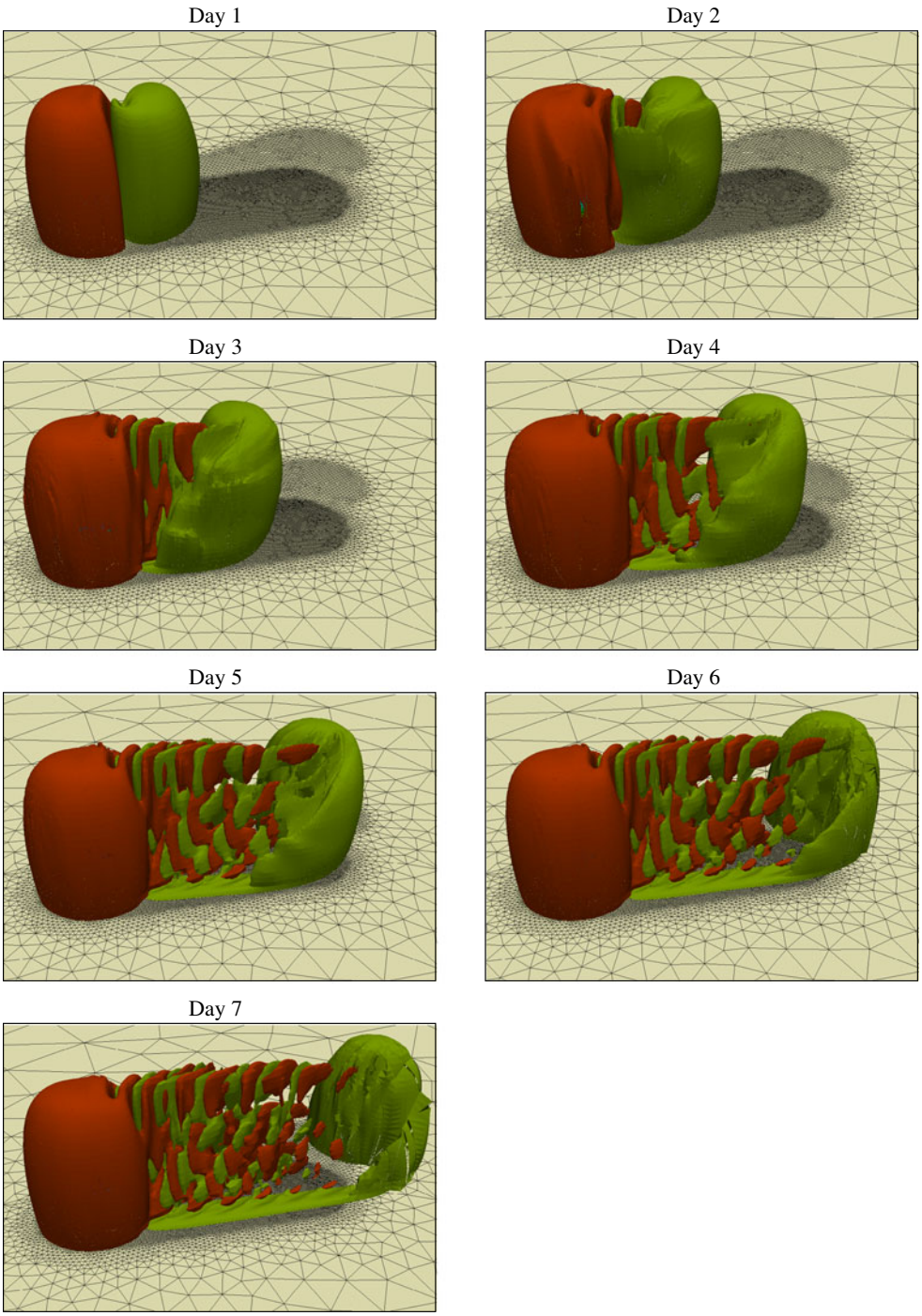

Fig. 6 Time evolution of the isosurfaces of the density perturbation. Isovalues for the density perturbation of -0.001 kg m^{-3} are in *green*. Isovalues for the density perturbation of 0.001 kg m^{-3} are in *orange*. The two-dimensional mesh is given on the sea bottom

perturbation is defined as the density deviation field ρ' from which the initial density deviation has been removed. The density perturbation can be considered as a good diagnostics: As the flow is dominated by geostrophy, it is directly an image of the vorticity induced by the fact that the flow is impulsively started. The free surface is raised upstream of the seamount and lowered downstream, leading thanks to geostrophic adjustment to an upstream clockwise vortex, and a downstream counterclockwise vortex, which are both visible in Fig. 6. The flow is strong enough to directly shed the counterclockwise eddy away from the seamount. The mean flow is deviated rightward downstream of the seamount. The clockwise eddy is trapped over the seamount. In the zone between the two eddies, internal waves are generated. Rather than being radiated away from the seamount, they are trapped in the lee of the seamount.

In Fig. 7, we also see that these waves have a particular structure. In the plots of the time evolution of the density perturbation at a 400-m depth, we observe that the waves are generated by the shedded eddy, propagate westward, and stack in the lee of the seamount. Again, the upstream clockwise vortex and the downstream counterclockwise vortex are both clearly visible in Fig. 7.

At the seventh day, two well-separated modes in the density perturbation field are clearly visible as shown in Fig. 8. In the left side of the lee, looking downstream, an internal mode with two extrema appears. In the

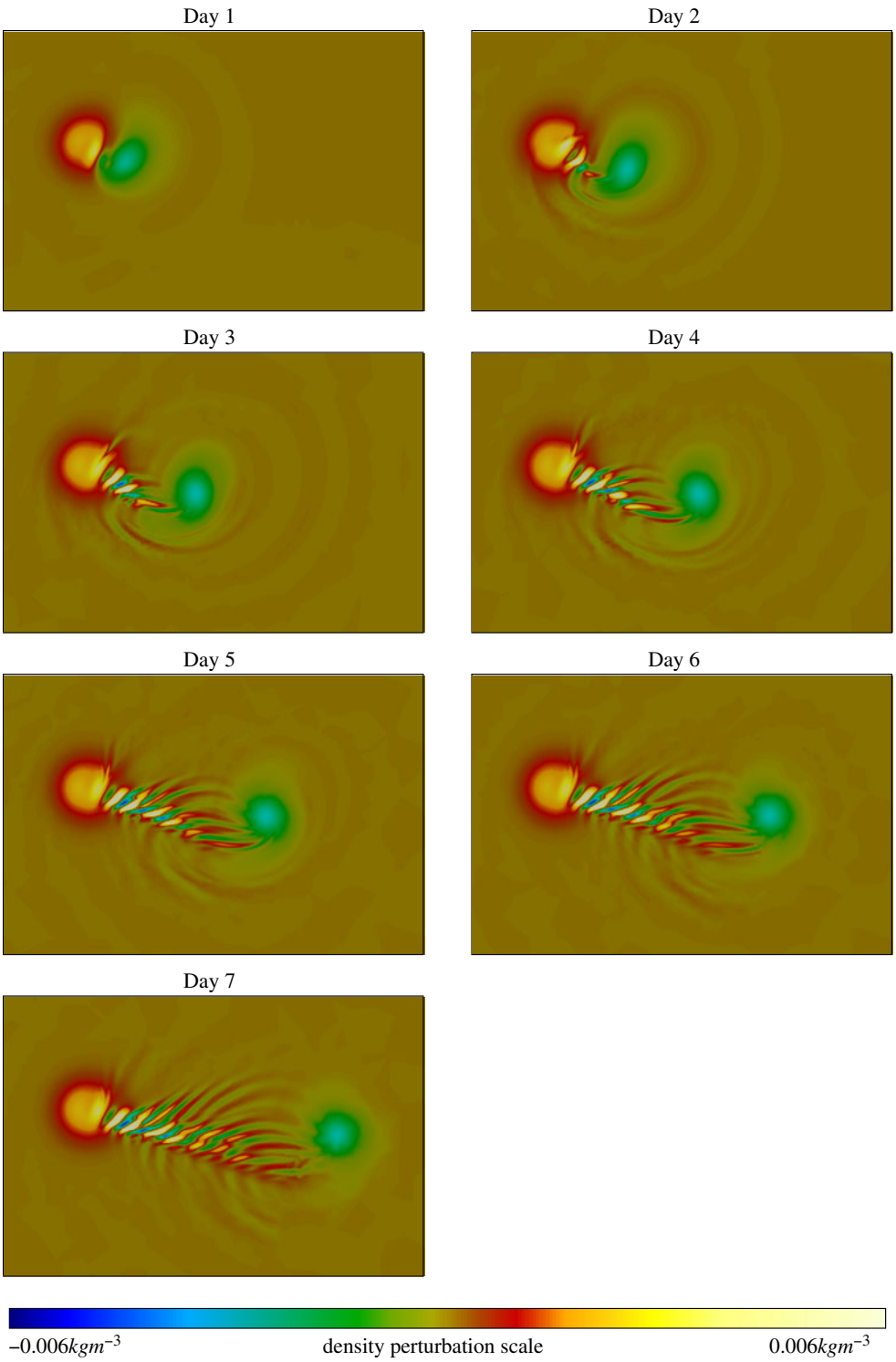

Fig. 7 Time evolution of the density perturbation field at a 400-m depth. The internal waves lying between the two vortices are clearly visible

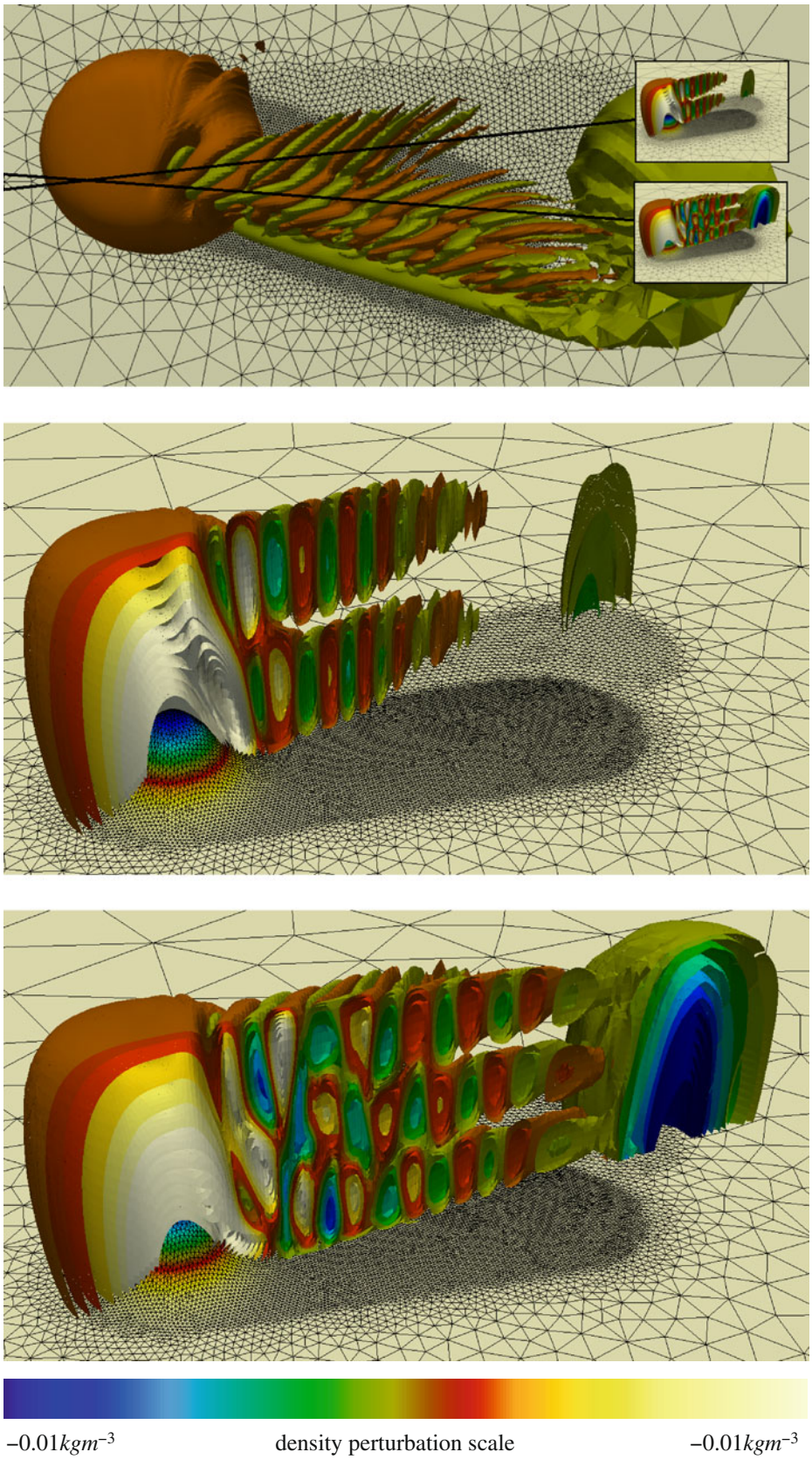

Fig. 8 Illustration of the two well-separated modes at day 7. The *top panel* shows a view of the isocontours of the density perturbation. Two cuts are defined and are given on the two *lower panels*. Isocontours values range from -0.02 to 0.02 kg m^{-3} with an interval of 0.002. Isovalues of -0.005 and 0.005 are added, while the zero isovalue is removed

right side of the lee, an internal wave mode with three extrema appears. These numerical results can be compared with the theoretical analysis of the internal waves given in Lecture 17 of the book of Pedlosky (2003). The theory of the internal waves in a flat bottom ocean with uniform stratification implies the occurrence of eigenmodes. The vertical wave number is:

$$m = \frac{i\pi}{H}, \tag{33}$$

with the integer i being the number of extrema of the vertical wave profile. In a linear analysis, these modes are independent and each of these modes behaves as a shallow water layer, with an equivalent depth defined as:

$$h_i = \frac{N^2 H^2}{i^2 \pi^2 g}. \tag{34}$$

As for the usual shallow water equations, Kelvin waves along coastlines, Poincaré waves, and Rossby waves can be observed. However, as we are interested in a flow over a relatively short period of time on a aquaplanet without coastlines, only the Poincaré waves are relevant here. The phase speed of the Poincaré waves is given by:

$$c_i = \sqrt{g h_i + \frac{f^2}{k^2}} = \sqrt{\frac{N^2 H^2}{i^2 \pi^2} + \frac{f^2}{k^2}}, \tag{35}$$

where k is the horizontal wavenumber. For the flow to support standing waves, the propagation speed of internal waves in the opposite direction of the mean flow must be equal to the mean flow speed:

$$c_i = U \cos(\alpha), \tag{36}$$

where α is the angle between the mean flow speed and the direction of the wave propagation. The direction of wave propagation is normal to the wave crests.

In Fig. 9, we represent the wave modes and propagation speeds. The mean velocity vector is not aligned

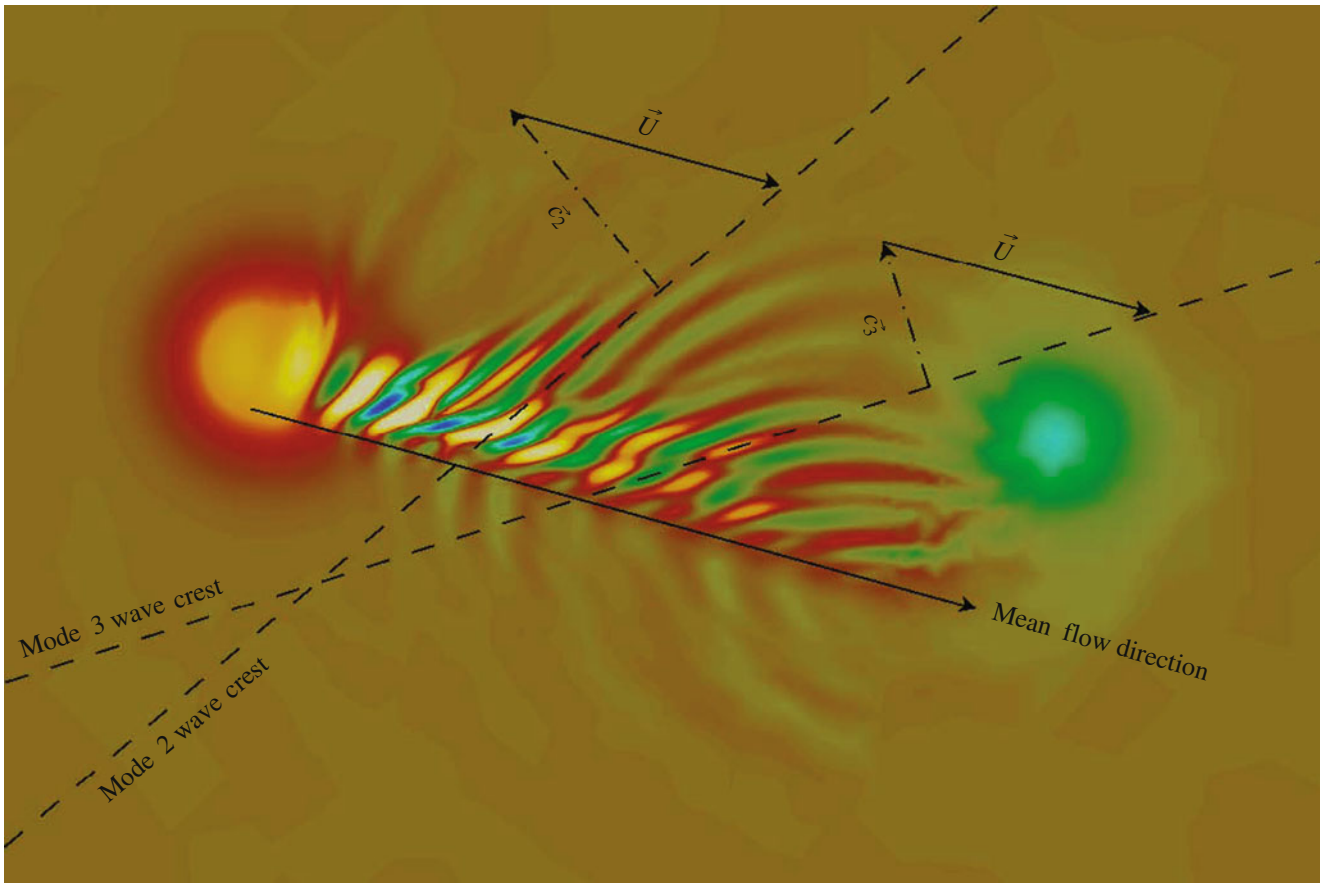

Fig. 9 Sketch of the wave modes and propagation speeds. $\vec{U}$ denotes the mean velocity vector. It is not horizontal because the mean flow is deviated to the right by the seamount. $\vec{c}_2$ and $\vec{c}_3$ denote the phase speed vector for mode 2 and mode 3 internal waves, respectively. The amplitude of the phase speed vectors is taken for the waves to be stationary

because the mean flow is deviated to the right by the seamount. The phase speed vector for mode 2 and mode 3 internal waves are also given, with the amplitude of the phase speed vectors selected for stationary waves. From such a picture, the angle α between the mean flow speed and the propagation of the wave propagation can be deduced. For mode 2 waves, the observed angle between the mean flow velocity and the wave propagation direction is about $\alpha \approx 35°$ in Fig. 9. Taking advantage of the theoretical linear analysis, we use Eqs. 35 and 36 to estimate the horizontal wavenumber of the waves from the observed angle α and the vertical wavenumber (i.e., the number of modes)

$$c = 0.42 \text{ m s}^{-1},$$

$$k = 1.03\ 10^{-3} \text{ m}^{-1}.$$

This speed is close to the threshold value 0.410 m s^{-1}, under which the flow is subcritical for mode 2, and stationary lee waves cannot exist. From the calculated horizontal wavenumber, the wave length is estimated to be $2\pi/k = 6.1$ km and should correspond to the crest to crest distance. However, in Fig. 9, we observe a crest to crest distance of 18.75 km for mode 2. Such a discrepancy between those two results can be explained by the fact that the phase speed is very close to the threshold value for which no wave can exist. Therefore, the phase speed does not depend so strongly on the wavenumber. Moreover, if we perform a vertical slice in the density deviation field ρ', we observe that the amplitude of the waves is significant compared to the total depth in Fig. 10. Therefore, considering the linear regime for the theoretical computation of the dispersion relation is probably a bad assumption and can induce significant errors.

7 Conclusions

The spatial discretization of a three-dimensional baroclinic free-surface marine model is introduced. This model relies on a discontinuous Galerkin method with a mesh of prisms extruded in several layers from an unstructured two-dimensional mesh of triangles. As the prisms are vertically aligned, the calculation of the vertical velocity and the baroclinic pressure gradient can be implemented in an efficient and accurate way. All discrete fields are defined in discontinuous finite element spaces, to take advantage of the good numerical properties of the discontinuous Galerkin methods for advection dominated problems and for wave problems. To be able to use the Riemann solver of the shallow water equations, the discretization of the three-dimensional horizontal momentum and the continuity equations are defined in such a way that their discrete integration along the vertical axis provides a stable $P_1^{DG} - P_1^{DG}$ formulation of the shallow water equations. Therefore, we can stabilize the discrete equations by using the exact Riemann solver of the linear shallow water for the gravity waves. For internal waves, an additional stabilizing term is derived from a Lax–Friedrichs solver. In the baroclinic dynamics, the vertical velocity acts as a source term, while the role of

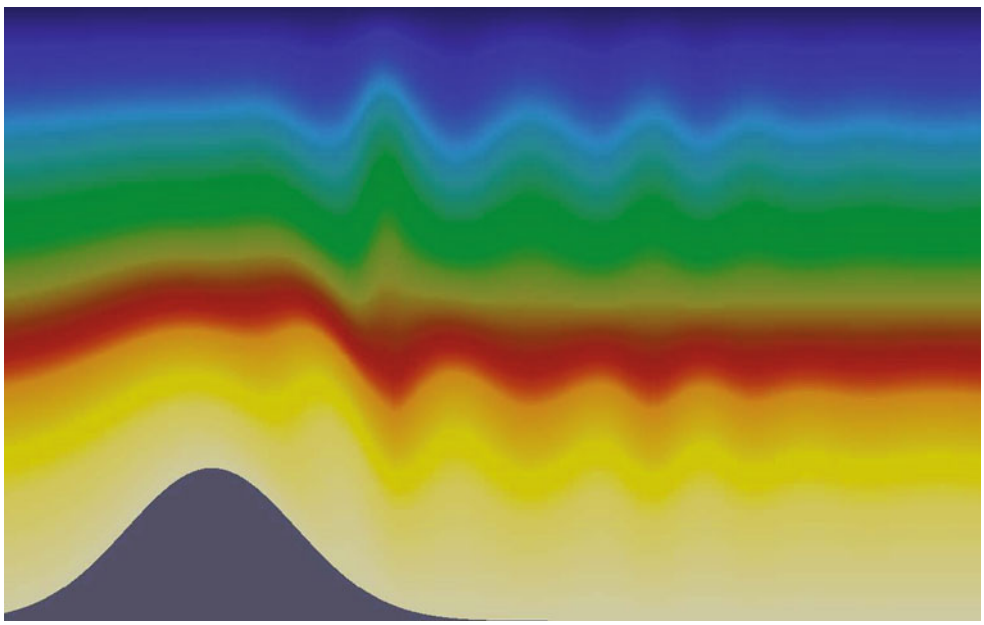

Fig. 10 Density field at day 7, along the plane for mode 2 defined in Fig. 8

the approximate Riemann solver is to penalize inter-element jumps to recover optimal accuracy. Consistency is ensured. The model is able to advect exactly a tracer with a constant concentration, meaning that the discrete transport term is compatible with the continuity equation.

A key advantage of discontinuous Galerkin finite elements is their ability to naturally handle higher order discretizations. The discrete formulation with the same approximate Riemann solvers can be used for high-order shape functions. As an illustrative example, we simulate the same problem as in Section 6 with second-order shape functions and a mesh two times coarser involving 12 layers. The triangles are twice larger than in the previous calculation. A comparison of the density perturbation field after 2 days is shown in Fig. 11. We observe the same behavior, as both simulations are performed on a sufficiently fine mesh. However, some small oscillations are induced by the subparametric representation of the bottom topography. They appear above the seamount for the quadratic shape functions computation. Indeed, the bathymetry is still represented using piecewise linear polynomials, while the fields are represented using piecewise quadratic polynomials. But we think that such a discretization of a three-dimensional baroclinic finite element marine model is an effective way for higher-order elements, paving the way for high-order ocean models based on discontinuous Galerkin methods. Finally, even if this contribution can be viewed as important effort to the development of unstructured mesh ocean models, our model has only a dynamic core and is not ready for realistic applications in this sense.

In conclusion, we use a three-dimensional finite element baroclinic free-surface model to represent accurately the complex structure of the internal waves in the lee of an isolated seamount, using either linear or quadratic shape functions. The model does not yet handle internally supercritical flows that occur for instance when internal waves break or in steed gravity currents. Including a limiting strategy to handle shockwaves would be the next required step. Finally, the second key ingredient for an efficient three-dimensional marine model is the definition of a good time integration procedure. This will be the topic of the second part of this contribution.

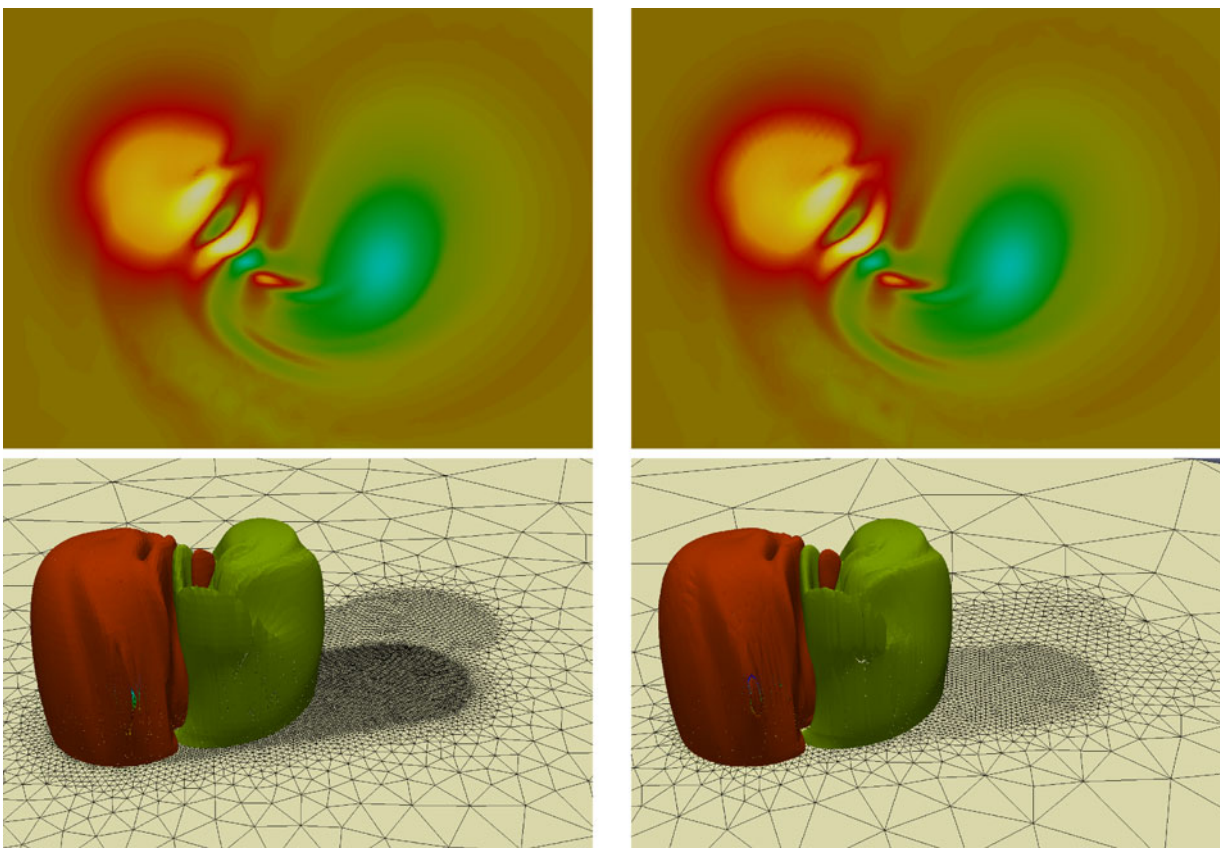

Fig. 11 Density perturbation isocontours after 2 days. Linear (*left*) and quadratic (*right*) discretizations for the 30% height case

Acknowledgements Sébastien Blaise and Jonathan Lambrechts are research fellows with the Belgian Fund for Research in Industry and Agriculture (FRIA). Richard Comblen is a research fellow with the Belgian National Fund for Scientific Research (FNRS). Eric Deleersnijder is a research associate with the Belgian National Fund for Scientific Research (FNRS). The research was conducted within the framework of the Interuniversity Attraction Pole TIMOTHY (IAP 6.13), funded by the Belgian Science Policy (BELSPO), and the programme ARC 04/09-316, funded by the Communauté Française de Belgique.

References

Adcroft A, Hill C, Marshall J (1997) Representation of topography by shaved cells in a height coordinate ocean model. Mon Weather Rev 95:2293–2315

Aizinger V, Dawson C (2002) A discontinuous Galerkin method for two-dimensional flow and transport in shallow water. Adv Water Resour 25:67–84

Aizinger V, Dawson C (2007) The local discontinuous Galerkin method for three-dimensional shallow water flow. Comput Methods Appl Mech Eng 196:734–746

Arakawa A, Lamb VR (1977) Computational design of the basic dynamical processes of the UCLA general circulation model. In: Chang J (ed) Methods in computational physics. Academic, New York, pp 173–265

Arnold DN, Brezzi F, Cockburn B, Marini LD (2002) Unified analysis of discontinuous Galerkin methods for elliptic problems. SIAM J Numer Anal 39:1749–1779

Bernard P-E, Remacle J-F, Legat V, (2008a) Boundary discretization for high order discontinuous Galerkin computations of tidal flows around shallow water islands. Int J Numer Methods Fluids 59:535–557

Bernard P-E, Remacle J-F, Legat V (2008b) Modal analysis on unstructured meshes of the dispersion properties of the $P_1^{NC} - P_1$ pair. Ocean Model 28:2–11

Blaise S, de Brye B, de Brauwere A, Deleersnijder E, Delhez EJM, Comblen R (2010) Capturing the residence time boundary layer—application to the Scheldt Estuary. Ocean Dyn 60:535–554

Bryan K (1969) A numerical method for the study of the circulation of the world ocean. J Comput Phys 4:347–376

Casulli V, Walters RA (2000) An unstructured grid, three-dimensional model based on the shallow water equations. Int J Numer Methods Fluids 32:331–348

Chapman DC, Haidvogel DB (1992) Formation of Taylor caps over a tall isolated seamount in a stratified ocean. Geophys Astrophys Fluid Dynam 64:31–65

Chapman DC, Haidvogel DB (1993) Generation of internal lee waves trapped over a tall isolated seamount. Geophys Astrophys Fluid Dynam 69:33–54

Chen C, Liu H, Beardsley RC (2003) An unstructured grid, finite-volume, three-dimensional, primitive equations ocean model: applications to coastal ocean and estuaries. J Atmos Ocean Technol 20:159–186

Chevaugeon N, Hillewaert K, Gallez X, Ploumhans P, Remacle J-F (2007) Optimal numerical parametrization of discontinuous Galerkin method applied to wave propagation problems. J Comput Phys 223:188–207

Cockburn B, Shu C (1998) The local discontinuous Galerkin finite element method for time-dependent convection-diffusion systems. SIAM J Numer Anal 35:2440–2463

Comblen R, Blaise S, Legat V, Remacle J-F, Deleersnijder E, Lambrechts J (2010a) A discontinuous finite element baroclinic marine model on unstructured prismatic meshes. Part II: implicit/explicit time discretization. Ocean Dyn. doi: 10.1007/s10236-010-0357-4

Comblen R, Lambrechts J, Remacle J-F, Legat V (2010b) Practical evaluation of five part-discontinuous finite element pairs for the non-conservative shallow water equations. Int J Numer Methods Fluids 73:701–724

Comblen R, Legrand S, Deleersnijder E, Legat V (2009) A finite element method for solving the shallow water equations on the sphere. Ocean Model 28:12–23

Cotter CJ, Ham DA, Pain CC (2009a) A mixed discontinuous/continuous finite element pair for shallow-water ocean modelling. Ocean Model 26:86–90

Cotter CJ, Ham DA, Pain CC, Sebastian Reich S (2009b) LBB stability of a mixed Galerkin finite element pair for fluid flow simulations. J Comput Phys 228:336–348

Dawson C, Aizinger V (2005) A discontinuous Galerkin method for three-dimensional shallow water equations. J Sci Comput 22–23:245–267

Deleersnijder E, Beckers J-M (1992) On the use of the sigma-coordinate in regions of large bathymetric variations. J Mar Syst 3:381–390

Deleersnijder E, Lermusiaux PFJ (2008) Multi-scale modeling; nested-grid and unstructured-mesh approaches. Ocean Dyn (special issue) 58:335–498

Ford R, Pain CC, Piggott M, Goddard A, de Oliveira CR, Umpleby A (2004a) A non-hydrostatic finite element model for three-dimensional stratified oceanic flows, part I: model formulation. Mon Weather Rev 132:2816–2831

Ford R, Pain CC, Piggott M, Goddard A, de Oliveira CR, Umpleby A (2004b) A non-hydrostatic finite element model for three-dimensional stratified oceanic flows, part II: model validation. Mon Weather Rev 132:2832–2844

Fringer OB, Gerritsen M, Street RL (2006) An unstructured-grid, finite-volume, nonhydrostatic, parallel coastal ocean simulator. Ocean Model 14:139–173

Giraldo FX (2006) High-order triangle-based discontinuous Galerkin methods for hyperbolic equations on a rotating sphere. J Comput Phys 214:447–465

Griffies SM, Adcroft AJ, Banks H, Böning CW, Chassignet EP, Danabasoglu G, Danilov S, Deleersnijder E, Drange H, England M, Fox-Kemper B, Gerdes R, Gnanadesikan A, Greatbach RJ, Hallberg RW, Hanert E, Harrison MJ, Legg SA, Little CM, Madec G, Marsland S, Nikurashin M, Piran A, Simmons HL, Schröter J, Treguier BLSA-M, Toggweiler JR, Tsujino H, Vallis GK, White L (2009) Problems and prospects in large-scale ocean circulation models. OceanObs'09. http://www.oceanobs09.net/blog/?p=88

Griffies SM, Böning C, Bryan FO, Chassignet EP, Gerdes R, Hasumi H, Hirst A, Treguier A-M, Webb D (2000) Developments in ocean climate modeling. Ocean Model 2: 123–192

Ham DA, Pietrzak J, Stelling GS (2005) A scalable unstructured grid 3-dimensional finite volume mode for the shallow water equations. Ocean Model 10:153–169

Hanert E, Roux DYL, Legat V, Deleersnijder E (2005) An efficient Eulerian finite element method for the shallow water equations. Ocean Model 10:115–136

Haney RL (1991) On the pressure gradient force over steep topography in sigma coordinate ocean models. J Phys Oceanogr 21:610–619

Hua B-L, Thomasset F (1984) A noise-free finite element scheme for the two-layer shallow water equations. Tellus 36A:157–165

Huppert HE, Bryan K (1976) Topographically generated eddies. Deep-Sea Res 23:655–679

Kubatko EJ, Westerink JJ, Dawson C (2006) *hp* discontinuous Galerkin methods for advection dominated problems in shallow water flows. Comput Meth Appl Mech Eng 196:437–451

Labeur RJ, Wells GN (2009) Interface stabilised finite element method for moving domains and free surface flows. Comput Meth Appl Mech Eng 198(5–8):615–630

Le Roux DY, Rostand V, Pouliot B (2007) Analysis of numerically induced oscillations in 2D finite-element shallow-water models part I: inertia–gravity waves. SIAM J Sci Comput 29:331–360

Le Roux DY, Rostand V, Pouliot B (2008) Analysis of numerically induced oscillations in 2D finite-element shallow-water models part II: free planetary waves. SIAM J Sci Comput 30:1971–1991

Le Roux DY, Staniforth A, Lin CA (1998) Finite elements for shallow-water equation ocean model. Mon Weather Rev 126:1931–1951

LeVeque RJ (2002) Finite volume methods for hyperbolic problems. Cambridge texts in applied mathematics. Cambridge University Press, Cambridge

Nair RD, Thomas SJ, Loft RD (2005) A discontinuous Galerkin global shallow water model. Mon Weather Rev 133:876–888

Pedlosky J (2003) Waves in the ocean and atmosphere. Springer, New York

Piggott MD, Gorman GJ, Pain CC, Allison PA, Candy AS, Martin BT, Wells MR (2008) A new computational framework for multi-scale ocean modelling basid on adapting unstructured meshes. Int J Numer Methods Fluids 56:1003–1015

Redi MH (1982) Oceanic isopycnal mixing by coordinate rotation. J Phys Oceanogr 12:1154–1158

Riviere B (2008) Discontinuous Galerkin methods for solving elliptic and parabolic equations: theory and implementation. Frontiers in mathematics, vol 35. SIAM, Philadelphia

Rostand V, Le Roux D (2008) Raviart–Thomas and Brezzi–Douglas–Marini finite element approximations of the shallow-water equations. Int J Numer Methods Fluids 57:951–976

Rostand V, Le Roux D, Carey G (2008) Kernel analysis of the discretized finite difference and finite element shallow-water models. SIAM J Sci Comput 31:531–556

Shahbazi K (2005) An explicit expression for the penalty parameter of the interior penalty method. J Comput Phys 205:401–407

Stuhne GR, Peltier WR (2006) A robust unstructured grid discretization for 3-dimensional hydrostatic flows in spherical geometry: a new numerical structure for ocean general circulation modeling. J Comput Phys 213:704–729

Timmermann R, Danilov S, Schröter J, Böning C, Sidorenko D, Rollenhagen K (2009) Ocean circulation and sea ice distribution in a finite element global sea ice–ocean model. Ocean Model 27(3–4):114–129

Toro E (1997) Riemann solvers and numerical methods for fluid dynamics, a practical introduction. Springer, Berlin

Wang Q, Danilov S, Schröter J (2008a) Comparison of overflow simulations on different vertical grids using the finite element ocean circulation model. Ocean Model 20:313–335

Wang Q, Danilov S, Schröter J (2008b) Finite element ocean circulation model based on triangular prismatic elements, with application in studying the effect of topography representation. J Geophys Res 113

White L, Legat V, Deleersnijder E (2008) Tracer conservation for three-dimensional, finite-element, free-surface, ocean modeling on moving prismatic meshes. Mon Weather Rev 136:420–442

Williamson DL, Drake JB, Hack JJ, Jakob R, Swarztrauber PN (1992) A standard test set for numerical approximations to the shallow water equations in spherical geometry. J Comput Phys 102:211–224

Ocean Dynamics (2010) 60:1395–1414
DOI 10.1007/s10236-010-0357-4

A discontinuous finite element baroclinic marine model on unstructured prismatic meshes

Part II: implicit/explicit time discretization

Richard Comblen · Sébastien Blaise · Vincent Legat · Jean-François Remacle · Eric Deleersnijder · Jonathan Lambrechts

Received: 6 July 2010 / Accepted: 27 October 2010 / Published online: 26 November 2010

Abstract We describe the time discretization of a three-dimensional baroclinic finite element model for the hydrostatic Boussinesq equations based upon a discontinuous Galerkin finite element method. On one hand, the time marching algorithm is based on an efficient mode splitting. To ensure compatibility between the barotropic and baroclinic modes in the splitting algorithm, we introduce Lagrange multipliers in the discrete formulation. On the other hand, the use of implicit–explicit Runge–Kutta methods enables us to treat stiff linear operators implicitly, while the rest of the nonlinear dynamics is treated explicitly. By way of illustration, the time evolution of the flow over a tall isolated seamount on the sphere is simulated. The seamount height is 90% of the mean sea depth. Vortex shedding and Taylor caps are observed. The simulation compares well with results published by other authors.

Keywords Baroclinic marine model · Discontinuous Galerkin methods · Time discretization

Responsible Editor: Pierre Lermusiaux

R. Comblen · S. Blaise · V. Legat (✉) · J.-F. Remacle · E. Deleersnijder · J. Lambrechts
Institute of Mechanics, Materials and Civil Engineering (IMMC), Université catholique de Louvain, 4 Avenue G. Lemaître, 1348 Louvain-la-Neuve, Belgium
e-mail: vincent.legat@uclouvain.be

R. Comblen · S. Blaise · V. Legat · J.-F. Remacle · E. Deleersnijder · J. Lambrechts
Georges Lemaître Centre for Earth and Climate Research (TECLIM), Université catholique de Louvain, 4 Avenue G. Lemaître, 1348 Louvain-la-Neuve, Belgium

R. Comblen · S. Blaise · V. Legat · J.-F. Remacle · E. Deleersnijder · J. Lambrechts
Centre for Systems Engineering and Applied Mechanics (CESAME), Université catholique de Louvain, 4 Avenue G. Lemaître, 1348 Louvain-la-Neuve, Belgium

E. Deleersnijder
Earth and Life Institute (ELI), Euler, Université catholique de Louvain, 4 Avenue G. Lemaître, 1348 Louvain-la-Neuve, Belgium

Present Address:
S. Blaise
Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, 1850 Table Mesa Drive, Boulder, CO 80305, USA

1 Introduction

The spatially discretized ocean system is a dynamical system with a very large number of unknowns. It is the case for many computational fluid dynamics problems. However, oceanic problems typically consider very large time scales compared to those of the rapidly varying dynamics of local flows. Indeed, for climate modeling, centuries are considered, while it only takes a few hours for a surface gravity wave to propagate around the world.

Early models used the rigid-lid approximation, where the sea surface is assumed to be a rigid horizontal, impermeable boundary. This approximation filters out the fast surface gravity waves. The two-dimensional mean problem is solved either using a streamfunction formulation (Bryan 1969) or a surface pressure formulation (Dukowicz et al. 1993). The computation of the surface pressure or streamfunction leads to elliptic two-dimensional problems, the rest of the dynamics being computed with explicit methods such as leap-frog

with filtering (Griffies et al. 2000). Explicit methods for hyperbolic problems are subject to the Courant–Friedrichs–Lewy stability condition: The time step must be sufficiently small. The information in a cell only influences its direct neighbors, or the time step scales as the ratio of the grid size to the fastest wave speed. For rigid-lid models, the fastest phenomena are internal gravity waves. When free surface is taken into account, much faster phenomena occur: Surface gravity waves propagate roughly two orders of magnitude faster than internal gravity waves. Reducing the baroclinic time step by a factor of a hundred was not an option for the first free-surface models (Blumberg and Mellor 1987; Killworth et al. 1991), and specific algorithms were designed to overcome this problem.

The purely explicit mode-splitting procedure, used in a large number of models (Kubatko et al. 2008), consists in integrating the two-dimensional barotropic equations with many explicit time steps while the three-dimensional baroclinic equations are solved with a single, much larger time step. For long-term computations, the numerical model must be consistent by being able to advect a constant concentration of a given tracer exactly, up to machine accuracy. To achieve this so-called consistency requirement, the advection term of the tracer equation must degenerate to the continuity equation when a constant tracer concentration is considered (White et al. 2008). Compatibility between the two-dimensional and the three-dimensional approximations of the velocity field is also a mandatory requirement to ensure consistency (Deleersnijder 1993). Therefore, the three-dimensional velocities are a posteriori corrected so that their averages match the velocities of the barotropic mode (Blumberg and Mellor 1987), i.e., to obtain compatibility. Averaging in time the two-dimensional quantities in the three-dimensional dynamics is usually needed to ensure stability (Griffies et al. 2000; Higdon and de Szoeke 1997; Hallberg 1997).

To get rid of this a posteriori correction step of the three-dimensional velocities, the two-dimensional barotropic mode can be time-stepped implicitly, using the same time step as the three-dimensional baroclinic mode (Dukowicz and Smith 1994). Therefore, no correction step is needed, when all three-dimensional terms are advanced explicitly in time. The model will be both compatible and consistent, but the time step will be small. However, some terms in the three-dimensional momentum equation can be advanced implicitly in time, if those implicit terms have only a small influence on the two-dimensional barotropic mode (Wang 2007).

An efficient strategy to discretize implicitly the free-surface equation is to solve a smaller system corresponding to the Schur complement of the system. If the linear discrete system corresponding to the horizontal three-dimensional momentum and free-surface equations reads:

$$\begin{bmatrix} M_{\boldsymbol{u}} & G \\ D & M_\eta \end{bmatrix} \begin{bmatrix} X_{\boldsymbol{u}} \\ X_\eta \end{bmatrix} = \begin{bmatrix} F_{\boldsymbol{u}} \\ F_\eta \end{bmatrix},$$

where $M_{\boldsymbol{u}}$, M_η and $X_{\boldsymbol{u}}$, X_η are the mass matrices and vectors of degrees of freedom for velocities and elevation, respectively, G and D correspond to the elevation gradient term of the horizontal momentum equation and the velocity divergence term of the free-surface elevation equation, while $F_{\boldsymbol{u}}$ and F_η are their right-hand sides.

An equivalent smaller system for the elevation is obtained substituting $X_{\boldsymbol{u}}$ in the last line of the system:

$$\left[M_\eta - D M_{\boldsymbol{u}}^{-1} G\right] X_\eta = F_\eta - D M_{\boldsymbol{u}}^{-1} F_{\boldsymbol{u}}.$$

Such a methodology was used not only by Dukowicz and Smith (1994) as well as by Marshall et al. (1997) for global-scale models but also by Giraldo et al. (2003) for shallow water problems on the sphere. Discontinuous finite elements are ideally suited for such a procedure, as the mass matrix is block diagonal. Such an implicit free-surface procedure is compatible and consistent if the three-dimensional mode is time-stepped explicitly. Unfortunately, vertical diffusion deduced from a turbulence scheme or used as a convective adjustment algorithm is often large enough for the corresponding constraint on the time step to be much more severe than the one deduced from internal gravity waves. Vertical diffusion and advection terms must often be treated implicitly.

The Finite Element Ocean Model (FEOM) uses a similar approach with an implicit vertical diffusion (Wang 2007; Wang et al. 2008). The momentum equation is split in two steps, and an intermediate velocity is introduced to perform the implicit calculation of the elevation. Implicit vertical viscosity is neglected in the correction step, meaning that this term is computed using this intermediate velocity rather than the final corrected value. This is needed to derive the equation associated with the elevation. Such a method of substitution is similar to the Schur complement approach used by Dukowicz and Smith (1994), but the substitution is performed in the continuous space rather than at the discrete level. Working within a continuous framework, the inverse of the mass matrix $M_{\boldsymbol{u}}^{-1}$ disappears in what corresponds to the Schur complement. Therefore, the discrete operator is not the same, and the two-dimensional discrete system for the elevation reads:

$$\left[M_\eta - \Delta_t L\right] X_\eta = F_\eta - \Delta_t D F_{\boldsymbol{u}},$$

where L is the discrete Laplacian matrix and Δ_t is the time step. The operator $\Delta_t L$ is used rather than $DM_{\boldsymbol{u}}^{-1}G$ and can be viewed as an approximation or an alternative choice. The same remark applies for $\Delta_t DF_{\boldsymbol{u}}$ that acts as a substitute of $DM_{\boldsymbol{u}}^{-1}F_{\boldsymbol{u}}$.

In this paper, we present an implicit mode splitting procedure used for a marine model, called Second-generation Louvain-la-Neuve Ice–ocean Model (SLIM; [1]) that should be able to deal with problems ranging from local and regional scales to global scales. In a first step of the time stepper, the new elevation field is implicitly computed, and afterward, we use this value and the corresponding two-dimensional velocities in the baroclinic mode to compute the three-dimensional velocity. The main contribution of the time marching procedure of SLIM is that a correction term is embedded in the three-dimensional momentum equation to ensure compatibility between two- and three-dimensional velocities. Identity between depth-averaged three-dimensional velocity and two-dimensional velocity is enforced using Lagrange multipliers in the three-dimensional system. It provides an implicit accurate coupling between two- and three-dimensional modes.

To illustrate the accuracy of this time discretization, we consider the simulation of Taylor caps in the wake of a tall seamount. We define the parameters of the simulation to compare the transient dynamics with results published by Chapman and Haidvogel (1992) and Ford et al. (2004). The main reason to analyze this problem is the fact that subgrid-scale parametrization is not required to generate complex baroclinic phenomena (Chapman and Haidvogel 1992).

In a companion paper (Blaise et al. 2010), we introduced the detailed description of the space discretization. Equal-order discontinuous interpolations for the elevation and velocity fields are used. The discontinuous Galerkin method is selected in order to accurately simulate the advection dominated processes. The model operates on prismatic meshes, obtained by extruding vertically a triangular surface grid. It relies on approximate Riemann solvers based on the wave dynamics of the system. The consistency, the accuracy, and the stability of the spatial discretization were analyzed. Herein, the time integration procedure will be provided. Section 2 describes the partial differential equations considered. Section 3 defines a new time-splitting procedure with compatible discrete barotropic 2D and the baroclinic 3D problems. Implicit–explicit (IMEX) Runge–Kutta (RK) time integrators used in the three-dimensional baroclinic marine model are explained in Section 5. A first validation of the dynamics of the model is given in Section 6. Revisiting a modified version of the flow over a tall seamount described in Blaise et al. (2010), we analyze the dynamics behavior and compare our results with previous publications. Concluding remarks are given in Section 7.

[1] http://www.climate.be/slim

2 Governing equations

Using material parameters and notations defined in Table 1, the set of partial differential equations of the three-dimensional baroclinic free-surface model reads:

$$\frac{\partial \boldsymbol{u}}{\partial t} + \nabla_h \cdot (\boldsymbol{u}\boldsymbol{u}) + \frac{\partial w\boldsymbol{u}}{\partial z} + f\mathbf{e}_z \wedge \boldsymbol{u} + \frac{\mathbf{p}}{\rho_0} + g\nabla_h \eta = \nabla_h \cdot (\nu_h \nabla_h \boldsymbol{u}) + \frac{\partial}{\partial z}\left(\nu_v \frac{\partial \boldsymbol{u}}{\partial z}\right), \tag{1}$$

Table 1 Notations for the three-dimensional baroclinic free-surface marine model

Coordinates and spatial operators	
x, y	Horizontal coordinates
z	Vertical coordinate, pointing upwards with its origin at the sea surface at rest
∇_h	Horizontal gradient operator
$\mathbf{e}_z$	Upward unit normal
$\wedge$	Cross product symbol
Material parameters and functions	
g	Gravitational acceleration
ρ_0	Reference density
f	Coriolis parameter
h	Depth at rest
ν_h	Horizontal turbulent viscosity parameter
ν_t	Vertical turbulent viscosity parameter
κ_h	Horizontal turbulent diffusivity parameter
κ_t	Vertical turbulent diffusivity parameter
$\boldsymbol{U}$	Two-dimensional horizontal mean velocity vector
Variables	
$\boldsymbol{u}$	Horizontal three-dimensional velocity vector
w	Vertical three-dimensional velocity vector
$\boldsymbol{u}^{\eta}$	Surface horizontal three-dimensional velocity vector
w^{η}	Surface vertical three-dimensional velocity vector
$\boldsymbol{u}^{-h}$	Bottom horizontal three-dimensional velocity vector
w^{-h}	Bottom vertical three-dimensional velocity vector
η	Sea surface elevation
p	Baroclinic pressure
$\boldsymbol{p}$	Baroclinic pressure gradient
c	Three-dimensional tracer, can be S of T
S	Salinity
T	Temperature

$$\frac{\partial \mathbf{p}}{\partial z} = -g\nabla_h \rho'(T, S) \tag{2}$$

$$\nabla_h \cdot \boldsymbol{u} + \frac{\partial w}{\partial z} = 0, \tag{3}$$

$$\frac{\partial \eta}{\partial t} + \nabla_h \cdot \int_{-h}^{\eta} \boldsymbol{u} dz = 0, \tag{4}$$

$$\frac{\partial c}{\partial t} + \nabla_h \cdot (\boldsymbol{u}c) + \frac{\partial wc}{\partial z} = \nabla_h \cdot (\kappa_h \nabla_h c) + \frac{\partial}{\partial z}\left(\kappa_v \frac{\partial c}{\partial z}\right). \tag{5}$$

where the unknown fields are the horizontal velocity $\boldsymbol{u}(x, y, z, t)$, the baroclinic pressure gradient $\boldsymbol{p}(x, y, z, t) = \nabla_h p(x, y, z, t)$, the vertical velocity $w(x, y, z, t)$, the sea surface elevation $\eta(x, y, t)$, and the tracer concentrations $c(x, y, z, t)$ that can be the temperature and/or the salinity.

Now, let us define the two-dimensional depth-averaged horizontal mean velocity:

$$\boldsymbol{U}(x, y, t) = \frac{1}{h(x, y) + \eta(x, y, t)} \int_{-h(x,y)}^{\eta(x,y,t)} \boldsymbol{u}(x, y, z, t)\, dz. \tag{6}$$

and the corresponding two-dimensional depth-averaged barotropic equations:

$$\frac{\partial \boldsymbol{U}}{\partial t} + f\mathbf{e}_z \wedge \boldsymbol{U} + g\nabla\eta = \mathbf{f}_{\mathbf{U}}, \tag{7}$$

$$\frac{\partial \eta}{\partial t} + \nabla \cdot \left[(h + \eta)\boldsymbol{U}\right] = 0, \tag{8}$$

where $\mathbf{f}_{\mathbf{U}}$ includes all the remaining terms resulting from the integration of Eq. 1. As the free-surface evolution only depends on the two-dimensional velocity $\boldsymbol{U}$, a *mode-splitting procedure* is often introduced. The two-dimensional barotropic equations and the three-dimensional baroclinic equations are advanced in time with different schemes and/or steps. As the surface gravity waves propagate roughly two orders of magnitude faster than internal gravity waves, the three-dimensional baroclinic time step could be quite a lot larger than the two-dimensional barotropic time step. Then, it also appears attractive to use an implicit time stepper for the two-dimensional barotropic problem and an explicit time stepper for the three-dimensional baroclinic mode.

The mode-splitting procedure consists in integrating the two-dimensional barotropic equations with many time steps or an implicit scheme while the three-dimensional baroclinic equations are solved with a single step. In this procedure, the three-dimensional velocities must be calculated in such a way that their average matches the velocities of the barotropic mode to obtain compatibility. Two critical conditions must be fulfilled:

- The two-dimensional $\boldsymbol{U}$ and three-dimensional $\boldsymbol{u}$ discrete representations of the velocities must be compatible. In other words, using the discrete version of Eq. 4 or Eq. 8 must produce exactly the same result. It means that $\boldsymbol{U}$ must be recovered by performing at a discrete level the depth average of $\boldsymbol{u}$. The basic idea is that the equivalence properties that exist in the continuous realm must be preserved in the discrete realm. It is the so-called *compatibility condition*.
- The discrete numerical mode must be consistent by being able to advect a constant concentration of a given tracer exactly, up to machine accuracy. To achieve this, the advection term of the tracer equation must degenerate to the continuity equation when a constant tracer concentration is considered (White et al. 2008). As the compatibility of $\boldsymbol{u}$ and $\boldsymbol{U}$ is needed to obtain the compatibility of w and η, ensuring impermeability at the sea surface, the compatibility can be viewed as a mandatory requirement to ensure this *consistency condition*. Obviously, this is not the usual consistency meaning that the discrete formulation converges to the continuous equation as the mesh size goes for zero.

3 Compatible discrete barotropic and baroclinic problems

In the current SLIM model, we use an implicit mode splitting procedure. Firstly, the elevation and two-dimensional velocities are implicitly computed: It is the two-dimensional barotropic problem. Then, we use this elevation and those two-dimensional velocities to obtain the three-dimensional velocities: It is the three-dimensional baroclinic problem.

The major novelty of the implemented approach of the SLIM model is that the correction term is embedded in the three-dimensional momentum equation to ensure compatibility between two-dimensional and three-dimensional velocities. Equality between depth-averaged three-dimensional velocity and two-dimensional velocity is enforced using Lagrange multipliers in the three-dimensional baroclinic problem. The implicit terms in the momentum equation are

computed using a three-dimensional velocity in agreement with the two-dimensional mode. This ensures for instance that, if the Coriolis term is treated semi-implicitly, it influences the dynamics of both modes in the same way.

In order to describe the whole procedure, let us first define the baroclinic and the barotropic problems.

– The two-dimensional barotropic problem consists in finding $(\boldsymbol{U}, \eta)$ such that:

$$\frac{\partial \boldsymbol{U}}{\partial t} + f\mathbf{e}_z \wedge \boldsymbol{U} + g\nabla\eta = \mathbf{f}_{\mathbf{U}}, \tag{9}$$

$$\frac{\partial \eta}{\partial t} + \nabla \cdot \left[(h+\eta)\boldsymbol{U}\right] = 0, \tag{10}$$

– The three-dimensional baroclinic problem consists in finding $(\boldsymbol{u}, \lambda)$ such that:

$$\frac{\partial \boldsymbol{u}}{\partial t} + f\mathbf{e}_z \wedge \boldsymbol{u} + \frac{\partial w\boldsymbol{u}}{\partial z} - \frac{\partial}{\partial z}\left(\nu_v \frac{\partial \boldsymbol{u}}{\partial z}\right) + \lambda, = \mathbf{f}_{\mathbf{u}}, \tag{11}$$

$$\int_{-h}^{\eta} \boldsymbol{u}\, dz = \boldsymbol{U}, \tag{12}$$

where all terms explicitly time-integrated are included in the right-hand side $\mathbf{f}_{\mathbf{U}}$ and $\mathbf{f}_{\mathbf{u}}$, respectively. In the three-dimensional momentum equation, we add a volume force λ. This force will act to ensure compatibility between both velocity representations. Equations 11 and 12 are the usual Euler–Lagrange equations of the saddle-point problem.

Now, we introduce the finite element mesh and the discrete discontinuous approximations of the field variables of the model $(\eta, \boldsymbol{u}, \boldsymbol{U}, \lambda)$ involved in the barotropic and the baroclinic modes. The three-dimensional mesh is made up of prismatic elements, as illustrated in Fig. 1, and is obtained from the extrusion of triangular two-dimensional elements. The vertical length scale is typically much smaller than the horizontal length scale. In other words, the prisms are thin. We choose prismatic elements to obtain a mesh unstructured in the horizontal direction and structured in the vertical direction. The two-dimensional fields η, $\boldsymbol{U}$, and λ are discretized with P_1^{DG} elements onto the two-dimensional mesh of triangles. The three-dimensional fields $(\boldsymbol{u})$ are discretized on the mesh of prisms, and the corresponding shape functions are obtained as the tensorial product of the linear discontinuous triangle P_1^{DG} and the linear one-dimensional element L_1^{DG}.

In the space discretization of the SLIM model (Blaise et al. 2010), the discrete free-surface equation is obtained as the aggregation of the discrete horizontal divergence of the three-dimensional horizontal velocity with a bottom boundary term. A similar approach can

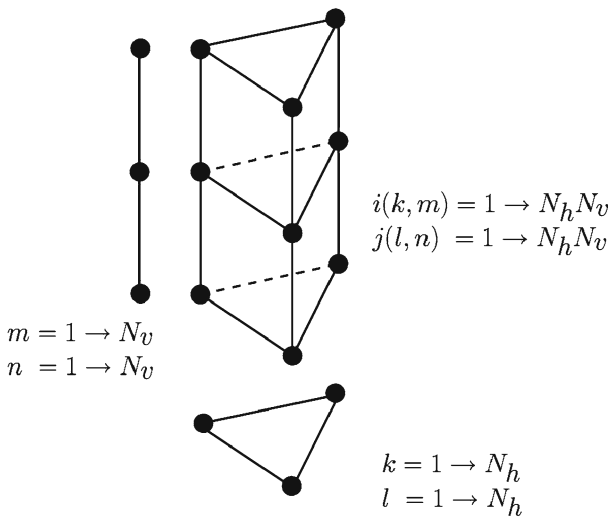

Fig. 1 Summary of the finite element spaces used for each field. Triangular linear elements are noted P_1 while vertical linear elements are noted L_1. Indices indicated on the mesh correspond to the nodal discrete values. The global indices $i(k,m)$ and $j(k,m)$ are a function of the horizontal position and the vertical position

Field		Finite element space
Free surface elevation	η	P_1^{DG}
Lagrang multiplier	λ	P_1^{DG}
Horizontal two-dimensional velocity vector	$\boldsymbol{U}$	P_1^{DG}
Horizontal three-dimensional velocity vector	$\boldsymbol{u}$	$P_1^{DG} \times L_1^{DG}$

be done for the barotropic and baroclinic momentum equations. We aggregate the three-dimensional horizontal momentum equation (Eq. 11), without the Lagrange multipliers λ. This will lead to Eq. 9, which corresponds to the non-conservative form of the momentum equation of the shallow water equations. To exactly obtain this Eq. 9, the test function of the three-dimensional momentum equation is divided by the depth.

Then we add to the linear system corresponding to the horizontal three-dimensional momentum equation for the degrees of freedom of a column of prisms, six lines and columns corresponding to two Lagrange multipliers for each of the three surface nodes (considering linear shape functions). The lines correspond to the compatibility constraint:

$$\sum_{e=1}^{N_e} < \widehat{\lambda} \cdot (\boldsymbol{u} - \boldsymbol{U}) >_e = 0, \tag{13}$$

while the columns correspond to the fictitious force:

$$\sum_{e=1}^{N_e} \left[< \widehat{\boldsymbol{u}} \cdot \frac{\partial \boldsymbol{u}}{\partial t} >_e + \cdots + < \widehat{\boldsymbol{u}} \cdot \boldsymbol{\lambda} >_e \right] = 0, \tag{14}$$

with λ the Lagrange multiplier field and $\widehat{\boldsymbol{\lambda}}$ the corresponding test functions. The baroclinic mode is compatible with the barotropic mode, and mass conservation is ensured. The inconsistency is only due to the commutation between spatial discretization and depth integration needed to make the vertical dynamics terms disappear in the two-dimensional momentum equation.

Both discrete barotropic and baroclinic problems can be then written with matrix notations.

- The two-dimensional discrete barotropic problem reads:

$$\begin{bmatrix} M_{\boldsymbol{U}} & G \\ D & M_\eta \end{bmatrix} \begin{bmatrix} X_{\boldsymbol{U}} \\ X_\eta \end{bmatrix} = \begin{bmatrix} F_{\boldsymbol{U}} \\ 0 \end{bmatrix}, \tag{15}$$

- The three-dimensional discrete baroclinic problem reads:

$$\begin{bmatrix} M_{\boldsymbol{u}} & E^T \\ E & 0 \end{bmatrix} \begin{bmatrix} X_{\boldsymbol{u}} \\ X_\lambda \end{bmatrix} = \begin{bmatrix} F_{\boldsymbol{u}} \\ F_\lambda \end{bmatrix}, \tag{16}$$

where E is the matrix associated with the discrete compatibility constraint. The effect of these Lagrange multipliers is to correct the discrepancy due to the different treatment of vertical terms in the baroclinic and barotropic modes.

4 Global time-stepping algorithm

The global time-stepping algorithm be summarized as follows:

1. Evaluate the terms common for both 2D and 3D problems
2. Solve the 2D barotropic problem to obtain $\boldsymbol{U}$ and η
3. For each column of prisms, solve the 3D baroclinic problem
   (a) Evaluate the implicit terms for the momentum equation
   (b) Evaluate the terms for the compatibility constraint
   (c) Solve the local linear system to obtain $\boldsymbol{u}$ and λ
4. Integrate the continuity equation to obtain w
5. For each column of prisms, solve the tracer equations
   (a) Evaluate the implicit terms
   (b) Solve the local linear system to obtain T and/or S
6. Calculate the density ρ
7. Integrate the baroclinic pressure gradient $\boldsymbol{p}$

We build the global matrix for the two-dimensional barotropic mode and solve the corresponding linear system with a generalized minimal residual method iterative solver preconditioned with a block factorization combined with an additive-Schwartz coupling. An efficient implementation will avoid to duplicate computations for the two-dimensional barotropic mode and the three-dimensional baroclinic mode and will therefore compute all common term in a preliminary step.

For the three-dimensional momentum equations, the terms related to surface gravity waves, vertical advection, vertical diffusion, and Coriolis are treated (semi-) implicitly, while horizontal advection and diffusion are explicit. However, the Coriolis term could be stepped with about 1 h explicitly, and the necessity to treat vertical advection implicitly or explicitly depends on applications (coastal or global) and resolution as well. Moreover, horizontal advection is not always a much slower phenomenon compared with vertical advection. Therefore, this selection strongly depends on the application. We do not assemble the linear system for the three-dimensional momentum equation because the mass matrix for the discontinuous Galerkin methods is block diagonal per element and all implicit terms are local on vertically aligned prisms. Therefore, the linear system is block diagonal for each column of prisms. Each block is then solved locally using a sparse direct solver. The memory usage is not larger than for an explicit method, and this solution strategy is intrinsically scalable. As time-integration scheme, we use IMEX methods. Such an approach allows us to treat implicitly the linear terms corresponding to the stiff part of the problem, while treating explicitly the other terms.

The continuity equation is then locally solved because only stacked prisms are coupled together. Further, the information only goes from bottom to top, as we treat this equation as a steady advection equation, so that block per element Gauss–Seidel sweeping from bottom to top gives the exact result in a single iteration. The matrix for a column of prisms is block triangular. The equation for the pressure gradient force can also be solved for each column of prisms independently.

5 Implicit–explicit Runge–Kutta methods

In IMEX methods (Ascher et al. 1995, 1997), only the most critical terms are integrated implicitly. In our three-dimensional baroclinic free-surface model, we use IMEX Runge–Kutta methods. Those methods are self-starting. Moreover, the combination of discontinuous Galerkin methods with Runge–Kutta methods is known to be efficient (Cockburn and Shu 2001).

The spatial discretization of the three-dimensional baroclinic problem and the two-dimensional barotropic problem leads to systems of ordinary differential equations of the form:

$$\boldsymbol{y}'(t) = \boldsymbol{f}(\boldsymbol{y}(t), t), \tag{17}$$

where $\boldsymbol{y}(t)$ denotes the vector of all discrete degrees of freedom of a step (barotropic, baroclinic, or tracer problems) of the time marching algorithm. To integrate such an ordinary differential equation, explicit Runge–Kutta methods are quite popular. As a typical example, the second-order explicit method of Heun consists of calculating $\boldsymbol{y}_{n+1} \approx \boldsymbol{y}(t_{n+1})$ from $\boldsymbol{y}_n = \boldsymbol{y}(t_n)$ with the following sequence:

$$\begin{cases} \boldsymbol{K}_1 = \boldsymbol{f}\left(\boldsymbol{y}_n, t_n\right), \\ \boldsymbol{K}_2 = \boldsymbol{f}\left(\boldsymbol{y}_n + \Delta_t \boldsymbol{K}_1, t_n + \Delta_t\right), \\ \boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \Delta_t \left(\boldsymbol{K}_1 + \boldsymbol{K}_2\right)/2, \end{cases} \tag{18}$$

where Δ_t is the time interval. The accuracy of the discrete time integration performed with the Heun scheme is $\mathcal{O}(\Delta_t^2)$. Typically, the accuracy is often directly related to the number of stages (the number of $\boldsymbol{K}_i$ to be computed). In a more general way, an explicit or implicit Runge–Kutta method with k stages is defined by the following procedure:

$$\begin{cases} \boldsymbol{K}_i = \boldsymbol{f}\left(\boldsymbol{y}_n + \sum_{j=1}^{k} a_{ij} \Delta_t \boldsymbol{K}_j, t_n + c_i \Delta_t\right), \quad i = 1 \ldots k, \\ \boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \Delta_t \left(\sum_{j=1}^{k} b_j \boldsymbol{K}_j\right). \end{cases} \tag{19}$$

A very convenient and compact way to define a Runge–Kutta method consists in having recourse to the three arrays a_{ij}, b_j, and c_i, usually represented as the so-called *Butcher tableau* defined by:

$$\left[\begin{array}{c|c} c_i & a_{ij} \\ \hline & b_j \end{array}\right] \tag{20}$$

In explicit Runge–Kutta schemes, the non-vanishing entries of the array a are only located in the left lower triangular part of the matrix with zeros on the diagonal and the right upper triangular part. As an example, the Butcher tableau of the Heun explicit method (Eq. 18) is given by:

$$\left[\begin{array}{c|c} c_i & a_{ij} \\ \hline & b_j \end{array}\right] = \left[\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}\right]$$

In implicit Runge–Kutta (IRK) methods, the upper triangular of the array a contains non-zero entries. In

these implicit schemes, it is possible to consider two classes: IRK methods and diagonally implicit Runge–Kutta (DIRK). For IRK schemes, the array a is full and it is required to solve all the stages at the same time. Obviously, it can be extremely expensive, and it is not very popular. In general, DIRK schemes are usually resorted to, where the upper right triangular part of the array a is empty. In this case, each stage can be solved in an independent way. Moreover, single diagonally implicit Runge–Kutta (SDIRK) are often used when the diagonal coefficients are equal. Accordingly, for a linear problem, the matrix of the corresponding linear system will be the same for all stages.

IMEX Runge–Kutta schemes simply require the splitting of the function $\boldsymbol{f}$ into a part to be integrated explicitly and a part that will be handled by an implicit scheme. We decompose the ordinary differential equation (Eq. 17) in the following way:

$$\boldsymbol{y}'(t) = \overbrace{\boldsymbol{f}^{\text{expl}}(\boldsymbol{y}(t),t) + \boldsymbol{f}^{\text{impl}}(\boldsymbol{y}(t),t)}^{\boldsymbol{f}(\boldsymbol{y}(t),t)}, \tag{21}$$

where $\boldsymbol{f}^{\text{expl}}$ and $\boldsymbol{f}^{\text{impl}}$ represent the terms treated explicitly and implicitly, respectively. The explicit terms of the three-dimensional baroclinic mode are typically the nonlinear advection and the horizontal diffusion terms, while the terms corresponding to the gravity waves, the Coriolis force, and the vertical diffusion are solved with the (semi-)implicit method.

The IMEX method of order k consists in using a SDIRK method with $k-1$ stages combined with an explicit RK scheme with k stages. This time stepper scheme is defined by:

$$\boldsymbol{K}_1^{\text{expl}} = \boldsymbol{f}^{\text{expl}}(\boldsymbol{y}_n, t_n)$$

For $i = 2 \ldots k$

$$\begin{cases} \boldsymbol{K}_i^{\text{impl}} = \boldsymbol{f}^{\text{impl}}\Big(\underbrace{\boldsymbol{y}_n + \Delta_t \left(\sum_{j=1}^{i} a_{ij}^{\text{impl}} \boldsymbol{K}_j^{\text{impl}} + \sum_{j=1}^{i-1} a_{ij}^{\text{expl}} \boldsymbol{K}_j^{\text{expl}} \right)}_{\widehat{\boldsymbol{y}}_i}, t_n + c_i \Delta_t \Big), \\ \boldsymbol{K}_i^{\text{expl}} = \boldsymbol{f}^{\text{expl}}\left(\widehat{\boldsymbol{y}}_i, t_n + c_i \Delta_t\right), \end{cases} \tag{22}$$

$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \Delta_t \left(\sum_{j=1}^{k} b_j^{\text{impl}} \boldsymbol{K}_j^{\text{impl}} + b_j^{\text{expl}} \boldsymbol{K}_j^{\text{expl}} \right).$$

IMEX schemes can also be defined with two Butcher tableau corresponding to the implicit and the explicit part, respectively. In order to synchronize the stages, a unique c array applies to both methods and an initial empty stage is added to the implicit scheme, and the corresponding arrays are padded with zeros.

To obtain a suitable discretization for linear finite elements, it seems logical to have the same accuracy in time and space and to consider a second order scheme in time. Spatial and temporal discretization errors will then converge at the same rate when the mesh is refined, the time step being adapted in accordance with the CFL condition. The implicit explicit Runge–Kutta methods used in the three-dimensional baroclinic model were derived in Ascher et al. (1997), and his Butcher tableaux read:

$$\left[\begin{array}{c|c} c_i & a_{ij}^{\text{impl}} \\ \hline & b_j^{\text{impl}} \end{array}\right] = \left[\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \gamma & 0 & \gamma & 0 \\ 1 & 0 & 1-\gamma & \gamma \\ \hline & 0 & 1-\gamma & \gamma \end{array}\right], \tag{23}$$

$$\left[\begin{array}{c|c} c_i & a_{ij}^{\text{expl}} \\ \hline & b_j^{\text{expl}} \end{array}\right] = \left[\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \gamma & \gamma & 0 & 0 \\ 1 & \delta & 1-\delta & 0 \\ \hline & \delta & 1-\delta & 0 \end{array}\right], \tag{24}$$

with $\gamma = (2-\sqrt{2})/2$ and $\delta = 1 - 1/(2\gamma)$. The IMEX method consists in using a SDIRK method with two

stages combined with an explicit RK scheme with three stages. Some simplifications in the calculation can be deduced from those Butcher tableau. As the last line of the matrix a exactly corresponds to the line b, the final update can be obtained directly from the last estimate $\widehat{\boldsymbol{y}}_3$ obtained for the vector $\boldsymbol{y}$. Finally, the last entry of b^{expl} is zero and the last explicit stage is not needed. In short, we only need to calculate two times the explicit part $\boldsymbol{f}^{\text{expl}}$ and to solve two times the implicit system associated with $\boldsymbol{f}^{\text{impl}}$. This system has the same matrix but a different right-hand side. In a systematic way, the IMEX procedure for the three-dimensional baroclinic model can be cast in the following sequence:

1. Calculate $\boldsymbol{K}_1^{\text{expl}} = \boldsymbol{f}^{\text{expl}}(\boldsymbol{y}_n, t_n)$
2. Obtain $\boldsymbol{K}_2^{\text{impl}}$ and $\widehat{\boldsymbol{y}}_2$ by solving:
$$\boldsymbol{K}_2^{\text{impl}} = \boldsymbol{f}^{\text{impl}}(\underbrace{\boldsymbol{y}_n + \Delta_t(\gamma \boldsymbol{K}_1^{\text{expl}} + \gamma \boldsymbol{K}_2^{\text{impl}})}_{\widehat{\boldsymbol{y}}_2}, t_n + \Delta_t \gamma)$$
3. Calculate $\boldsymbol{K}_2^{\text{expl}} = \boldsymbol{f}^{\text{expl}}(\widehat{\boldsymbol{y}}_2, t_n + \Delta_t \gamma)$
4. Obtain $\boldsymbol{K}_3^{\text{impl}}$ and $\widehat{\boldsymbol{y}}_3$ by solving:
$$\boldsymbol{K}_3^{\text{impl}} = \boldsymbol{f}^{\text{impl}}(\underbrace{\boldsymbol{y}_n + \Delta_t(\delta \boldsymbol{K}_1^{\text{expl}} + (1-\delta)\boldsymbol{K}_2^{\text{expl}} + (1-\gamma)\boldsymbol{K}_2^{\text{impl}} + \gamma \boldsymbol{K}_3^{\text{impl}})}_{\widehat{\boldsymbol{y}}_3}, t_n + \Delta_t)$$
5. Set $\boldsymbol{y}_{n+1} = \widehat{\boldsymbol{y}}_3$

6 Numerical results

Internal waves in the lee of a moderately tall seamount were simulated in Blaise et al. (2010). In this paper, we consider a similar setup to compare the transient dynamics with results published by Chapman and Haidvogel (1992) and in Section 3c of Ford et al. (2004). This setup is selected because there is no need for subgrid-scale parametrization to create complex baroclinic phenomena. The flow is stratified but subcritical: There is no internal wave breakup. No boundary layer appears, as a slip condition at the seabed is prescribed. In the first part of this work (Blaise et al. 2010), the height of the seamount was 30% of the total depth, and a complicated internal wave structure developed in the wake of the seamount. Now, we consider that the height of the seamount is 90% of the total depth in order to observe some recirculation patterns in the wake of the seamount.

The first computation of a three-dimensional linearly stratified flow over a Gaussian seamount was done by Huppert and Bryan (1976) with the model of Bryan (1969). A detailed numerical study of flows past Gaussian seamounts can be found in Chapman and Haidvogel (1992, 1993). The strengths and weaknesses of a few other models have been assessed by simulating flow past seamounts: MIT general circulation model (Adcroft et al. 1997), Imperial College Ocean Model (Ford et al. 2004), and FEOM (Wang et al. 2008). Our numerical simulations will be performed in order to draw some comparisons with previous computations.

The computational domain is an aquaplanet, as it allows us to avoid open boundary conditions. Figure 2 shows a close-up view of the mesh and the bathymetry near the seamount. The mesh resolution is refined in the lee of the seamount, to allow for an accurate representation of the shedded vortex described in Chapman and Haidvogel (1992) and Ford et al. (2004). The edge length in the most refined zone is 2 km. This mesh is made up of 13,836 triangles extruded into 20σ layers.

The geometry of the problem is defined by a Gaussian seamount located at a latitude of 45° N. The bathymetry reads:

$$1 - \frac{h(x,y)}{H} = \delta \exp\left(\frac{\left(x-\frac{R}{2}\right)^2 + \left(y-\frac{R}{2}\right)^2 + \left(z-\frac{R}{\sqrt{2}}\right)^2}{-L^2}\right), \tag{25}$$

where $H = 4.5$ km is the total depth, $\delta = 0.9$ is the relative height of the seamount, $R = 6\ 372$ km is Earth radius, and $L = 25$ km is the length scale of the seamount. The coordinates x, y, and z are relative to the global Cartesian reference coordinates axis located in the center of the sphere. The flow simulation is initiated with a global zonal geostrophic equilibrium ignoring the seamount. In other words, the initial guess of the calculation is the same as in the test case 5 of Williamson et al. (1992) where the velocity field only exhibits a non-vanishing east component u_e. In this test

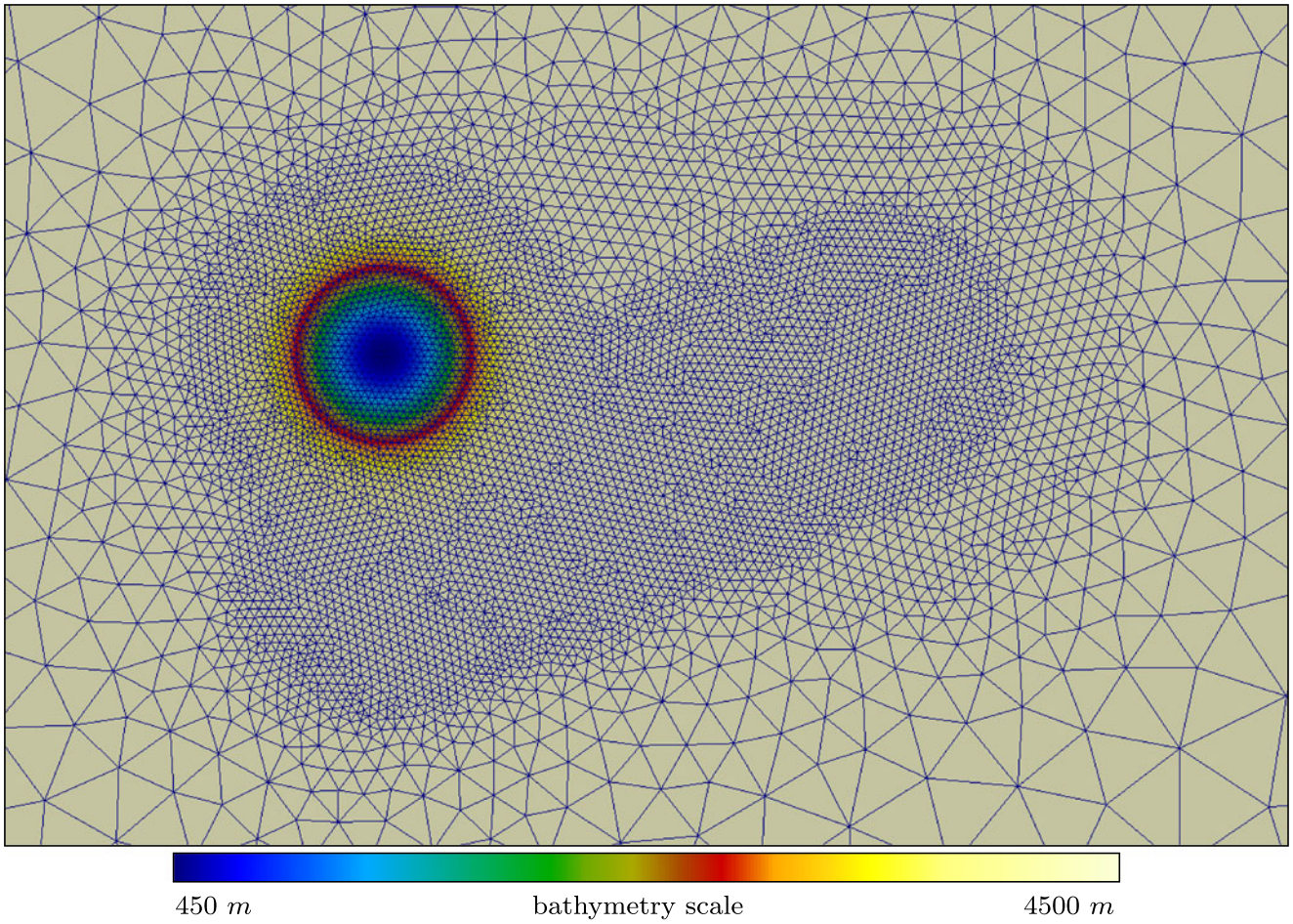

Fig. 2 Close-up view on the mesh and the bathymetry around the seamount. The mesh is refined in the lee of the seamount

case, the elevation and velocity fields are respectively given by

$$\frac{\eta}{U^2/g} = -\frac{z^2}{R^2}\left(1+\sqrt{2}\,\frac{R\Omega}{U}\right), \tag{26}$$

$$\frac{u_e}{U} = \sqrt{\frac{x^2+y^2}{R^2}}, \tag{27}$$

where $U = 0.258$ m s^{-1} is the velocity scale at a latitude of 45° N, $\Omega = 7.292 \times 10^{-5}$ s^{-1} is Earth rotation rate, and $g = 9.81$ m s^{-2} is the gravitational acceleration. We only consider the density deviation ρ' as the unique tracer of the model, and the initial value of the density deviation is a linear function of the vertical coordinate, with vanishing mean. The derivative of the density with respect to the vertical coordinate is given by $\partial\rho/\partial z = -3.43 \times 10^{-5}$ kg m^{-4}, and the reference density is selected as $\rho_0 = 1{,}025$ kg m^{-3}. The turbulent viscosities and diffusivities are given by: $\nu_h = \kappa_h = 6.45$ m^2 s^{-1}, $\nu_v = 0.0001$ m^2 s^{-1} and $\kappa_v = 0$. With those parameters, we consider that the flow is characterized by the same four dimensionless numbers as that in Section 3.c of Ford et al. (2004). These dimensionless number are defined as follows:

- Seamount ratio $\quad \delta = 0.9$
- Rossby number $\quad Ro = \dfrac{U}{fL} = 0.1$
- Reynolds number $\quad Re = \dfrac{UL}{\nu_h} = 1000$
- Burger number $\quad Bu = \dfrac{NH}{fL} = \sqrt{\dfrac{-g}{\rho_0}\dfrac{\partial_\rho}{\partial_z}}\dfrac{H}{fL} = 1$

where N is the Brunt–Väisälä frequency.

The critical numerical parameter in the three-dimensional baroclinic model is the jump penalty coefficient γ of the Lax–Friedrichs solver. For this problem, we select $\gamma = 6$ m s^{-1}. Here, we select a slightly higher γ than for the simulations presented in Blaise et al. (2010) because the height of the seamount is quite larger. Above the seamount, the density profile may be significantly altered, and this parameter must

be an upper bound of the phase speed of the fastest wave. For discontinuous linear elements combined with the second-order explicit Runge–Kutta time stepper (Chevaugeon et al. 2007) used in this simulation, the relevant CFL conditions reads:

$$\Delta_t < \frac{\Delta_x}{3\gamma} \tag{28}$$

The smallest edge length is 2 km and the relevant length is the inradius of this smallest triangle. Therefore, $\Delta_x = 0.29 \times 2$ km and the greatest time step to avoid instabilities is 32 s. In this simulation, we use a time step of 20 s.

The two-dimensional dynamics of flows past isolated obstacles is already complicated. Verron and

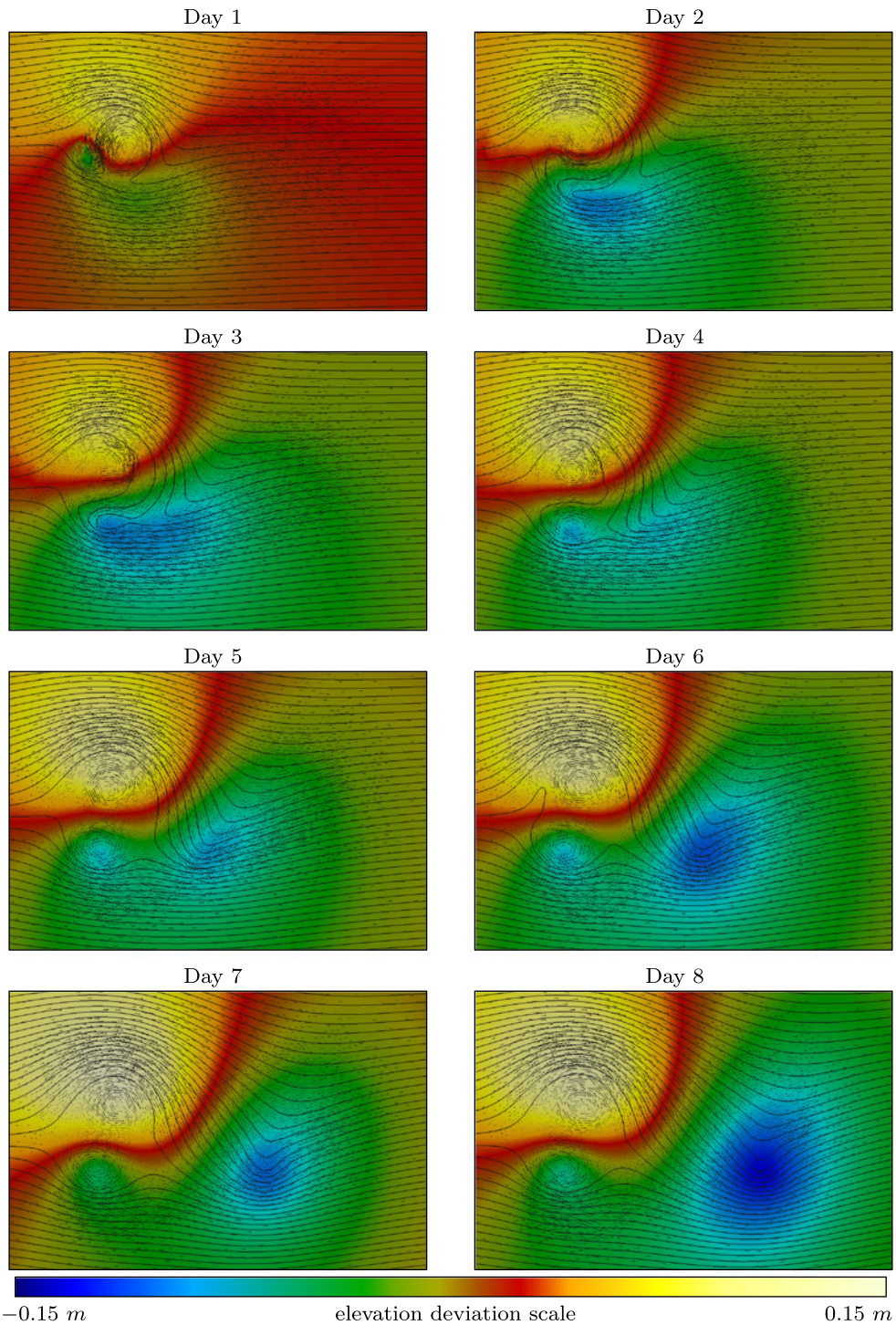

Fig. 3 Two-dimensional flow ($\delta = 0.9$). *Colors* denote sea surface deviation with respect to initial geostrophic equilibrium. *Glyphs* represents two-dimensional mean velocities. The *black continuous lines* are the instantaneous streamlines

Le Provost (1985) give a detailed analysis of the flows that occurs over an isolated seamount, using a two-dimensional model of the quasi-geostrophic equations. A clockwise vortex is always trapped over the seamount, but several transient regimes can occur. For strong flows, such as the configuration given in Blaise et al. (2010), the counterclockwise vortex generated in the initiation flow phase is directly advected downstream. For weak flows, such as the current configuration, a stronger interaction between the two eddies occurs and the counterclockwise vortex is shifted to the right and trapped in the vicinity of the seamount, leading to a double vortex structure.

Two-dimensional daily depth-averaged velocities and the sea surface deviation are shown in Fig. 3. The sea surface deviation is defined as the difference between the sea-surface elevation and initial elevation corresponding to the geostrophic elevation. As the flow

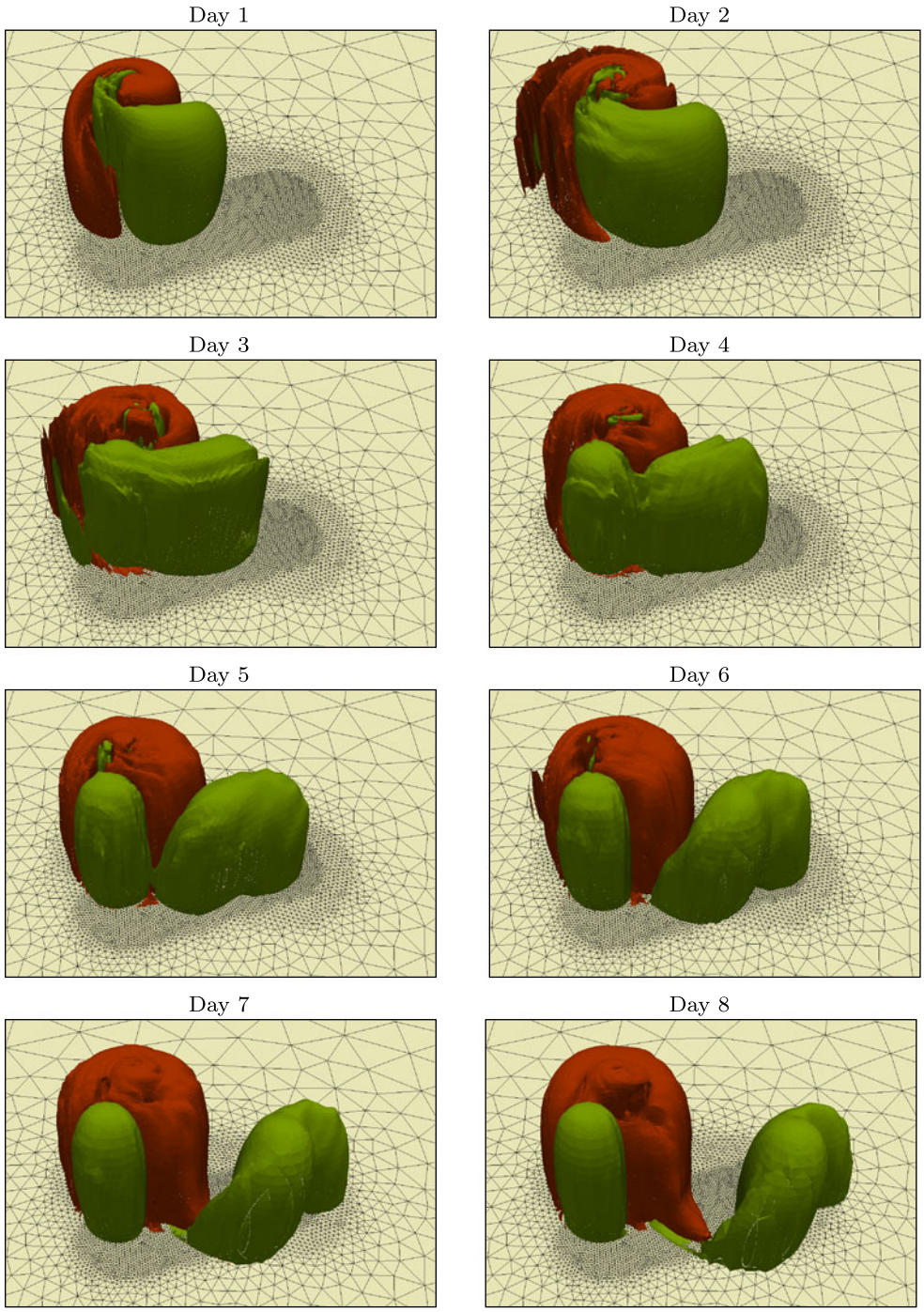

Fig. 4 Time evolution of the isosurfaces of the density perturbation. Isovalues of density perturbation of -0.001 kg m^{-3} are in *green*. Isovalues of density perturbation of 0.001 kg m $^{-3}$ are in *red*. The two-dimensional mesh is given on the sea bottom

is impulsively started, the free-surface is raised in front of the seamount and lowered behind it. Geostrophic adjustment induces two counter-rotating eddies, the one in front of the seamount being clockwise. Under a rigid-lid approximation, this adjustment can be interpreted in terms of vortex compression and stretching (Verron and Le Provost 1985). These two vortices progress clockwise around the seamount, with a time scale much smaller than the advective one. In Fig. 3, we see that at day 1, the two vortices have already rotated almost half a turn clockwise. The rotation of this vortex pair can be explained in terms of topographic Rossby waves. Depth variation induces effects similar to Coriolis parameter variation, i.e., the β effect. These waves

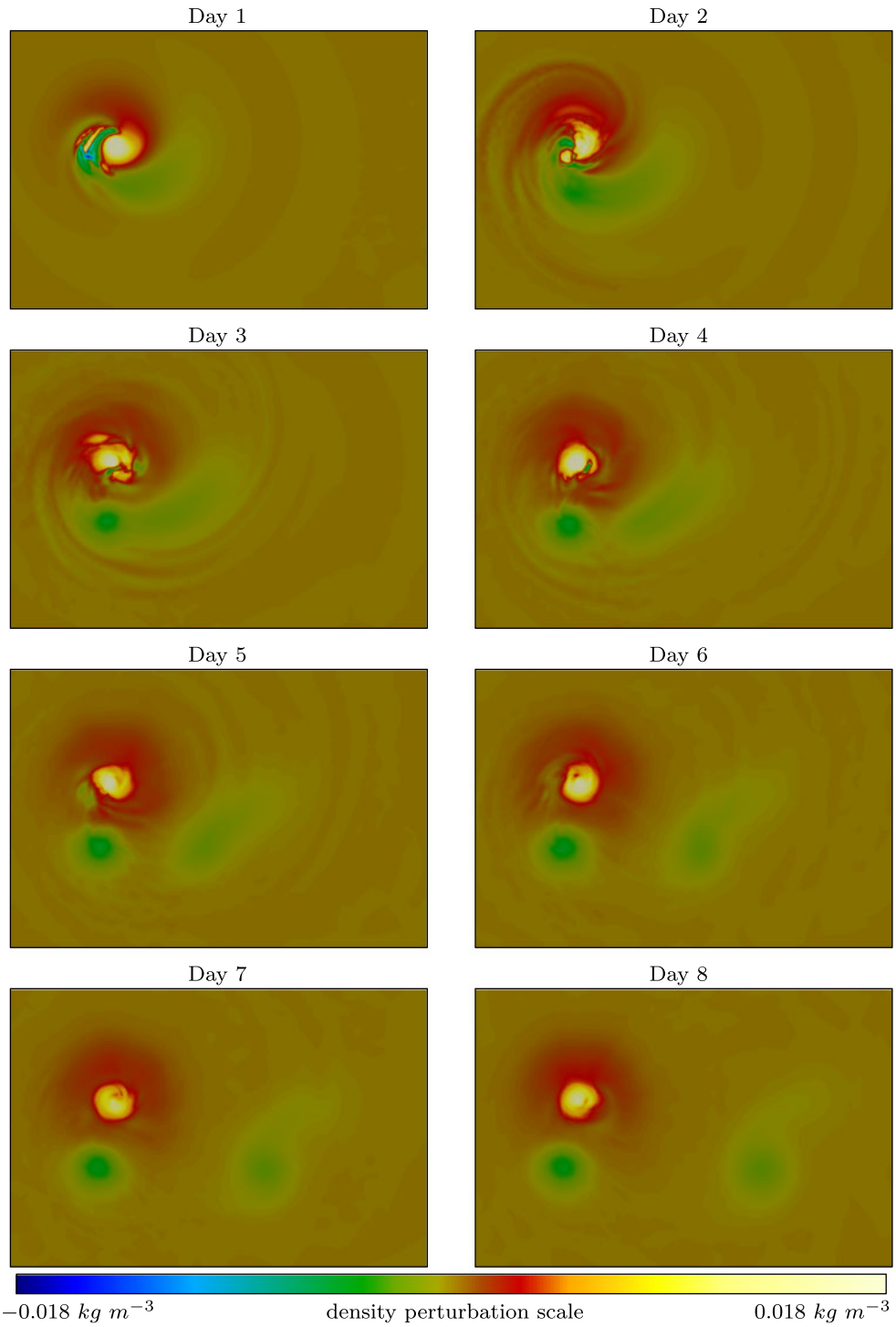

Fig. 5 Density perturbation for a horizontal cut at 400 m depth

propagate with the shallowest area on their right (Cushman-Roisin 1994, Section 6.5). This leads to a clockwise progression of the vortex pair. The phase speed aligned with isobaths is proportional to the bottomslope. For a Gaussian-shaped bathymetry, the maximum speed will therefore occur at a radius corresponding to the inflexion point of the Gaussian, which in this case is $L/\sqrt{2}$. The initially circularly shaped vortices tend to become spiral-shaped, as explained in Johnson (1984). The flow still exhibit some global coherent structures and have been modeled with two-dimensional approximations such as quasi-geostrophic

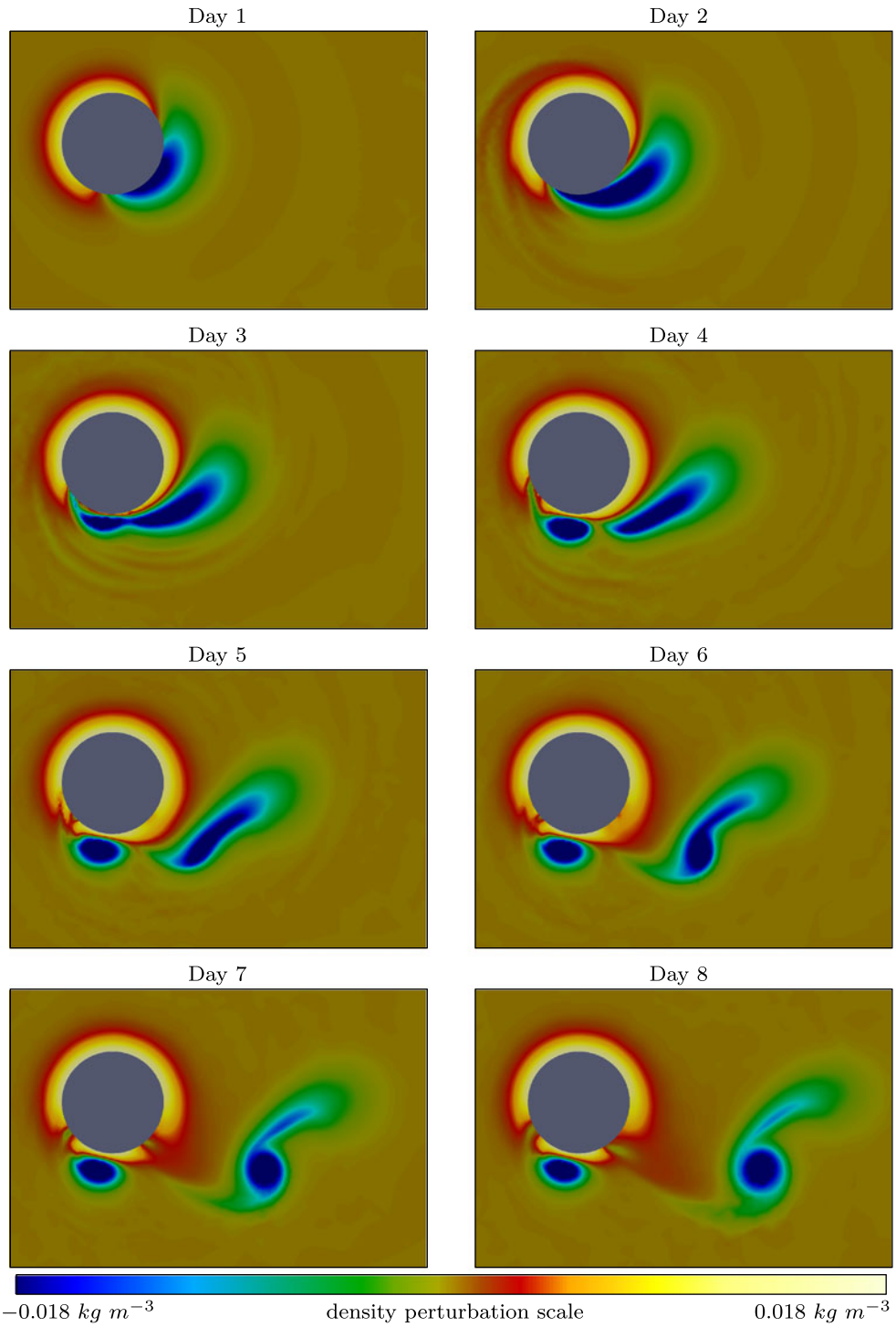

Fig. 6 Density perturbation for a horizontal cut at 4,000 m depth

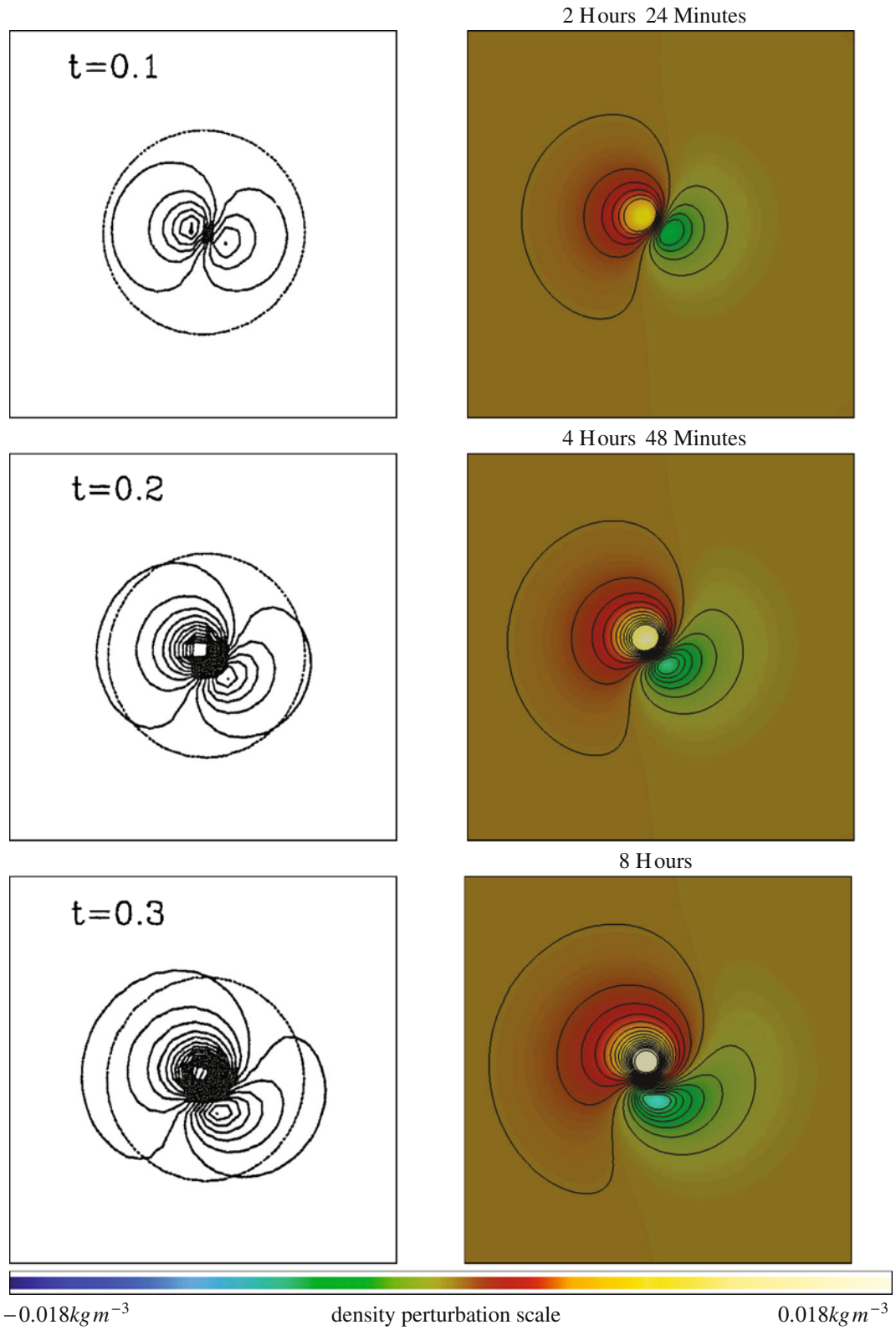

Fig. 7 Comparison of density perturbation field at 400 m depth with results obtained by Chapman and Haidvogel (1992), during the startup of the calculations. The *dashed circle* in the reference data is the 4,000-m isobath. For $t = 0.1$, isolines range from -0.0039 to 0.0051. For $t = 0.2$, isolines range from -0.0059 to 0.0101. For $t = 0.3$, isolines range from -0.0069 to 0.0161. Isolines interval is 0.001 kg m^{-3}. The same values are used for *all panels*

equations (Johnson 1984; Verron and Le Provost 1985).

The vertical structure of the flow can be observed from the isosurfaces of the density perturbation given in Fig. 4. The density perturbation is defined as the difference between density deviation field ρ' and the initial density deviation. As the flow is quasi-geostrophic, the density perturbation is an image of the vorticity, as the elevation deviation. In fact, a complex interaction takes place, where the counterclockwise vortex undergoes a stretching and breaking sequence that generates internal waves in the lee of the seamount. This can be observed in Fig. 5 where a horizontal slice in the density perturbation at a depth of 400 m is displayed.

Using the quasi-geostrophic equations, Johnson (1984) shows that the starting flow over a smooth ob-

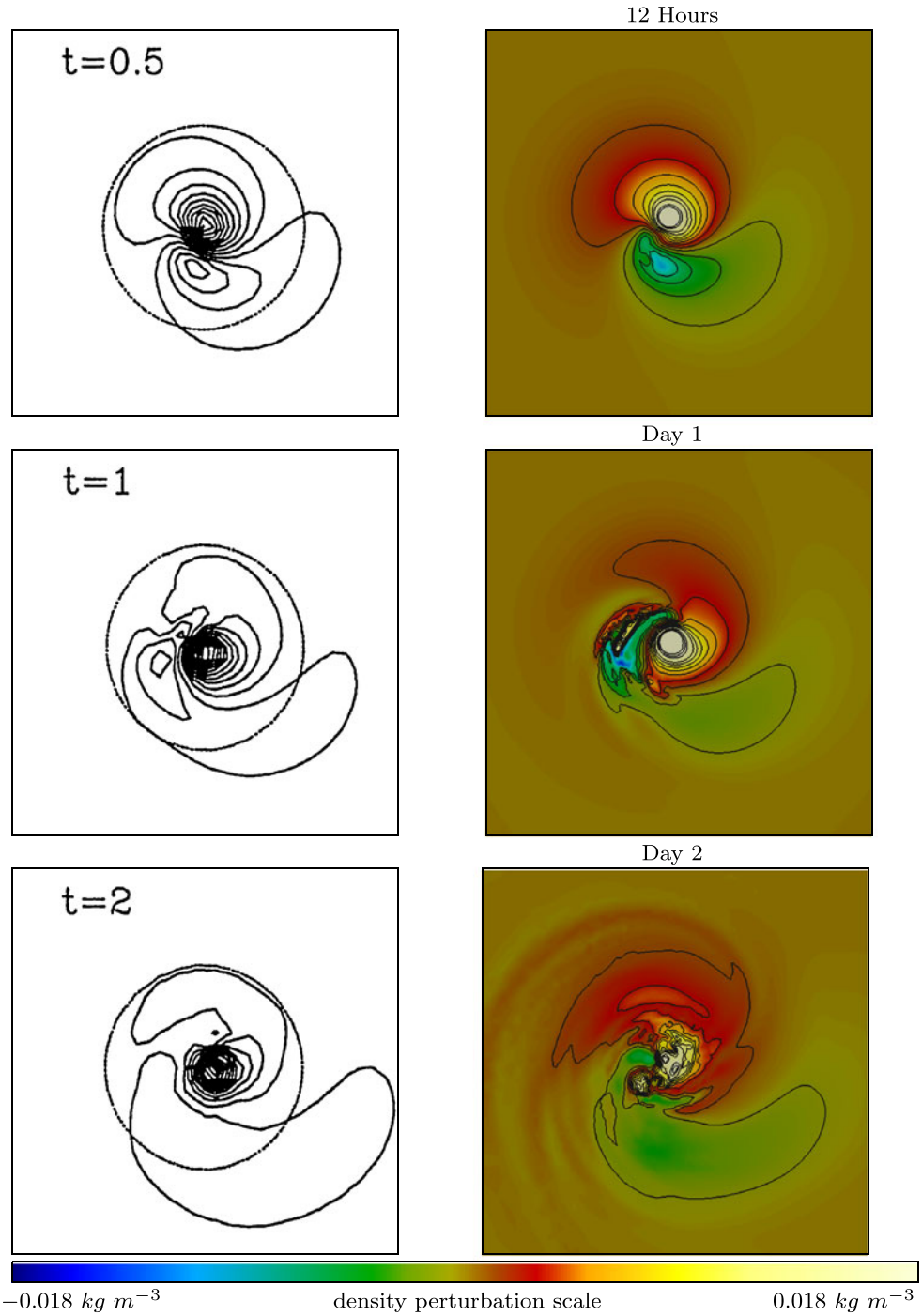

Fig. 8 Comparison of density perturbation field at 400 m depth with results obtained by Chapman and Haidvogel (1992) during the two first days. The *dashed circle* in the reference data is the 4,000-m isobath. For $t = 0.5$, isolines range from −0.0069 to 0.0191. For $t = 1.0$, isolines range from −0.0049 to 0.0211. For $t = 2.0$, isolines range from −0.0009 to 0.0231. Isolines interval is 0.002 kg m^{-3}. The same values are used for *all panels*

stacle leads to topographic Rossby waves that rotate clockwise around the obstacle. For a parabolic obstacle, spiral waves are observed. These internal topographic Rossby waves are progressing as spirals between days 2 and 3. These waves can be clearly detected from the density perturbation isovalues at a depth of 4,000 m in Fig. 6. The counterclockwise vortex then breaks into two well-separated parts. The first one is trapped on the right side of the seamount (looking downstream), while the second one is ejected and transported at the mean speed of the flow. Indeed, two recirculation cells exist. One is trapped over the seamount, and the other one is located on the right side of the seamount.

Finally, it is instructive to perform some comparisons between our numerical simulations and some previous calculations. A quite similar flow in a rectangular domain with constant Coriolis parameter was simulated by several authors:

- Chapman and Haidvogel (1992) use a rigid lid model with finite difference horizontal discretization and spectral vertical discretization and along sigma levels hyperviscous dissipation.
- Adcroft et al. (1997) use a rigid lid finite volume model but only provide results after 10 days.
- Ford et al. (2004) use a finite element non-hydrostatic model, with rigid-lid approximation and Laplacian dissipation.

The problem is not exactly the same as we introduce the full Coriolis term on the sphere, while the f-plane

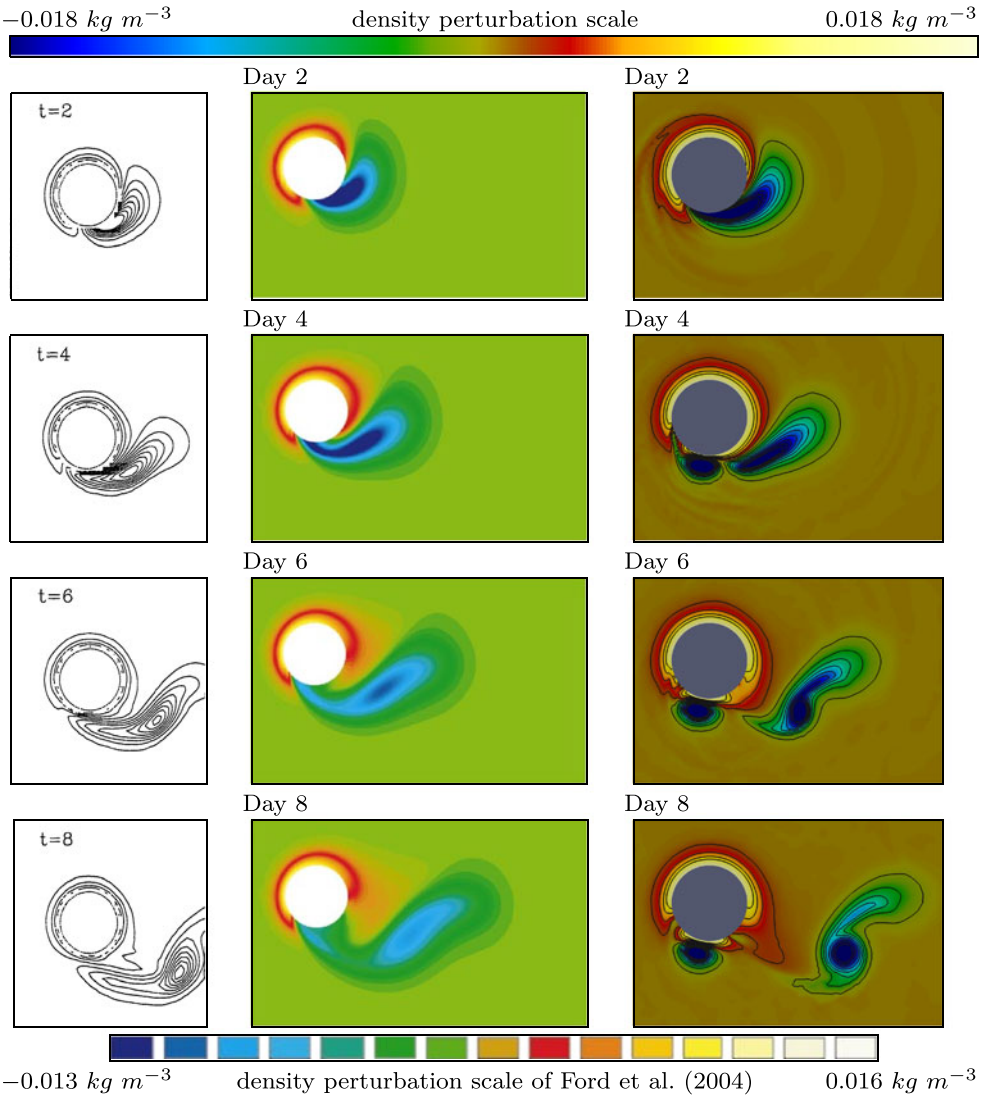

Fig. 9 Comparison of density perturbation field at 4,000 m depth with results of Chapman and Haidvogel (1992) (*left*) and of Ford et al. (2004) (*middle*) during the first 8 days. Isolevels are the same for the *left* and *right sides* and range from −0.0212 to 0.0088 with a interval of 0.0025 kg m^{-3}

approximation is considered in those previous calculations. However, this should not induce significant flow discrepancy, as we focus on the small scale, where the variations of the Coriolis parameter are negligible compared to other effects, such as the influence of the bathymetry.

For the early stages of the flow, our model exhibits numerous wave phenomena. In Figs. 7 and 8, we show a detailed comparison with the results obtained by Chapman and Haidvogel (1992). Both simulations produce a quite similar behavior during the first 2 days. However, our calculation exhibits significant internal spiral waves that do not appear in Chapman and Haidvogel (1992) and Ford et al. (2004). This observation could be explained by the rather smaller dissipation introduced by our numerical scheme. Those waves could also be due to an interaction with the free surface, while Chapman and Haidvogel (1992) and Ford et al. (2004) use the rigid lid approximation. In Fig. 9, we observe that the counterclockwise eddy is stretched and breaks up, one part being trapped near the seamount while the other is shedded. In our computation, the breakup of the eddy happens much faster. The trapped eddy is much larger, and the shedded eddy tends to become much more circular. This difference can be explained easily: Ford et al. (2004) and Chapman and Haidvogel (1992) carry out their computation in a box domain, with lateral walls. Indeed, these walls are too close to the seamount for their influence to be negligible. Figure 10 sketches the elevation deviation after 3 and 7 days, along with black lines located where the lateral walls are found in Ford et al. (2004). It can be seen that the flow clearly varies along those boundaries. When those boundaries are present, the counterclockwise vortex cannot develop laterally, and this prevents its breakup.

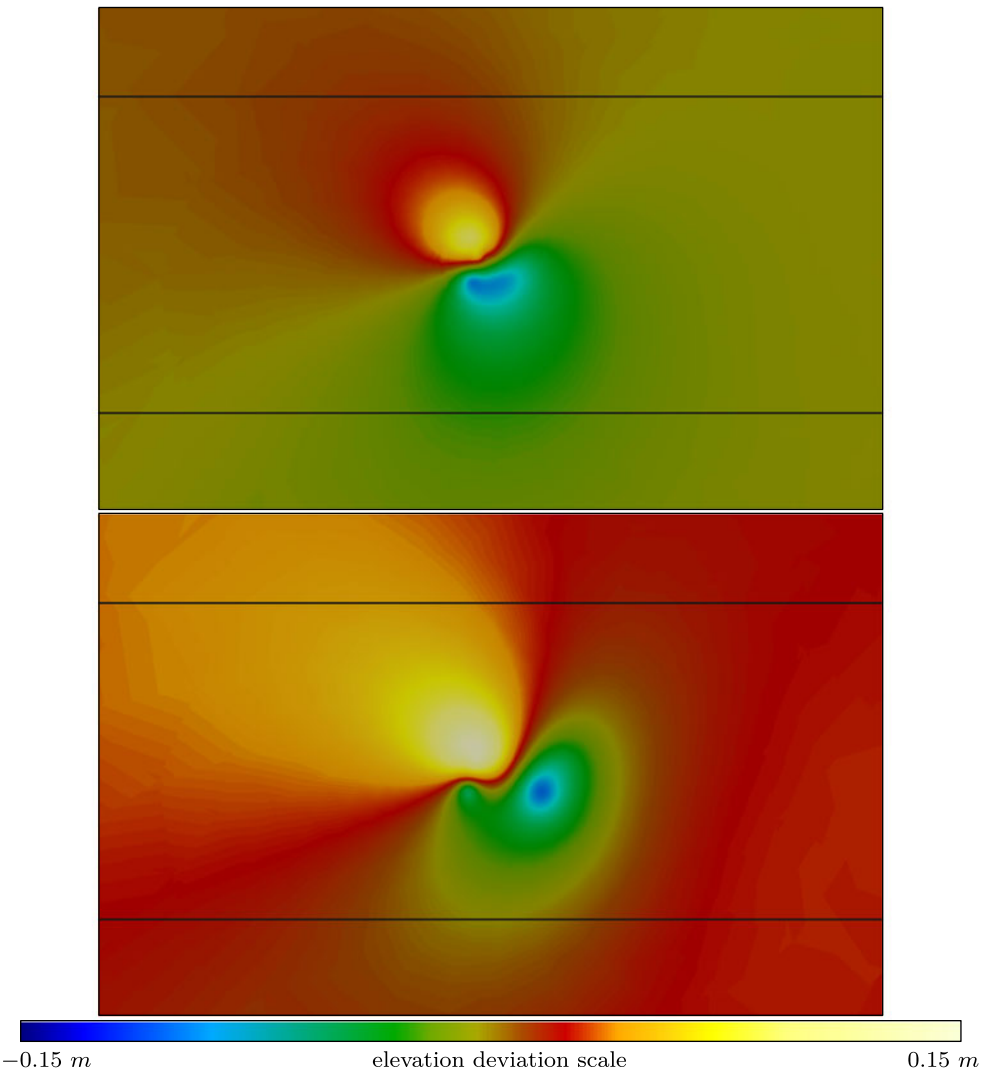

Fig. 10 Sea surface elevation deviation after 3 days (*top*) and 7 days (*bottom*). The *black lines* denote the location of lateral boundary in Ford et al. (2004). It can be observed that the flow significantly varies outside of the domain denoted by the *black lines*

7 Conclusions

An implicit–explicit time discretization for the three-dimensional free-surface baroclinic marine model described in Blaise et al. (2010) is proposed. The major contribution consists in the definition of a new implicit mode-splitting procedure with compatible barotropic and baroclinic problems. To achieve this, the two-dimensional barotropic problem is discretely obtained from the three-dimensional baroclinic problem. Compatibility between the two modes is enforced in a weak way by introducing Lagrange multipliers. The transports in the two- and three-dimensional problems are constrained by explicitly incorporating this compatibility constraint in the three-dimensional horizontal momentum equation. Combined with IMEX Runge–Kutta methods, such an approach sounds very attractive. The order of accuracy can be selected as required. On the one hand, we take advantage of the stability of the implicit method that will damp the unresolved or poorly resolved modes. On the other hand, we could also benefit from the total variation diminishing property of the explicit part of some methods.

Revisiting the benchmark flow over an isolated seamount of Blaise et al. (2010), we simulate the complex spiral wave dynamics that previous calculations were not able to capture either because of the rigid-lid assumption or their numerical methods. Thanks to the unstructured nature of the mesh, the resolution is refined in the lee of the seamount, enabling a detailed representation of the wave dynamics in this region. Further, vortex shedding is observed. The early stages of the simulation compare well with the two previous calculations.

For such an implicit–explicit approach to be interesting, the discrete operators for the dynamics handled implicitly must be significantly stiffer than those for the explicit dynamics. Indeed, the time step allowed by the IMEX scheme must be significantly larger than the time step of a purely explicit discretization. It is definitely the case when the vertical mixing parameters are deduced from a turbulence closure. For the simulation of the internal waves in the lee of a moderately tall seamount, it is not really the case as the stratification is rather strong. The internal waves are fast, and the vertical viscosity is still reasonable. The time steps are only 20 times larger than the explicit time step and are much more expensive, as local linear systems are solved. In our opinion, the way to faster computations is twofold. On one hand, the computation of the discrete terms can still be improved, by recasting most of the operations into efficient matrix–matrix products computed with highly optimized linear algebra subroutines. It is the classical optimization procedure of a numerical model. On the other hand, the time-stepping strategy can itself be improved. Indeed, most ocean models have resorted to a mode splitting approach to avoid solving three-dimensional linear systems. It may be necessary to go beyond this paradigm and investigate a full implicit approach. To be efficient, it must be scalable. Multigrid methods have the potential to provide scalable solutions to large-scale discrete problems. Further such multigrid methods do not need the matrix of the linear system to be assembled, significantly reducing the memory footprint of the algorithm. However, the design of an efficient multigrid algorithm is in itself a whole domain of research and goes much beyond the scope of this work.

Acknowledgements Sébastien Blaise and Jonathan Lambrechts are research fellows with the Belgian Fund for Research in Industry and Agriculture (FRIA). Richard Comblen is a research fellow with the Belgian National Fund for Scientific Research (FNRS). Eric Deleersnijder is a research associate with the Belgian National Fund for Scientific Research (FNRS). The research was conducted within the framework of the Interuniversity Attraction Pole TIMOTHY (IAP 6.13), funded by the Belgian Science Policy (BELSPO), and the programme ARC 04/09-316, funded by the Communauté Française de Belgique.

References

Adcroft A, Hill C, Marshall J (1997) Representation of topography by shaved cells in a height coordinate ocean model. Mon Weather Rev 95:2293–2315

Ascher UM, Ruuth SJ, Spiteri RJ (1997) Implicit–explicit Runge–Kutta methods for time-dependent partial differential equations. Appl Numer Math 25:151–167

Ascher UM, Ruuth SJ, Wetton BT (1995) Implicit–explicit methods for time-dependent partial differential equations. SIAM J Numer Anal 32:797–823

Blaise S, Comblen R, Legat V, Remacle J-F, Deleersnijder E, Lambrechts J (2010) A discontinuous finite element baroclinic marine model on unstructured prismatic meshes. Part I: space discretization. Ocean Dyn (in this issue)

Blumberg AF, Mellor GL (1987) A description of three-dimensional coastal ocean circulation model. In: Heaps NS (ed) Three dimensional coastal ocean model. American Geophysical Union, Washington, DC, pp 1–16

Bryan K (1969) A numerical method for the study of the circulation of the world ocean. J Comput Phys 4:347–376

Chapman DC, Haidvogel DB (1992) Formation of Taylor caps over a tall isolated seamount in a stratified ocean. Geophys Astrophys Fluid Dyn 64:31–65

Chapman DC, Haidvogel DB (1993) Generation of internal lee waves trapped over a tall isolated seamount. Geophys Astrophys Fluid Dyn 69:33–54

Chevaugeon N, Hillewaert K, Gallez X, Ploumhans P, Remacle J-F (2007) Optimal numerical parametrization of discontinuous Galerkin method applied to wave propagation problems. J Comput Phys 223:188–207

Cockburn B, Shu C-W (2001) Runge–Kutta discontinuous Galerkin methods for convection-dominated problems. J Sci Comput 16:173–261

Cushman-Roisin B (1994) Introduction to geophysical fluid dynamics. Prentice Hall, New York

Deleersnijder E (1993) Numerical mass conservation in a free-surface sigma coordinate marine mode with mode splitting. J Mar Syst 4:365–370

Dukowicz JK, Smith RD (1994) Implicit free-surface method for the Bryan–Cox–Semtner ocean model. J Geophys Res 99:7991–8014

Dukowicz JK, Smith RD, Malone RC (1993) A reformulation and implementation of the Bryan–Cox–Semtner ocean model on the connection machine. J Atmos Ocean Technol 10:195–208

Ford R, Pain CC, Piggott M, Goddard A, de Oliveira CR, Umpleby A (2004) A non-hydrostatic finite element model for three-dimensional stratified oceanic flows, part II: model validation. Mon Weather Rev 132:2832–2844

Giraldo FX, Perot JB, Fischer PF (2003) A spectral element semi-Lagrangian (SESL) method for the spherical shallow water equations. J Comput Phys 190:623–650

Griffies SM, Böning C, Bryan FO, Chassignet EP, Gerdes R, Hasumi H, Hirst A, Treguier A-M, Webb D (2000) Developments in ocean climate modeling. Ocean Model 2:123–192

Hallberg R (1997) Stable split time stepping schemes for large-scale ocean modeling. J Comput Phys 135:54–56

Higdon RL, de Szoeke RA (1997) Barotropic–baroclinic time splitting for ocean circulation modeling. J Comput Phys 135:30–53

Huppert HE, Bryan K (1976) Topographically generated eddies. Deep-Sea Res 23:655–679

Johnson ER (1984) Starting flow for an obstacle moving transversely in a rapidly rotating fluid. J Fluid Mech 149:71–88

Killworth PD, Stainforth D, Webb DJ, Paterson SM (1991) The development of a free-surface Bryan–Cox–Semtner ocean model. J Phys Oceanogr 21:1333–1348

Kubatko E, Dawson C, Westerink J (2008) Time step restrictions for Runge–Kutta discontinuous Galerkin methods on triangular grids. J Comput Phys 227:9697–9710

Marshall JC, Adcroft AJ, Hill CN, Perelman L, Heisey C (1997) A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. J Geophys Res 102:5753–5766

Verron J, Le Provost C (1985) A numerical study of quasi-geostrophic flow over isolated topography. J Fluid Mech 154:231–252

Wang Q (2007) The finite element ocean model and its aspect of vertical discretization. Ph.D. thesis, Bremen University

Wang Q, Danilov S, Schröter J (2008) Finite element ocean circulation model based on triangular prismatic elements, with application in studying the effect of topography representation. J Geophys Res 113:C05015

White L, Legat V, Deleersnijder E (2008) Tracer conservation for three-dimensional, finite-element, free-surface, ocean modeling on moving prismatic meshes. Mon Weather Rev 136:420–442

Williamson DL, Drake JB, Hack JJ, Jakob R, Swarztrauber PN (1992) A standard test set for numerical approximations to the shallow water equations in spherical geometry. J Comput Phys 102:211–224

Ocean Dynamics (2010) 60:1415–1445
DOI 10.1007/s10236-010-0351-x

High-order schemes for 2D unsteady biogeochemical ocean models

Mattheus P. Ueckermann · Pierre F. J. Lermusiaux

Received: 20 June 2010 / Accepted: 19 October 2010 / Published online: 16 November 2010

Abstract Accurate numerical modeling of biogeochemical ocean dynamics is essential for numerous applications, including coastal ecosystem science, environmental management and energy, and climate dynamics. Evaluating computational requirements for such often highly nonlinear and multiscale dynamics is critical. To do so, we complete comprehensive numerical analyses, comparing low- to high-order discretization schemes, both in time and space, employing standard and hybrid discontinuous Galerkin finite element methods, on both straight and new curved elements. Our analyses and syntheses focus on nutrient–phytoplankton–zooplankton dynamics under advection and diffusion within an ocean strait or sill, in an idealized 2D geometry. For the dynamics, we investigate three biological regimes, one with single stable points at all depths and two with stable limit cycles. We also examine interactions that are dominated by the biology, by the advection, or that are balanced. For these regimes and interactions, we study the sensitivity to multiple numerical parameters including quadrature-free and quadrature-based discretizations of the source terms, order of the spatial discretizations of advection and diffusion operators, order of the temporal discretization in explicit schemes, and resolution of the spatial mesh, with and without curved elements. A first finding is that both quadrature-based and quadrature-free discretizations give accurate results in well-resolved regions, but the quadrature-based scheme has smaller errors in under-resolved regions. We show that low-order temporal discretizations allow rapidly growing numerical errors in biological fields. We find that if a spatial discretization (mesh resolution and polynomial degree) does not resolve the solution, oscillations due to discontinuities in tracer fields can be locally significant for both low- and high-order discretizations. When the solution is sufficiently resolved, higher-order schemes on coarser grids perform better (higher accuracy, less dissipative) for the same cost than lower-order scheme on finer grids. This result applies to both passive and reactive tracers and is confirmed by quantitative analyses of truncation errors and smoothness of solution fields. To reduce oscillations in un-resolved regions, we develop a numerical filter that is active only when and where the solution is not smooth locally. Finally, we consider idealized simulations of biological patchiness. Results reveal that higher-order numerical schemes can maintain patches for long-term integrations while lower-order schemes are much too dissipative and cannot, even at very high resolutions. Implications for the use of simulations to better understand biological blooms, patchiness, and other nonlinear reactive dynamics in coastal regions with complex bathymetric features are considerable.

Responsible Editor: Vincent Legat

Electronic supplementary material The online version of this article (doi:10.1007/s10236-010-0351-x) contains supplementary material, which is available to authorized users.

M. P. Ueckermann · P. F. J. Lermusiaux (✉)
Massachusetts Institute of Technology,
77 Mass. Avenue, Cambridge, MA 02139, USA
e-mail: pierrel@mit.edu

M. P. Ueckermann
e-mail: mpuecker@mit.edu

Keywords Ocean straits · Shelfbreaks · Biogeochemical dynamics · High order · Curved boundary · Quadrature · Ocean modeling · Discontinuous Galerkin

1 Introduction

Accurate modeling of biogeochemical–physical ocean dynamics is required for multiple scientific and societal applications, covering a wide range of time and space scalves. With the increased understanding of biogeochemical interactions (Lalli and Parsons 1997; Robinson et al. 2002; Fennel and Neumann 2004), ecosystems models have substantially improved in the past decades (Fasham et al. 1990; Hofmann and Lascara 1998; Robinson and Lermusiaux 1999; Hofmann and Friedrichs 2002; Lynch et al. 2009). Coupled biogeochemical–physical models have been used from coastal regions (e.g., Anderson et al. 2005; Spitz et al. 2005; Ji et al. 2008; Stow et al. 2009) to basins and global ocean domains (e.g., Oschlies and Garcon 1998; Rothstein et al. 2006; Doney et al. 2009). However, in light of the strong nonlinearities observed in biological processes, an important subject that has been largely overlooked is the numerical requirements for such simulation studies. One of the major objectives of our work is to address such computational questions for reactive ocean tracers, directly including the latest advances in computational fluid dynamics (e.g., Chung 2002; Ferziger and Peric 2002; Lomax et al. 2003; Cebeci et al. 2005; Karniadakis and Sherwin 2005) and multiscale ocean modeling (Deleersnijder and Lermusiaux 2008).

Previous numerical ocean studies related to ours have primarily focused on passive or dynamic (density-related) tracer advections. The most significant progress include the results of Hecht et al. (1995), Hanert et al. (2004), and Budgell et al. (2007), but none of these advances has dealt with higher-order advection of reactive tracers on unstructured meshes with curved geometries. Iskandarani et al. (2005) applied and studied high-order schemes for passive tracer and density dynamics in two dimensions, including Hecht et al. (1995)'s test and the gravitational adjustment of density in a channel of constant depth (Haidvogel and Beckmann 1999), but they did not consider curved elements. Lévy et al. (2001) assessed five different low-order finite volume advection schemes for biological modeling and found a 30% difference in new production estimates, highlighting the need for careful numerical studies. In Bernard et al. (2009), high-order discontinuous Galerkin (DG) methods are used to solve tidal flows around shallow water islands with non-trivial geometries and using curved triangular meshes. Here, we are interested in biogeochemical tracers with possibly highly nonlinear reactive or source terms, and we compare a set of low- to high-order schemes, both in time and in space. We employ the DG finite element method (Cockburn 1998), using both straight and curved elements, and we study a varied set of numerical properties. As in previous computational studies, we restrict our numerical analyses to 2D flows, focusing on coupled dynamics in idealized straits.

Our ultimate dynamics motivation is to allow quantitative simulation studies of fundamental nonlinear biological–physical dynamics in coastal regions with complex bathymetric features such as straits, sills, ridges, and shelfbreaks. Such features strongly affect flows, and if they are shallow enough, one can expect biological responses in the euphotic zone. Multiple physical scales are possible, from rapid tidal effects to slow water-mass-driven overflows, and biological resonances at some of these scales are likely. Our focus is on the numerical requirements prerequisite to such studies. Our work is partly inspired by our experience in coastal regions with complex geometries (Haley and Lermusiaux 2010), especially with steep shelfbreaks such as the Massachusetts Bay and Stellwagen Bank (Besiktepe et al. 2002), Middle Atlantic Bight shelfbreak (Lermusiaux 1999), Monterey Bay shelfbreak (Haley et al. 2009), Taiwan region shelfbreak (Lermusiaux and Xu 2010), and Philippine Archipelago Straits (Haley and Lermusiaux 2010). The latter effort particularly motivated the present work, within the context of the Philippines Experiment (PhilEx) which is a 5-year joint research project focused on interdisciplinary modeling, data assimilation, and dynamical studies in the straits regions of the Philippine Archipelago to better understand, model, and predict sub-mesoscale and mesoscale physical and biogeochemical dynamics in complex regions. For realistic PhilEx simulations, we employ our MIT Multidisciplinary Simulation, Estimation, and Assimilation Systems (MSEAS-Group 2010). It includes a free surface hydrostatic ocean model over complex geometries with novel implicit schemes for telescoping nesting (Haley and Lermusiaux 2010). This physical model is coupled to biological models (Besiktepe et al. 2002), forced with multiscale barotropic tides (Logutov and Lermusiaux 2008), and initialized with new objective mapping schemes specific for multiconnected domains (Agarwal 2009; Agarwal and Lermusiaux 2010, in press). The multiresolution nested domains cover very shallow regions with strong tides, steep bathymetries, and the deep ocean. The MSEAS system was employed in real time, assimilating data sets from ships, gliders, and satellite remote

sensing and issuing daily physical–biological forecasts with dynamical descriptions and adaptive sampling guidance (Lermusiaux et al. 2009). The complex, nonlinear, and multiscale biology in the region confirmed the need for the present computational studies.

Our work is part of an incubation for the next generation of ocean modeling systems, focusing on key numerical questions for biogeochemical dynamics. The biological model we employ is based on Flierl and McGillicuddy (2002), Burton (2009), and Ueckermann (2009). We restrict ourselves to a relatively simple model to focus on the numerics. However, the model is complex enough to reveal important characteristics and to complete a large number of parameter sensitivity studies which we can synthesize. We study three biological regimes, one with single stable points at all depths and two with stable limit cycles. We examine interactions that are dominated by the biology, by the advection, or that are balanced. We also consider idealized simulations of biological patchiness which is commonly observed in the coastal ocean. For these regimes and interactions, we study a wide range of temporal and spatial discretizations. In what follows, we give our dynamical problem statement, definitions, and notation in Section 2. Our new numerical schemes and discretization are formulated and studied in Section 3. The results of our varied numerical and scientific investigations are described in Section 4. Finally, our conclusions are stated in Section 5.

2 Dynamical problem statement, definitions, and notation

2.1 Dynamical problem statement

The biological dynamics are governed by the following advection–diffusion–reaction (ADR) equations:

$$\frac{\partial \Phi}{\partial t} + \nabla \cdot (\mathbf{u}\Phi) - \kappa \nabla^2 \Phi = \mathbf{S}(\Phi, \mathbf{x}, t), \quad \text{in } \Omega \tag{1}$$

with boundary conditions

$$\Phi = \mathbf{g}_{\mathrm{D}}, \quad \text{on } \Gamma_{\mathrm{D}}$$
$$(\mathbf{u}\Phi - \kappa \nabla \Phi) \cdot \hat{\mathbf{n}} = \mathbf{g}_{\mathrm{N}}, \quad \text{on } \Gamma_{\mathrm{N}} \tag{2}$$

where $\Phi(\mathbf{x}, t) = [\phi^1(\mathbf{x}, t), \ldots, \phi^{N_c}(\mathbf{x}, t)]$ is the vector of N_c biological components, $\mathbf{u}$ is the prescribed velocity field, κ is a positive diffusivity coefficient, $\mathbf{S}(\Phi, \mathbf{x}, t)$ is the biological reaction terms, and $\mathbf{g}_{\mathrm{D}}$, $\mathbf{g}_{\mathrm{N}}$ are the boundary conditions for the Dirichlet and Neumann boundaries, respectively. Equations 1 and 2 are solved on the domain $\Omega \in \mathbb{R}^d$, where d is the dimension of the problem, with boundary $\partial\Omega = \Gamma_{\mathrm{D}} \cup \Gamma_{\mathrm{N}}$ such that $\Gamma_{\mathrm{D}} \cap \Gamma_{\mathrm{N}} = \emptyset$.

Since we are interested in strait dynamics, for the flowfield $\mathbf{u}$, we assume that earth rotational effects are negligible, which is true if the ratio of the strait width to the Rossby radius is small (Pedlosky 1987; Signell 1989; Cushman-Roisin 1987; Bourgault and Kelley 2004). Additionally, for uniform geometry across the strait with a rigid lid approximation, a small Froude number, and a homogeneous density, the velocity field can be approximated as a potential flow field. A similar setup was used by Signell (1989) for tidal flows. The potential velocity $\mathbf{u}$ is obtained by solving for the stream function

$$\nabla^2 \psi = 0, \quad \text{in } \Omega \tag{3}$$

$$\mathbf{u} = \nabla \times \psi \tag{4}$$

with boundary conditions

$$\psi = h_{\mathrm{D}}, \quad \text{on } \partial\Omega. \tag{5}$$

2.2 Finite element definitions and notation

We discretize Eqs. 1–5 using the DG finite element method. The first reported use of DG FEM was by Reed and Hill (1973) where DG was used to solve the steady-state neutron transport equation. However, DG drew little attention until a series of papers (Cockburn and Shu 1989, 1998b; Cockburn et al. 1989, 1990), where the Runge–Kutta DG methods were described. The extension of DG to higher-order derivatives by Bassi and Rebay (1997) made the method applicable to solving advection–diffusion equations, which can be extended to solving the Navier–Stokes equations. Since the late 1990s, DG has seen a number of realistic applications in aerospace, solid mechanics, and electromagnetism to name a few. For a review on the use of such schemes in next-generation physical ocean models, we refer to Pain et al. (2005), Slingo et al. (2009), and Ueckermann (2009).

In this section, we first describe the notation used for the domain and the discrete elements. Then we define the notation used for the solution on the element interior and on the element interfaces. Next, we described the discontinuous polynomial spaces and the necessary inner products for the DG discretization, followed by the discontinuous finite element space and inner product for the hybridized discontinuous Galerkin (HDG) method. A set of terms is also defined.

The basic domain notation is illustrated in Fig. 1. We let $\mathcal{T}_h = \cup K_i$ be a finite collection of nonoverlapping elements, K_i, that discretizes the domain Ω, where h denotes the characteristic size of an element. Also, let $\partial\mathcal{T}_h = \{\partial K : K \in \mathcal{T}_h\}$ be the set of interfaces of all

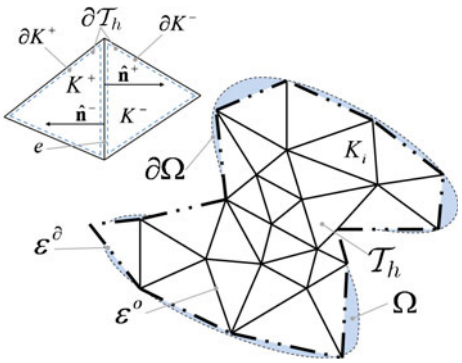

Fig. 1 Notation for domain and triangular elements

elements. For two elements K^+ and K^- belonging to $\mathcal{T}_h$, we define $e = \partial K^+ \cap \partial K^- \neq \emptyset$ as the unique interior interface between elements K^+ and K^-. For a single element K belonging to $\mathcal{T}_h$, $e = \partial K \cap \partial\Omega \neq \emptyset$ is a boundary interface. Let ε_h° and ε_h^∂ denote the set of unique interior and boundary interfaces, respectively, such that $\varepsilon_h = \varepsilon_h^\partial \cup \varepsilon_h^\circ$. We note that in the interior $\partial\mathcal{T}_h$ contains two interfaces, ∂K^+ and ∂K^-, at the same location (one for each element), whereas the set ε_h only contains a single interface, e, at the same location.

K^+ and K^- have outward pointing normals $\hat{\mathbf{n}}^+$ and $\hat{\mathbf{n}}^-$, respectively. We then let vector and scalar quantities $(\mathbf{q}^\pm, u^\pm)$ be the traces of $(\mathbf{q}, u)$ on the interface e from the interior of $K^\pm$. The mean value $\{\!\{.\}\!\}$ and jumps $[\![.]\!]$ on the interior interface $e \in \varepsilon_h^\circ$ for scalar and vector quantities are then defined as

$$\begin{aligned} \{\!\{\mathbf{q}\}\!\} &= (\mathbf{q}^+ + \mathbf{q}^-)/2 \qquad \{\!\{u\}\!\} = (u^+ + u^-)/2 \\ [\![\mathbf{q}\cdot\hat{\mathbf{n}}]\!] &= \mathbf{q}^+\cdot\hat{\mathbf{n}}^+ + \mathbf{q}^-\cdot\hat{\mathbf{n}}^- \quad [\![u\hat{\mathbf{n}}]\!] = u^+\hat{\mathbf{n}}^+ + u^-\hat{\mathbf{n}}^-. \end{aligned}$$

On the set of boundary interfaces $e \in \varepsilon_h^\partial$ (with outward facing normal $\hat{\mathbf{n}}$ on $\partial\Omega$), we set

$$\begin{aligned} \{\!\{\mathbf{q}\}\!\} &= \mathbf{q} \qquad \{\!\{u\}\!\} = u \\ [\![\mathbf{q}\cdot\hat{\mathbf{n}}]\!] &= \mathbf{q}\cdot\hat{\mathbf{n}} \quad [\![u\hat{\mathbf{n}}]\!] = u\hat{\mathbf{n}}. \end{aligned}$$

since here $\mathbf{q}$ and u are single-valued. Note that the jump in a vector is a scalar (involving only the normal component of the vector), whereas the jump in a scalar is a vector. Additionally, the jump will be zero for a continuous function.

The main difference between continuous Galerkin (CG) and DG lies in the approximation subspaces used. DG uses bases that are in normed space $L^2(\Omega)$ while CG uses bases that are in the Hilbert space $H^1(\Omega)$, that is, the function has to be continuous across elements. For a function $f(\mathbf{x})$ to be in $L^2(\Omega)$, it has to satisfy $\int_\Omega f(\mathbf{x})^2 d\Omega < \infty$, whereas a function in $H^1(\Omega)$ has to belong to a smaller space satisfying $\int_\Omega f(\mathbf{x})^2 + \nabla f(\mathbf{x}) \times \nabla f(\mathbf{x}) d\Omega < \infty$. Let $\mathcal{P}^p(D)$ denote the set of polynomials of maximum degree p existing on a domain D. For example, we will be using $p2$ to denote a second degree polynomial basis, which will result in a third-order accurate scheme. We introduce the discontinuous finite element spaces defined as

$$\begin{aligned} W_h^p &= \left\{ w \in L^2(\Omega) : w\,|_K \in \mathcal{P}^p(K), \forall K \in \mathcal{T}_h \right\} \\ \mathbf{V}_h^p &= \left\{ \mathbf{v} \in (L^2(\Omega))^d : \mathbf{v}\,|_K \in (\mathcal{P}^p(K))^d, \forall K \in \mathcal{T}_h \right\} \end{aligned}$$

where W_h^p is a scalar space, $\mathbf{V}_h^p$ is a vector space of dimension d, and $L^2(D)$ is the space of square integrable functions $f(\mathbf{x})$ such that $\int_D f(\mathbf{x})^2 dD < \infty$ on domain D.

Finally, we define the inner products over continuous domains $D \in \mathbb{R}^d$ and $\partial D \in \mathbb{R}^{d-1}$ as

$$\begin{aligned} (\mathbf{q}, \mathbf{v})_D &= \textstyle\int_D \mathbf{q}\cdot\mathbf{v} dD \qquad (u, w)_D = \int_D uw dD \\ \langle \mathbf{q}, \mathbf{v}\rangle_{\partial D} &= \textstyle\int_{\partial D} \mathbf{q}\cdot\mathbf{v} d\partial D \quad \langle u, w\rangle_{\partial D} = \int_{\partial D} uw d\partial D \end{aligned} \tag{6}$$

for vector functions $\mathbf{q}$, $\mathbf{v}$ and scalar functions u, w. Over discontinuous domains, we also define

$$(q, v)_{\mathcal{T}_h} = \sum_{K\in\mathcal{T}_h} (\mathbf{q}, \mathbf{v})_K, \quad \langle u, w\rangle_{\partial\mathcal{T}_h} = \sum_{K\in\mathcal{T}_h} \langle u, w\rangle_{\partial K}, \tag{7}$$

for vector or scalar functions q, v defined on $\mathcal{T}_h$, and u, v defined on $\partial\mathcal{T}_h$.

To use the HDG framework for solving Eqs. 3–5, we will require the traced finite element space existing on the interfaces ε_h

$$M_h^p = \left\{ \mu \in L^2(\varepsilon_h) : \mu|_e \in \mathcal{P}^p(e), \forall e \in \varepsilon_h \right\}.$$

We also set $M_h^p(g_D) = \{\mu \in M_h^p : \mu = \mathsf{P} g_D \text{ on } \Gamma_D\}$, where P is the L^2 projection into the space $\{\mu|_{\partial\Omega} \forall \mu \in M_h^p\}$. Note that M_h^p is continuous on the interface, e, shared by K^+ and K^-, but discontinuous at the borders between different interfaces. We will also require the additional inner product on this discontinuous domain

$$\langle \mu, \eta\rangle_{\varepsilon_h} = \sum_{e\in\varepsilon_h} \langle \mu, \eta\rangle_e \tag{8}$$

for vector or scalar functions μ, η defined on ε_h.

2.3 Comparing numerical codes: defining efficiency, accuracy, and performance

To be clear, we use the term "efficiency" or "cost" to refer exclusively to the computational resources (elapsed time, memory) required for a simulation, and we do not use "efficiency" to imply any degree of correctness of the solution. We reserve the term "accuracy" to refer to the correctness of the solution. Finally, here we also use the term "performance" as the combined consideration

between efficiency and accuracy (Chapra and Canale 2006).

Comparing different numerical schemes is not a straightforward task (see Kubatko et al. 2009). First, results are not universally applicable and are generally problem dependent. After focusing on a particular class of problems, the usual approach is to fix the computational efficiency of both schemes and then compare the accuracy, or vice versa. The scheme that performs better will then have a superior accuracy, since the efficiency will be the same for both. However, the efficiency of the scheme is dependent on its implementation, as well as the computer architecture on which the simulations are performed. A simple approach, then, is to fix the number of degrees of freedom (DOFs) of the different schemes, that is, having the same number of unconstrained parameters in both schemes. Because the DOFs are related to computational efficiency, this approach is useful for comparing similar numerical schemes with different implementations. However, it is not a good approach when comparing different numerical schemes where the computational cost per DOF is inherently and significantly different between the schemes, which is the case for comparisons between high-order and low-order schemes. Finally, conclusions drawn about the performance is also dependent on the particular definition of accuracy. The accuracy is normally defined in terms of a quantity useful to a particular researcher. Thus, researchers with different quantities of interest may draw different conclusions about the performance of a scheme. We address the efficiency issue by presenting results for multiple efficiencies, and we address the accuracy issue by using generic global error measures (see Section 3.5) and by using difference plots.

3 Numerical methodology

In this section, we first derive the basic FE formulations for Eqs. 1–5, using the notation from Section 2.2. Starting with Eqs. 1 and 2, we multiply each biological component by test function w and integrate over the domain. We seek approximations $\Phi_h = [\phi_h^1 \dots \phi_h^{N_c}]$ of Φ such that for all $K \in \mathcal{T}_h$,

$$\left(\frac{\partial \Phi_h}{\partial t}, w\right)_K + (\nabla \times (\mathbf{u}\Phi_h), w)_K - \left(\kappa \nabla^2 \Phi_h, w\right)_K = (\mathbf{S}(\Phi_h, \mathbf{x}, t), w)_K, \forall w \in \mathcal{P}(K). \quad (9)$$

where each component $\phi_h^i \in W_h^p$. We set $\phi_h^i = \sum_{j=1}^{N_P} \phi_j^i(t)\theta_j(\mathbf{x})$ where $\phi_j^i(t)$ are N_P time varying coefficients with N_P corresponding spatial basis functions, $\theta_j(\mathbf{x}) \in \mathcal{P}(K)$. For convenience, we use Einstein summation notation $\Phi_h = \Phi_j\theta_j$, where the sum over the repeated index j is implied and we let $\Phi_j \equiv \Phi_j(t)$, $\theta_j \equiv \theta_j(\mathbf{x})$.

To obtain our finite-element formulation of Eqs. 3–5, we multiply by test functions w and $\mathbf{v}$ and integrate over the domain. We seek approximations $\psi_h \in W_h^p$ of ψ and $\mathbf{u}_h \in \mathbf{V}_h^p$ of $\mathbf{u}$ such that for all $K \in \mathcal{T}_h$,

$$\left(\nabla^2 \psi_h, w\right)_K = 0, \qquad \forall w \in \mathcal{P}(K), \quad (10)$$

$$(\mathbf{u}_h, \mathbf{v})_K = (\nabla \times \psi_h, \mathbf{v})_K, \quad \forall \mathbf{v} \in (\mathcal{P}(K))^d \quad (11)$$

Next, we describe in detail the discretization of the source terms in Section 3.1, since one of the novel aspects of our work is the high-order schemes for ocean biogeochemical simulations. Our spatial discretizations are derived in Section 3.2 for the advection terms, Section 3.3 for the diffusive terms, and our temporal discretizations given in Section 3.4. The calculation of error norms is outlined in Section 3.5. Finally, brief details of our implementation are presented in Section 3.6, and our method for generating high-order curved meshes is described in Section 3.7.

3.1 Source term discretization

Quadrature-based integration approximates a definite integral by a weighted sum of function evaluations at discrete points (quadrature points). For example, Gaussian quadrature using N_g points can exactly evaluate polynomials of degree $2N_g - 1$. Quadrature-free methods avoid evaluating the weighted sum by using approximations and/or analytically evaluating definite integrals, and the solution to the definite integral is used directly in numerical implementations. For more details, see for example, Hesthaven and Warburton (2008). We further restrict the definition of "quadrature-free" to mean exact integration on the original polynomial basis. It is possible to use an expanded basis to evaluate the source terms, but additional cost is involved to interpolate the solution unto the higher degree basis. For the source term discretization, either approach could be used, and we will examine the impact of this choice.

To discretize the source terms with a quadrature-based approach, consider the source term for a single biological component i:

$$\begin{aligned}\left(S^i(\Phi_h, \mathbf{x}_i, t), w_k\right)_K &= \left(S^i(\Phi_j\theta_j, \mathbf{x}, t), w_k\right)_K \\ &\approx S^i\left(\Phi_j\theta_j(\mathbf{x}_i), \mathbf{x}_i, t\right) w_k(\mathbf{x}_i)\omega_i \\ &= \mathcal{W}_{ki}\left[S^i\left(\Phi_j\Theta_{ji}, \mathbf{x}_i, t\right)\omega_i J_i\right] \end{aligned} \quad (12)$$

where $\mathbf{x}_i$ are the locations of the N_g quadrature points with corresponding weights ω_i, $\mathcal{W}_{ki} = w_k(\mathbf{x}_i)$ are the values of the test functions evaluated at the quadrature points, and $\Theta_{ji} = \theta_j(\mathbf{x}_i)$ are the values of the basis functions evaluated at the quadrature points. To evaluate this integral numerically, we pre-compute the matrix $\mathcal{W} \in \mathbb{R}^{N_P \times N_g}$. Because a reference element is used, the corresponding Jacobians at the quadrature points J_i also need to be calculated, and these are multiplied together with the weights (ω). The integration over an element is performed as a matrix–vector multiplication with $\mathcal{O}(2N_gN_P)$ operations for a single biological component. In addition, we need to consider the interpolation of the values of Φ_h unto the quadrature points, $\Theta_{ji}\Phi_j$, resulting in an additional $\mathcal{O}(2N_gN_P)$ for each biological component. Finally, we have to evaluate the function describing the source terms at the quadrature points. The quadrature-based algorithm thus has a total of $\mathcal{O}(4N_cN_gN_P) + N_gC_s$ operations per element, where C_s is the cost of evaluating the source terms.

Alternatively, to discretize the source terms with a quadrature-free approach using the original basis, we use

$$\begin{aligned}\left(S^i(\Phi_h, \mathbf{x}, t), w_k\right)_K &\approx \left(S^i\left(\Phi_j, \mathbf{x}_j, t\right)\theta_j, w_k\right)_K \\ &= \left(\theta_j, w_k\right)_K S^i\left(\Phi_j, \mathbf{x}_j, t\right) \\ &= \mathcal{M}_{kj} S^i\left(\Phi_j, \mathbf{x}_j, t\right) \end{aligned} \tag{13}$$

where $\mathbf{x}_j$ are the locations of the N_P nodal points and we can pre-compute the element local mass matrix $\mathcal{M}_{kj} = (\theta_j, w_k)_K$ such that $\mathcal{M} \in \mathbb{R}^{N_P \times N_P}$. Note that, for straight-sided elements, the mass matrix computed on the reference matrix can be used and multiplied by the element local scalar Jacobian. However, on elements with curved boundaries, a mass matrix using the element local spatially variable Jacobian is computed. In the quadrature-free case, we approximate the integral by essentially fitting the function $S^i(\Phi, \mathbf{x}, t)$ on the polynomial space spanned by θ. For a nodal basis, the coefficients of the basis are the values of the source terms evaluated at discrete points. Since these source term values at nodal points do not depend continuously on space, the second equality in Eq. 13 follows. This approach introduces an aliasing error since $S^i(\Phi, \mathbf{x}, t)$ may contain complicated functions which are not captured in span$\{\theta\}$. The operation count for this approach scales as $\mathcal{O}(2N_cN_pN_p) + N_pC_s$ if we need to multiply through by the mass matrix. If the number of quadrature points are equal to the number of basis functions, $N_g = N_p$, and the cost of evaluating the source terms is small, then the quadrature-free algorithm is two times more efficient in terms of total operations.

However, we can normally eliminate the mass matrix multiplication in front of the source term since there is also a mass matrix in front of the $\frac{\partial \phi_h^i}{\partial t}$ term, $\mathcal{M}_{kj}\frac{\partial \phi_j^i}{\partial t} = \mathcal{M}_{kj} S^i\left(\Phi_j, \mathbf{x}_j, t\right) \rightarrow \frac{\partial \phi_j^i}{\partial t} = S^i\left(\Phi_j, \mathbf{x}_j, t\right)$. Note that the mass matrix cannot be eliminated for all the terms in the partial differential equation (PDE), for example, a matrix will remain in front of the discretized advection operator. However, the operation count for evaluating the source terms of the quadrature-free algorithm reduces to N_pC_s. Therefore, from an efficiency perspective, it is desirable to use a quadrature-free algorithm.

The accuracy of the quadrature-based integration is limited by the quadrature rule used and the number of quadrature points N_g. Choosing a greater number of quadrature points, any desirable accuracy can be obtained, at the cost of reduced efficiency. Conversely, the accuracy of the quadrature-free integration is limited by the order of basis function used, and potentially large errors can be introduced due to inexact integration.

The biological source terms are the origin of nonlinear and nonhomogeneous dynamics for the whole PDE. They can lead to high-accuracy discretization requirements in the other terms of the PDE in both space and time. This is discussed in the next sections (Sections 3.2–3.4).

3.2 Spatial discretization of advection operators

Integrating the advection terms $(\nabla \cdot (\mathbf{u}\Phi_h), w)_K$ by parts and using the divergence theorem, we obtain the weak form of the advection operator

$$(\nabla \cdot (\mathbf{u}\Phi_h), w)_K = -(\mathbf{u}\Phi_h, \nabla w)_K + \langle \widehat{\mathbf{u}\Phi_h} \cdot \hat{\mathbf{n}}, \nabla w \rangle_{\partial K}, \tag{14}$$

where the formulation is complete once we specify the value of the flux $\widehat{\mathbf{u}\Phi_h}$. Here we use the upwind flux

$$\widehat{\mathbf{u}\Phi_h} \cdot \hat{\mathbf{n}} = \mathbf{u} \cdot \hat{\mathbf{n}}\{\{\Phi_h\}\} - \frac{1}{2}|\mathbf{u}| \cdot [\![\Phi_h \hat{\mathbf{n}}]\!] \tag{15}$$

The same quadrature versus quadrature-free discussion is relevant to the advection terms. However, here we choose to use a quadrature-based scheme to ensure the accuracy of the advection part of the discretization. Since we are focusing on evaluating the accuracy of the source terms, we do not want the additional consideration about the accuracy of the advection operators to complicate the discussion. Note, however, that considerable efficiency can be gained for the advection

terms as well if a quadrature-free scheme is employed (Hesthaven and Warburton 2008). We then let

$$
\begin{aligned}
&-\left(\mathbf{u}\Phi_h, \nabla w_k\right)_K + \left\langle \widehat{\mathbf{u}\Phi_h}\cdot\hat{\mathbf{n}}, \nabla w_k\right\rangle_{\partial K} \\
&\quad \approx -\mathcal{D}_{ki}\left[\left(\mathbf{u}_i\Phi_j\Theta_{ji}\right)\omega_i J_i\right] \\
&\qquad + \sum_{e\in K}\mathcal{W}^e_{ki}\left[\left(\widehat{\mathbf{u}_i^e\Phi_j}\Theta^e_{ji}\right)\cdot\hat{\mathbf{n}}_i\omega^e_i J^e_i\right]
\end{aligned}
\tag{16}
$$

where $\mathcal{D}_{ki} = \nabla w_k(\mathbf{x}_i)$ and the superscript $(.)^e$ indicates that the quantity is defined on an element interface and $\mathbf{u}_i = \mathbf{u}_j\Theta_{ji}$ is the flow-field evaluated at the quadrature points.

3.3 Spatial discretization of diffusive operators

To discretize the diffusive operator $\left(\kappa\nabla^2\Phi_h, w\right)_K$, we follow the standard practice with DG and consider the canonical problem

$$-\nabla\cdot(\kappa\nabla)\phi = s.$$

Introducing the auxiliary variable $\mathbf{q} = -\kappa\nabla\phi$, this equation is re-written as two coupled first-order equations

$$\nabla\cdot\mathbf{q} = s$$

$$\mathbf{q} + \kappa\nabla\phi = 0.$$

Multiplying these equations by the appropriate test functions, integrating by parts, and applying the divergence theorem, we obtain the finite element formulation for all $K \in \mathcal{T}_h$

$$-\left(\mathbf{q}_h, \nabla w\right)_K + \left\langle\hat{\mathbf{q}}_h\cdot\hat{\mathbf{n}}, w\right\rangle_{\partial K} = \left(s_h, w\right)_K \quad \forall w \in \mathcal{P}(K) \tag{17}$$

$$
\begin{aligned}
&\left(\kappa^{-1}\mathbf{q}_h, w\right)_K - \left(\phi, \nabla\cdot\mathbf{v}\right)_K \\
&\quad + \left\langle\hat{\phi}_h, \mathbf{v}\cdot\hat{\mathbf{n}}\right\rangle_{\partial K} = 0 \qquad \forall\mathbf{v} \in (P(K))^d
\end{aligned}
\tag{18}
$$

where we have multiplied by κ^{-1} so that it does not appear as part of the flux (or interface) terms. This formulation is complete once we specify the form of the flux terms at the interfaces, $\hat{\mathbf{q}}$ and $\hat{\phi}_h$. The diffusive fluxes for DG schemes are normally reported in the form

$$\hat{\mathbf{q}}_\mathbf{h} = \{\{\mathbf{q}_h\}\} - C_{11}[\![\phi_h\hat{\mathbf{n}}]\!] + \mathbf{C}_{12}[\![\mathbf{q}_h\cdot\hat{\mathbf{n}}]\!] \tag{19}$$

$$\hat{\phi}_h = \{\{\phi_h\}\} - \mathbf{C}_{12}\cdot[\![\phi_h\hat{\mathbf{n}}]\!] - C_{22}[\![\mathbf{q}_h\cdot\hat{\mathbf{n}}]\!] \tag{20}$$

Biogeochemical diffusive fluxes Using explicit time integration to solve Eq. 1, we do not need to invert a matrix in Eq. 9, in which case we utilize the local discontinuous Galerkin fluxes (Cockburn and Shu 1998a)

$$C_{11} = \tau, \quad \mathbf{C}_{12} = \frac{\hat{\mathbf{n}}^{\pm}}{2}, \quad C_{22} = 0, \tag{21}$$

although many other choices exist.

Potential flows For solving Eq. 3, a matrix inversion is required in Eq. 10, in which case we discretize the diffusive operators using the novel HDG method. For the full derivation of these equations, including more specific implementation details, refer to Cockburn et al. (2009) and Nguyen et al. (2009). The premise of HDG is to recognize that one can solve Eqs. 17–18 locally on an element as long as the flux quantities are known. Within an HDG framework, the local element unknowns are parameterized in terms of a new variable $\lambda_h \in M_h^p(0)$, where the notation $M_h^p(0)$ refers to the space M_h^p that is zero-valued on the boundaries of the domain. The fluxes are expressed as

$$\hat{\phi}_h = \begin{cases} \mathsf{P}g_D, \text{ on } \varepsilon_h^\partial \\ \lambda_h, \text{ on } \varepsilon_h^\circ \end{cases} \tag{22}$$

$$\hat{\mathbf{q}}_h = \mathbf{q}_h + \tau(\phi_h - \hat{\phi}_h)\hat{\mathbf{n}}, \quad \text{on } \partial\mathcal{T}_h \tag{23}$$

where τ is a tune-able stabilization parameter and P is, again, the L^2 projection into the space $\{\mu|_{\partial\Omega}\forall\mu \in M_h^p\}$. Now, once λ_h is known, Eqs. 17–18 can be solved efficiently on each element independently. What remains is an equation for λ_h, which can be found by enforcing continuity of the normal diffusive flux

$$\langle[\![\hat{\mathbf{q}}\cdot\hat{\mathbf{n}}]\!], \mu\rangle_{\varepsilon_h} = \langle g_N, \mu\rangle_{\Gamma_N}. \tag{24}$$

Note that this is an equation with globally coupled unknowns. However, the number of unknowns is greatly reduced compared to the original system, since only unknowns on the interfaces ε_h are involved.

This solution method involves three steps:

1. The inversion of local operators on each element to form both the right-hand-side vector and the global matrix
2. The global solution to find λ_h
3. The local reconstruction of the solution on the element

The local operations are efficient because inversions are done on matrices which are of dimension $\mathbb{R}^{(1+d)N_p\times(1+d)N_p}$. This procedure dramatically increases the efficiency of solving elliptic problems with DG where implicit time integration is required. Additionally, when the stabilization parameter for this choice of fluxes is chosen optimally ($\tau \sim \mathcal{O}(1)$), the optimal convergence rate of $\mathcal{O}(p+1)$ is obtained for both the gradient ($\mathbf{q}$) and the solution (ϕ; Nguyen et al. 2009).

This allows a post-processing procedure that can produce a solution u_h^* which converges at $\mathcal{O}(p+2)$.

Reporting the fluxes for HDG in the standard form we have from Nguyen et al. (2009), the following

$$C_{11} = \frac{\tau^+\tau^-}{\tau^+ + \tau^-}, \quad \mathbf{C}_{12} = \frac{1}{2}\left(\frac{[\![\tau\hat{\mathbf{n}}]\!]}{\tau^+ + \tau^-}\right),$$
$$C_{22} = \frac{1}{\tau^+ + \tau^-}. \tag{25}$$

3.4 Temporal discretization

Motivated by strong nonlinearities in biogeochemical dynamics, an objective of this study is to evaluate effects of temporal discretization errors on the accuracy of numerical simulations. We investigate fourth-order, second-order, and first-order schemes in time. Specifically, for the majority of this work, we will use the following four-stage low-storage fourth-order accurate Runge–Kutta scheme for explicit time integration

$$\Phi^a = \Phi(t) + \frac{\Delta t}{4}\left.\frac{\partial \Phi}{\partial t}\right|_{\Phi(t)}$$
$$\Phi^b = \Phi(t) + \frac{\Delta t}{3}\left.\frac{\partial \Phi}{\partial t}\right|_{\Phi^a}$$
$$\Phi^c = \Phi(t) + \frac{\Delta t}{2}\left.\frac{\partial \Phi}{\partial t}\right|_{\Phi^b}$$
$$\Phi(t+\Delta t) = \Phi(t) + \Delta t\left.\frac{\partial \Phi}{\partial t}\right|_{\Phi^c}.$$

It is implemented in a four-stage for loop, where the solution at the initial time $\Phi(t)$ is saved in a temporary array and the array containing the solution is updated using the three intermediate Φ^{a-c} variables.

We evaluate the temporal discretization error by considering a second accurate explicit Runge–Kutta scheme

$$\Phi^a = \Phi(t) + \frac{\Delta t}{2}\left.\frac{\partial \Phi}{\partial t}\right|_{\Phi(t)}$$
$$\Phi(t+\Delta t) = \Phi(t) + \Delta t\left.\frac{\partial \Phi}{\partial t}\right|_{\Phi^a},$$

and the first-order-accurate explicit Euler scheme

$$\Phi(t+\Delta t) = \Phi(t) + \Delta t\left.\frac{\partial \Phi}{\partial t}\right|_{\Phi^t}.$$

In each case, $\frac{\partial \Phi}{\partial t}$ is evaluated using the right-hand-side spatial PDE.

3.5 Error norm calculation

Unless indicated otherwise, the global domain L^2 norm $\|e\|_2 = \left(\int_\Omega e^2 d\Omega\right)^{\frac{1}{2}}$ is calculated using the quadrature-based approach as described in Section 3.1. That is, the numerical solution is interpolated unto the quadrature points, the error $e = \phi_h - \phi$ is evaluated, and then multiplied by the quadrature weights and summed for an approximate integration. In some cases, we evaluate the global error using an interpolation approach (similar to quadrature-free), and this is mentioned when we do. In these cases, the error is evaluated at the nodal points, then the error is interpolated to the quadrature points, multiplied by quadrature weights, and summed. Where ambiguous, we indicate the quadrature-based error evaluation using $\|e\|_2^{\text{qp}}$ (quadrature points) and the interpolated error evaluation using $\|e\|_2^{\text{nd}}$ (nodal points).

The infinity norm $\|e\|_\infty = \max|e|$ is calculated by evaluating the error at nodal points and taking the maximum absolute value.

3.6 Implementation

The discretized equations were implemented for 2D triangular elements. Even though our codes are efficient, our implementation is not fully optimized, in particular, the higher-order simulations would most benefit from further optimization (e.g., see Lambrechts et al. 2010). This is fine for our purposes since if we find that our high-order implementations are more accurate for the same cost/efficiency than lower-order schemes, then further optimization would only accentuate this result.

The correctness of our implementation is verified by performing convergence studies using analytical test cases on curved and straight geometries. The implementation of each discretized operator is verified both separately and collectively. Results of some of the convergence studies are shown in Sections 4.1 and 4.2.

For the polynomial spaces restricted on each element $\mathcal{P}^p(K)$, a nodal basis with $N_p = \frac{(p+1)(p+2)}{2}$ nodal points is used in two dimensions and $N_p = (p+1)$ nodal points in one dimension. The total number of DOFs can then be calculated by multiplying the number of nodal points by the number of elements. The node locations are chosen according the method described in Hesthaven and Warburton (2008). More specifically, we have the basis θ_j such that

$$\theta_j(\mathbf{x}_i) = \delta_{ij}$$

where $\mathbf{x}_i$ are the nodal locations. This basis is constructed for the reference element [0, 0], [1, 0], [0, 1], and an isoparametric coordinate mapping is used for

arbitrary triangles. The isoparametric coordinate mapping is described through the element nodal locations in the problem reference frame.

To integrate polynomials of degree $p \geq 10$, cubature rules (multidimensional quadrature rules) are constructed on triangles by using tensor products of 1D Gauss quadrature rules. For $p < 10$, we use tabulated rules from Solin et al. (2003) and Strang and Fix (1973).

In all cases, we utilize the Galerkin approach, that is, we choose the test functions to be the same as the basis functions, $w_j = \theta_j$.

3.7 Higher-order mesh generation

Since higher-order DG schemes have more degrees of freedom per element, a coarse mesh with large elements is required to keep a similar performance across discretizations. To obtain an accurate solution with a coarse, high-order discretization, it is necessary to use curved boundary interfaces, as will be demonstrated in Section 4.3. Here we describe our new method for creating such a coarse, high-order curved mesh.

When curving the boundary of an element, care needs to be taken because it is possible to create an element where two of the interfaces cross. The left triangular element shown in Fig. 2 has the true circular geometry crossing one of the straight interfaces. To avoid this situation, we need to ensure that

$$h < 2\rho(x)\sin(\theta), \tag{26}$$

where h is the length of the element side bordering the boundary, $\rho(x) = \frac{|1+f'(x)|^{3/2}}{|f''(x)|}$ is the radius of curvature of the boundary described by $f(x)$, and θ is the minimum angle of the two angles on the edge bordering the boundary. The element shown on the right side of Fig. 2 illustrates this limiting case for an equilateral triangle, but our condition (26) is trivially extended to arbitrary triangles as shown by the dashed lines.

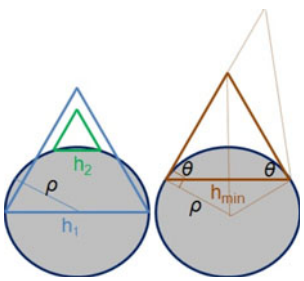

Fig. 2 Minimum triangle angle criterion (26) demonstrated on a circle with equilateral triangles. $h_1 = 2\rho$ does not satisfy the criterion, $h_2 = 2/3\rho$ satisfies the criterion, and $h_3 = 2\rho\sin(\pi/3)$ demonstrates the limiting case. This result can be extended to arbitrary triangles as shown by the *dashed lines*

Using our criteria (Eq. 26), we define the minimum edge spacing on the boundary as $h_{\min} = 2\rho\sin(30°) = \rho$. Then, we let the minimum edge length grow linearly by a certain percentage (fit to 12% here) away from the boundary up to a specified minimum edge length. Using these criteria, we create coarse base meshes, then uniformly refine these meshes to obtain finer discretizations. To create the meshes, we primarily used the free mesher *Distmesh* (Persson and Strang 2004), but we also used *Gmsh* (Geuzaine and Remacle 2009). *Distmesh* uses an implicit geometry representation, that is, we define the geometry by a distance function that gives the distance between a queried point and the nearest boundary. Using *Distmesh*, we create meshes with straight sides.

To curve the boundary interfaces, we use the same distance function provided to *Distmesh* and numerically calculate the gradient of the distance function to the boundary. The normalized gradient vector provides the direction of translation, but to determine the magnitude of the translation, a weight needs to be applied to the calculated distance. That is, $p_1^{new} = p_1^{old} + \mathcal{W} d \frac{\nabla d}{|\nabla d|}$, where d is the distance from the point p_1^{old} to the boundary and $\mathcal{W}$ is the weight. Now, points on the straight boundary interface are translated to the true, curved boundary with a weight 1, and points on interior interfaces are not translated, i.e., having weights 0. Points in the volume have weights defined by the same weighting functions used to create the nodal basis, that is

$$\mathcal{W}(e_1) = \left(\frac{2\lambda_3}{2\lambda_3 + \lambda_1}\right)\left(\frac{2\lambda_2}{2\lambda_2 + \lambda_1}\right)$$

where the point is defined by the barycentric coordinates λ_i corresponding to vertices i and e_1 is the curved boundary interface defined by vertices 2 and 3. For details of this blending function, see Hesthaven and Warburton (2008).

The base mesh with three mesh refinements is shown in Fig. 3, and details of the base mesh for a curved and straight mesh boundary are shown in Fig. 4. Using our criterion $h_{\min} = \rho$, the minimum theoretical edge length for our geometry, that is a Gaussian bump defined by $\tilde{H}(x) = e^{-x^2}$, is $h_{\min} = 0.25$. The mesh shown in Fig. 3a has a minimum edge length of $h_{\min} = 0.2418$, close to the theoretical value.

4 Numerical studies and scientific implications

Biogeochemical models may contain a large number of biological or chemical components (Hofmann and

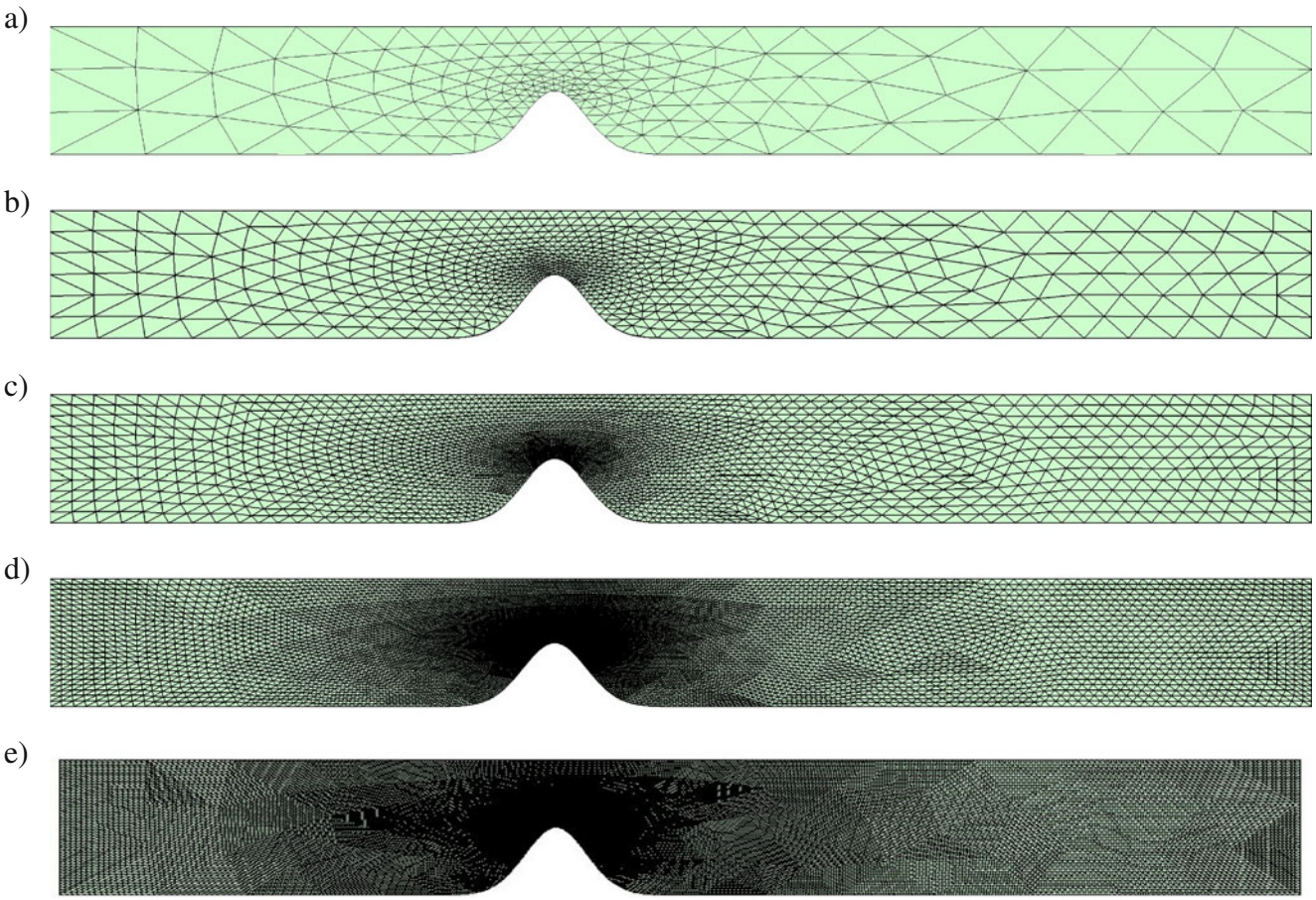

Fig. 3 **a** The base mesh ($g1$) with 350 elements. **b** First ($g2$) (1,400 elements). **c** Second ($g3$) (5,600 elements). **d** Third ($g4$) (22,400 elements) and fourth ($g5$) (89,600 elements) grid refinements. The more-refined meshes are used for lower-order schemes whereas less-refined meshes are used for higher-order schemes such that the cost of the two schemes are comparable

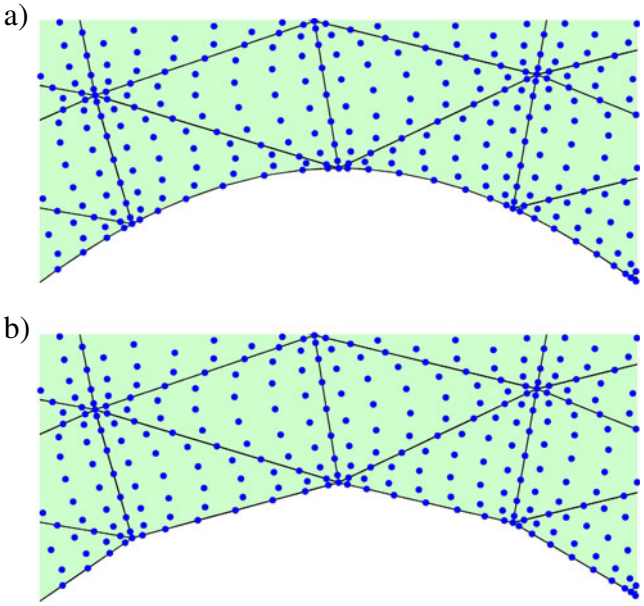

Fig. 4 Details of ($g1$) using a **a** curved and **b** straight mesh for a $p = 8$ nodal basis

Friedrichs 2002). The simplest models often only use nutrient, phytoplankton, and zooplankton as components and are commonly called NPZ models. More complicated models (Besiktepe et al. 2002) can be adaptive and contain many components. Each component requires the solution of an ADR equation of the form 1. The source terms describe the commonly nonlinear "reactions" and may lead to stationary, periodic, or chaotic dynamics. For this numerical work, a nondimensional version of a NPZ model (Flierl and McGillicuddy 2002) is used since it contains all characteristics required for our studies:

$$\frac{\partial \phi_N^*}{\partial t^*} + \nabla \times (\mathbf{u}^* \phi_N^*) - \nabla \times \frac{1}{P_e} \nabla \phi_N^* = -\mathcal{U}^* e^{z^*/h^*} \frac{\phi_P^* \phi_N^*}{\phi_N^* + k_s^*} + d_P^* \phi_P^* + d_Z^* \phi_Z^* + (1-a) g_v^* \phi_Z^* (1 - e^{-\nu^* \phi_P^*}) \tag{27}$$

$$\frac{\partial \phi_P^*}{\partial t^*} + \nabla \times \left(\mathbf{u}^* \phi_P^*\right) - \nabla \times \frac{1}{P_e} \nabla \phi_P^* = \mathcal{U}^* e^{z^*/h^*} \frac{\phi_P^* \phi_N^*}{\phi_N^* + k_s^*} - d_P^* \phi_P^* - g_\nu^* \phi_Z^* \left(1 - e^{-\nu^* \phi_P^*}\right) \quad (28)$$

$$\frac{\partial \phi_Z^*}{\partial t^*} + \nabla \times \left(\mathbf{u}^* \phi_Z^*\right) - \nabla \times \frac{1}{P_e} \nabla \phi_Z^* = -d_Z^* \phi_Z^* + a g_\nu^* \phi_Z^* \left(1 - e^{-\nu^* \phi_P^*}\right) \quad (29)$$

where $\phi^*_{(N,P,Z)} = \frac{\phi_{(N,P,Z)}}{\mathcal{N}_t}$; $u_x^* = \frac{u_x}{\bar{\mathbf{u}}}$; $u_z^* = \frac{u_z L}{\bar{\mathbf{u}} H}$; $x^* = \frac{x}{L}$; $z^* = \frac{z}{H}$; $t^* = \frac{t}{\bar{\tau}}$; the parameters are explained in Table 1; the nondimensional groups with values are given in Table 2 with P_e the Peclet number and D^* the aspect ratio; the subscripts $(.)_N, (.)_P, (.)_Z$ refer to nutrients, phytoplankton, and zooplankton, respectively; $\nabla = \frac{\partial}{\partial x^*} + \frac{\partial}{\partial z^*}$; and lowercase z^* refers to the depth coordinate which is positive upward with $z^* = 0$ at the surface. Note that not all three equations (Eqs. 27–29) are required since the biological model satisfies the following conservation law for total nutrients, assuming a closed ocean system:

$$\phi_N^* = 1 - \phi_P^* - \phi_Z^*. \quad (30)$$

The first equation (Eq. 27), for example, could be eliminated in favor of Eq. 30; however, here we still use Eq. 30 to check the conservation of the numerical schemes.

The domain setup is depicted in Fig. 5 for the geometric parameter values given in Table 2. With this setup, an upwelling of nutrients is created (see Section 4.4), and the study of idealized biological blooms, which may occur in straits or sills, can be studied. In total, we consider three sets of parameter values, differing by the nondimensional parameters ν^* and k_s^*. In the absence of advection and diffusion, they lead to Eq. 29 with at most one physically relevant steady-state solution (Burton 2009). The three sets of nondimensional parameters ν^* and k_s^* correspond to biological dynamics with single stable points at all depths (bio case 1: $k_s^* = 1/30$, $\nu^* = 0.3$), with stable limit cycles for depths around $z^* = 0.4 - 0.9$ and single stable points elsewhere (bio case 2: $k_s^* = 1/50$, $\nu^* = 0.5$) and stable limit cycles everywhere in the euphotic zone (bio case 3: $k_s^* = 1/100$, $\nu^* = 0.1$). The middle parameter values, bio case 2, correspond to those values used by Flierl and McGillicuddy (2002): They are idealized and not meant to represent a specific ocean region. We note that biological models with discontinuities in stable solutions are not always representative of nature. However, biology of interest is likely to have intrinsic oscillatory or chaotic time dependence, e.g., Flierl and McGillicuddy (2002). For our purposes, we address these issues by considering three sets of parameter values and so cover a range of biological dynamics. To handle nonphysical negative concentrations due to numerics, we use

Table 1 NPZ equation parameter descriptions and units

Parameter	Description	Units
$\mathcal{U}$	Phytoplankton uptake rate	1/day
k_s	Saturation concentration of phytoplankton	μmol/L
d_P	Mortality rate of phytoplankton	1/day
d_Z	Mortality rate of zooplankton	1/day
g	Grazing rate of zooplankton	L/(μmol day
a	Assimilation (efficiency) rate	
h	e-folding depth for light (photosynthesis)	m
ν	Parameter for Ivlev form of grazing function	L/μmol
$\mathcal{N}_T$	Total biomass	μmol/L
$\bar{\mathbf{u}}$	Average inlet velocity	km/day
H	Height of bathymetry	m
D	Total maximum depth	m
L	Effective width of bathymetry	km
$\underline{\underline{\kappa}}$	Diffusion tensor (vertical and horizontal diffusion different)	m^2/s

Table 2 Values of the dimensionless numbers entering the NPZ equations (Eq. 29) that are used in the examples for this manuscript

Bracketed triplets of values correspond to the three bio cases [1, 2, 3]. The other values are the same for the three cases

Parameter	Value
$\mathcal{U}^* = \mathcal{U}\bar{\tau}$	7.5
$k_s^* = \frac{k_s}{\mathcal{N}_T}$	$\left[\frac{1}{30}, \frac{1}{50}, \frac{1}{100}\right]$
$d_P^* = d_P \bar{\tau}$	0.2
$d_Z^* = d_Z \bar{\tau}$	1
$g_\nu^* = \frac{g\bar{\tau}}{\nu}$	12.5
$a^* = a$	0.4
$h^* = \frac{h}{H}$	0.34
$\nu^* = \mathcal{N}_T \nu$	[0.3, 0.5, 1]
$P_e = \frac{\bar{\mathbf{u}} L}{\kappa}$	∞
$D^* = \frac{D}{H}$	2

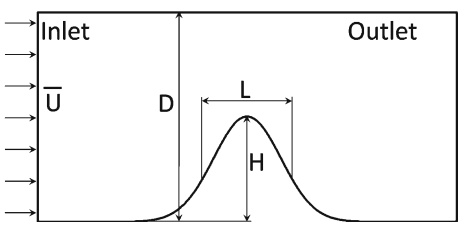

Fig. 5 Test case domain with idealized strait bottom geometry described by $\tilde{H}(x) = He^{-\frac{x^2}{L^2}}$

$\max(0, \phi^*_{(N,P,Z)})$ when evaluating the source terms. In the absence of advection, a timescale of $\bar{\tau} = 1$[days] is used, while in the presence of advection, the advective timescale $\bar{\tau}_a = L/\bar{\mathbf{u}}$ is used, where $\bar{\mathbf{u}} = \int_{\text{Inlet}} |\mathbf{u}| dz$ is the average inlet velocity.

In our numerical study, we need to characterize the three dynamical regimes and their behavior since these properties affect numerical errors. Specifically, for each dynamical regime, we study the three limiting balances of terms in Eqs. 27–29: biological terms dominating, advection terms dominating, and advection and biological terms balancing. When biological terms dominate, the advection is slow, and the problem reduces to a 1D problem, studied in Section 4.1. When advection is fast, the biology is unimportant, and we study this case in Section 4.3, with the generated flowfield studied in Section 4.2. Finally, the case where the advection and biological terms are approximately balanced is studied in Section 4.4, with the effect of biological patches demonstrated in Section 4.5. Since the timescale of biology varies in depth, the advection and biological terms can only be exactly balanced for one depth. While this results in many choices of approximately balanced terms, we will focus on one parameter set where $\bar{\tau}_a =$ 12.5 days. For more details on biological dynamics in straits, we refer to Burton (2009). Finally, capabilities of numerical filtering for higher-order schemes are examined in Section 4.6.

Results in this section will be reported for various grid and polynomial degree combinations, and the notation (grid number, polynomial degree) is used to denote this information. For example, ($g2$, $p4$) refers to the second grid (Fig. 3b) with a fourth degree basis function. We provide a table with the number of DOFs and computational time estimates for the test cases we completed Table 3. In what follows, we show the results of ($g1$, $p6$) and ($g2$, $p5$) for our high-order simulations and compare them to ($g4$, $p1$). Normalizing by the average computational time of ($g4$, $p1$), these simulations have relative computational times of 0.34, 1, and 1.2 for ($g1$, $p6$), ($g4$, $p1$), and ($g2$, $p5$), respectively (see Table 3). Following the discussion in Sections 2.3 and 3.6, we note that, in terms of efficiency, ($g4$, $p1$) and ($g2$, $p5$) are comparable (in fact, ($g2$, $p5$) would be cheaper if fully optimized, see Section 3.6). ($g1$, $p6$) is included because it is comparable in accuracy to ($g4$, $p1$) and it highlights the effect of under-resolution (here $g1$) when using higher-order schemes (here $p6$).

4.1 1D biogeochemical source terms studies

In this section, we first illustrate the convergence of our numerical implementation. Following this, we examine the numerical behavior of the biological source terms using three tests: *perturbations of steady states, vertical resolution*, and *high-order bases*.

Numerical convergence in space and time Since an analytical solution to Eq. 29 does not exist, we verify the spatial implementation of the quadrature-free and quadrature-based source terms using the analytical test problem $\frac{\partial \phi}{\partial t} - 2\frac{\partial^2 \phi}{\partial z^2} = S(z,t)$ on $\Omega \in [-100, 0]$ integrating until $t = \pi/200$, with solution $\phi = \sin(t)\cos^2(\pi/50z)$ (for an appropriately chosen $S(z,t)$). We use a sufficiently small timestep, such that the errors are dominated by the spatial discretization. The results are shown in Table 4, with the norm of the error $e = \phi_h - \phi$ calculated as described in Section 3.5. From Table 4, we note that our implementation converges at the optimal rates for both the quadrature-based and quadrature-free treatments. While the error numbers are for a special case and not those for Eqs. 27–29, they show that the solution using quadratures is more accurate than the solution without quadratures, and this result will be generally expected.

Table 3 Normalized run-times and DOFs for various grids/polynomial degree basis functions, for the simulations in Section 4.4

Degree of basis	Grid 1	Grid 2	Grid 3	Grid 4	Grid 5
1	0.0014 (1,050)	0.014 (4,200)	0.12 (16,800)	**1.0** (67,200)	8.2 (268,800)
2	0.007 (2,100)	0.057 (8,400)	0.51 (33,600)	4.2 (134,400)	
3	0.026 (3,500)	0.21 (14,000)	1.8 (56,000)		
4	0.062 (5,250)	0.51 (21,000)	4.1 (84,000)		
5	0.15 (7,350)	**1.2** (29,400)			
6	**0.34** (9,800)	3.1 (39,200)			

The times are normalized by the ($g4$, $p1$) run-time, and numbers in parentheses are the DOFs

Table 4 Spatial convergence of 1D DG solver used to evaluate the source terms using N_h elements

	Degree	$N_h = 10$		$N_h = 20$	
		$\|e\|_2$	Order	$\|e\|_2$	Order
Quadrature-based	$p = 1$	5.550e−003	1.9	1.409e−003	2.0
	$p = 2$	5.901e−004	2.9	7.491e−005	3.0
	$p = 3$	4.690e−005	3.9	3.019e−006	4.0
	$p = 4$	2.976e−006	4.9	9.673e−008	4.9
Quadrature-free	$p = 1$	1.340e−002	1.9	3.435e−003	2.0
	$p = 2$	1.068e−003	2.9	1.332e−004	3.0
	$p = 3$	7.464e−005	3.9	4.724e−006	4.0
	$p = 4$	4.282e−006	5.0	1.274e−007	5.1

The L^2 norm (see Section 3.5) of the error, $e = \phi_h - \phi$, is smaller for the quadrature-based scheme, but the order of convergence is the same for both. Order of convergence is computed in a standard way, e.g., Chapra and Canale (2006)

We verify the implementation of the fourth-order low-storage Runge–Kutta time integrator using the ordinary differential equation $\frac{\partial \phi}{\partial t} = \phi$ on $\Omega \in [-100, 0]$ integrating up to a time $t = 1$, with solution $\phi = \phi_0 e^t$. Here we choose $\phi_0 = 1$, such that the spatial discretization does not affect the error. The results are given in Table 5, from which we note that our implementation converges at the optimal rate for this low-storage Runge–Kutta scheme. While this test corresponds to exponential biological growth, as above, the error values are of course not those that would occur for Eqs. 27–29.

Perturbations of steady states The purpose of these studies is to examine how the biological dynamics behave as a result of perturbations away from the steady state. Perturbations will arise due to the forcing and dynamics and due to numerical reasons in the more complicated tests in Section 4.4, and it is important to understand how these perturbations will affect the biological dynamics. All three different biological regimes were examined in these perturbation tests. We focus on the behavior of the 1D dynamics for the time interval $t^* = [0, 250]$ because this corresponds to the residence time of the biology for the dynamics in Section 4.4. We initialize all tests using a perturbed or unperturbed exact steady state, which can be found by setting $\frac{\partial \Phi^*}{\partial t} + \nabla \cdot (\mathbf{u}^* \Phi^*) - \nabla \cdot \frac{1}{P_e} \nabla \Phi^* = 0$ in Eqs. 27–29. The steady-state solution is perturbed by setting $\Phi^*_{(P,Z)\text{perturb}} = (1 + \gamma)\Phi^*_{(P,Z)\text{steady}}$, where γ is some constant, and using Eq. 30. Where required, we impose $\phi^*_Z + \phi^*_P < 1$, by setting $\Phi^*_{(P,Z)\text{perturb}} = \frac{\Phi^*_{(P,Z)\text{steady}}}{(\phi^*_Z + \phi^*_P)_{\text{steady}}}$. This initialization is done numerically by setting the value of the numerical solution equal to the calculated solution *at the nodal points*.

First we ensure that the exact steady-state solution can be maintained, and then we initialize with a perturbation from the exact steady state, and the results are reported in Table 6. For these runs we used 100, second-order accurate linear elements ($p = 1$), which roughly corresponds to the resolution at the inlet for ($g5$, $p1$). We find that the steady solution can be maintained for all cases up to machine precision for the quadrature-free implementation when evaluating the error at the nodal points. This is because we initialized the numerical simulation using the exactly calculated steady state *at the nodal points*. Note that the quadrature-based scheme has a smaller difference than the quadrature-free version when evaluating the error at the quadrature points, except for the case with stable limit cycles in the euphotic zone (bio case 3). Because the quadrature version evaluates the source term at the quadrature points and the interpolation of the solution onto the quadrature points is not exactly at the analytical steady state, the source terms are nonzero, and the solution evolves. If the source-terms were polynomials of lower degree than the basis in the z direction, this would not happen.

Finally, Table 6 gives a rough description of the dynamical properties of the equations. Here the norm of the initial difference, $\|D_i\|_2$, should be compared to the norm of the final difference at quadrature points

Table 5 Temporal convergence of 1D DG solver using N_t timesteps (different values of N_t given only to show that the order does not vary with N_t but the absolute error of course changes)

Integration	$N_t = 16$		$N_t = 32$		$N_t = 64$	
scheme	$\|e\|_2$	Order	$\|e\|_2$	Order	$\|e\|_2$	Order
RK4	5.683e−006	3.9	3.646e−007	4.0	2.308e−008	4.0

Order is computed using Chapra and Canale (2006)

Table 6 Difference between analytical steady-state solution, and perturbed solution at $t^* = 250$

Stability	$\gamma \times 100\%$	$\|D_i\|_2^{\mathrm{qp}}$	$\|D_i\|_2^{\mathrm{nd}}$	$\|D_q\|_2^{\mathrm{qp}}$	$\|D_{\mathrm{qf}}\|_2^{\mathrm{qp}}$	$\|D_q\|_2^{\mathrm{nd}}$	$\|D_{\mathrm{qf}}\|_2^{\mathrm{nd}}$
Single stable points	0	0.361	0.000	0.296	0.361	0.117	0.000
	0.05	0.361	0.014	0.296	0.361	0.117	0.002
	0.50	0.377	0.126	0.295	0.363	0.122	0.015
	5.00	0.906	0.860	0.318	0.402	0.208	0.141
Stable limit cycles at bottom of euphotic zone	0	0.364	0.000	0.352	0.364	0.015	0.000
	0.05	0.366	0.017	0.353	0.365	0.019	0.011
	0.50	0.410	0.169	0.377	0.387	0.116	0.112
	5.00	1.360	1.29	1.01	0.974	0.945	0.915
Stable limit cycles in whole euphotic zone	0	0.109	0.000	0.112	0.109	0.038	0.000
	0.05	0.111	0.021	0.637	0.517	0.736	0.619
	0.50	0.234	0.205	1.40	1.20	1.51	1.36
	5.00	1.880	1.88	2.39	2.17	2.51	2.37

Here $D = \frac{\Phi - \Phi_h}{\int_\Omega 1 d\Omega} \times 100\%$ is the percent error per area in the domain. The column $\|D_i\|_2$ gives the initial difference, D_q indicates using quadratures, D_{qf} indicates quadrature-free, $\|.\|_2^{\mathrm{qp}}$ indicates the error evaluated at quadrature points, $\|.\|_2^{\mathrm{nd}}$ indicates the error evaluated at nodal points

for the quadrature-based treatment and at the nodal points for the quadrature-free implementation. For the case with single stable points (bio case 1), the initial difference of the perturbed solution to the analytical steady state is greater than the final difference, which indicates that the solution is approaching the calculated steady-state value. For the case with stable limit cycles in the euphotic zone (bio case 3), the final difference is greater than the initial perturbed difference, showing the solution is logically not approaching the steady state but instead spiraling outward toward the stable limit cycles present at each depth. Additionally, plotting the solution (Fig. 6) profile for the largest perturbation, we can see that the perturbed solution tends toward the steady state for the entire column for bio case 1, only for the top part of the water column for bio case 2 and nowhere for bio case 3. Thus, the parameter set with limit cycles in the euphotic zone (bio case 3) has the most structure in the vertical and will require the most resolution to model accurately. Also, numerical perturbations will be most important for bio case 3 because the differences will grow away from the calculated steady state, as opposed to decaying.

Vertical resolution By varying the resolution of the problem, we found that a minimum of 25 degrees of freedom were necessary to roughly capture the vertical structure of the biological model dynamics at the final time. For the tests in Section 4.4, $(g1, p6)$ has approximately 21 degrees of freedom, indicating that it will be under-resolved.

High-order bases We find that the quadrature-based treatment of the source terms results in large jumps of the solution between elements. This is illustrated in Fig. 7 after 500 time units of integration using a 15th degree polynomial and three elements. The problem is amplified when using a uniform nodal spacing, due to a larger interpolation error. Also, increasing the number of quadrature points used for integration did not solve this problem. The problem originates from the discontinuous jump in the solution, causing oscillations known as Gibbs phenomena. Note that both simulations are initialized in the same manner, but the Gibbs oscillations can only be "seen" when evaluating the initial condition at points other than the nodes. The quadrature-based integration, then, "sees" these oscillations because the source terms are evaluated at the quadrature points. Using the quadrature-free approach for this 1D problem essentially decouples the vertical nodes, so numerically, the quadrature-free version does not "see" the oscillations. The Gibbs oscillations would have occurred in the quadrature-free scheme if the initialization was done at the quadrature points instead. This example illustrates one of the drawbacks of using increasingly higher-order schemes, that is, without special treatment, large oscillations occur for nonsmooth functions. Using lower order but on a finer discretization (more elements) can be a better strategy if special treatment is not used. This issue is further addressed in Sections 4.4 and 4.6.

In this section, we showed that, with our implementation, both the quadrature-based and quadrature-free treatment of the source terms give accurate, convergent results (see Table 4), although the absolute error of the quadrature-based implementation is smaller than the quadrature-free implementation. Then we showed that the analytical steady-state solution could be maintained and illustrated the dynamical behavior of three different biological parameter sets through perturbations of the steady-state solutions. With the vertical

resolution tests, we found a minimum of 25 degrees of freedom necessary to roughly capture the vertical solution features of our particular setup. Finally, we showed that oscillations can occur solely due to numerics for a high-order discretization. While the quadrature-based algorithm was shown to be more accurate, the oscillations at element interfaces and the added numerical cost need to be considered. The additional accuracy may be warranted when a bifurcation of the solution could occur, or when the solution is under-resolved. As a whole, a key result is that, for any numerical scheme, careful numerical studies should be performed in one dimension to understand the potential errors arising from the nonlinear source term discretization before proceeding with advective models.

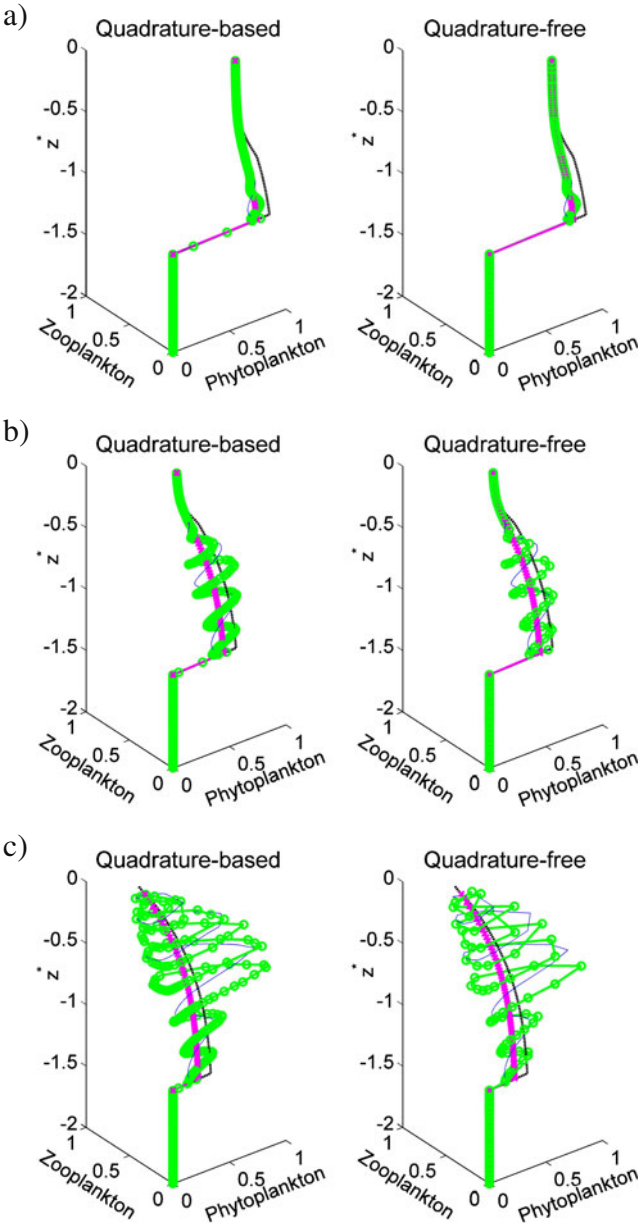

Fig. 6 Solution profiles at all depths with $\gamma = 5\%$. *Magenta crosses* show the analytical steady-state solution, the *thick black dashed lines* show the initial condition, *green circles* show the profile at $t^* = 250$, and *thin blue lines* show the profile at $t^* = 125$. Plotted for biological dynamics with **a** single stable points at all depths, **b** stable limit cycles at bottom of euphotic zone, and **c** stable limit cycles in entire euphotic zone. The quadrature-based solution is plotted at the quadrature points, whereas the quadrature-free solution is plotted at the nodal points

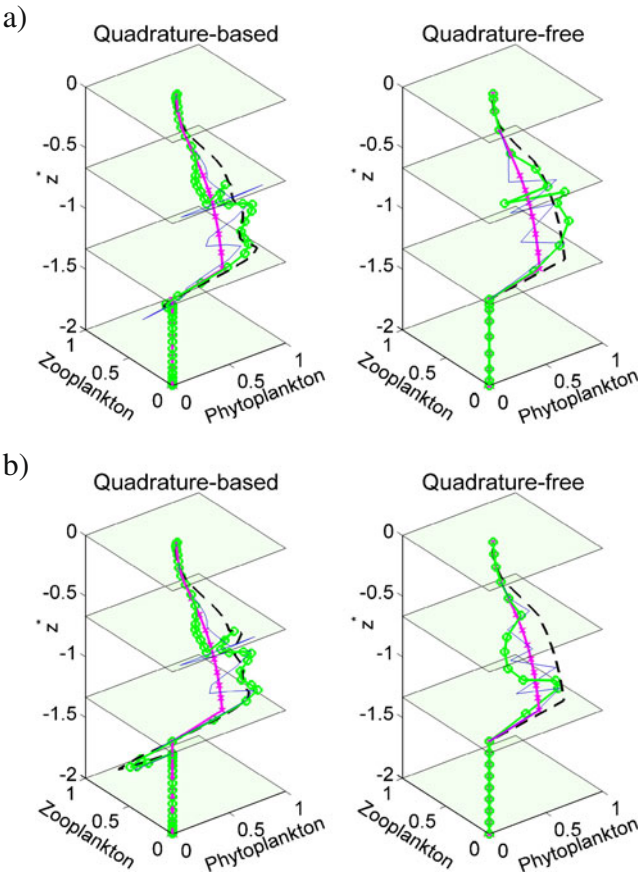

Fig. 7 Solution profiles for all depths at $t^* = 500$ using a 15th degree polynomial and three elements with $\gamma = 5\%$ for dynamics with stable limit cycles at the bottom of the euphotic zone. As in Fig. 6, the *magenta crosses* show the analytical steady-state solution, *thick black dashed lines* show the initial condition, *green circles* show the profile at $t^* = 250$, and *thin blue lines* give the profile at $t^* = 125$. The solution is plotted at the quadrature points for the quadrature version, and at the nodal points for the quadrature-free (i.e., where the source terms are evaluated). **a** Uses well-behaved (Gauss–Lobatto) nodal points, **b** uses uniform nodal points

4.2 Flow field convergence

A potential flow-field is calculated by solving Eq. 3 using HDG as described in Section 3.3. Once ψ is found, we take $\mathbf{u} = [\psi_z, -\psi_x]$. The value of ψ is specified

on all boundaries. The top and bottom boundaries are specified as constants $\psi = \psi_{\text{top}}$ and $\psi = \psi_{\text{bot}}$. The inlet and outlet stream functions are specified to vary linearly in z^*, i.e., $\psi = \psi_{\text{top}} + z^*(\psi_{\text{bot}} - \psi_{\text{top}})/D^*$. For the advection-dominating cases in Section 4.3, periodic boundaries are used, in which case the values of ψ are equal at the inlet and outlet, with no need for boundary conditions. A flowfield specific to the grid and polynomial degree is used for all simulations.

We perform a convergence study on steady flowfields to verify that we are indeed obtaining near-optimal $\mathcal{O}(p+1)$ convergence (Cockburn et al. 2009; Nguyen et al. 2009), for the gradients of ψ used for the flowfield $\mathbf{u}$. To evaluate the convergence, reference solutions using $(g5, p2)$ and $(g2, p8)$ are calculated, both giving similar results. The error was evaluated by considering the point-wise solution at all interior vertices of grid 1. That is, the solution was averaged across all element-local solutions touching the vertex and compared to the reference solutions. The point-wise error calculated using $(g2, p8)$ as the true solution is plotted in Fig. 8 for multiple grids and polynomial bases. Using this point-wise error calculation, we obtained near-optimal rate of convergence. When the domain boundary nodes are included, nonoptimal (smaller and larger) rates of convergence were found for $p > 1$. This may be due to variations in the discretizations of the domain boundary for the different grids/polynomial bases. Similarly, when the $\|.\|_2$ or $\|.\|_\infty$ norms are considered, we find near-optimal convergence for $p = 1$, but not for $p > 1$. When the domain boundary is not curved, the flowfields with $p > 1$ have both the same rate of convergence and error magnitude as $p = 1$ using any of the error metrics mentioned. This shows that curved boundaries are essential for an accurate high-order solution. Also, by using HDG, we have a velocity field defined on the same grid with the same order of accuracy as our complete solution scheme for the ADR equations.

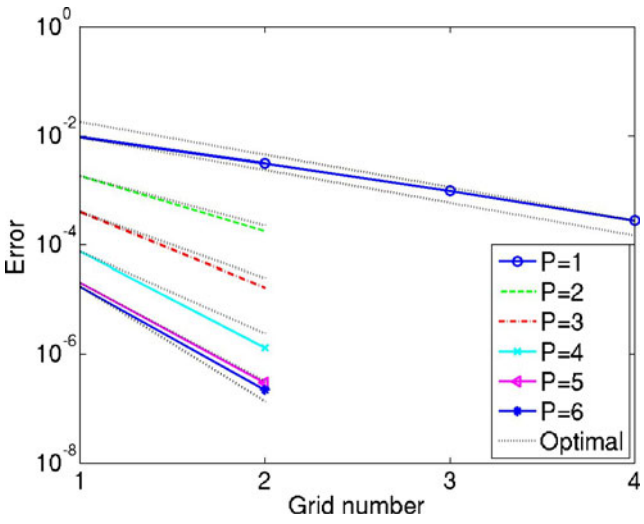

Fig. 8 Illustrating the convergence of the flow-field error. The point-wise error calculated using $(g2, p8)$ as the true solution for multiple grids and polynomial bases. Note that HDG gives near-optimal convergence for the derivative quantities $\mathbf{u} = [\psi_z, -\psi_x]$

Therefore, by using HDG with boundary-curved elements we obtain an accurate, high-order convergent potential flowfield for our geometry.

4.3 Tracer advection over bump test case results

We study tracer advection without source terms to evaluate the behavior of our numerical scheme when advection dominates. We do not illustrate here the convergence of the resolution (our results as shown in Sections 4.1 and 4.2). Instead we focus on numerical advection artifacts that may affect the behavior of the biology. We examine three cases: a uniform tracer, a tracer with a discontinuous jump in the vertical, and a tracer with a discontinuous jump in the horizontal and a linear horizontal gradient. The first case examines the divergence of the flowfield. The second case examines the behavior of the scheme in the presence of a horizontal jump, which occurs at the edge of the euphotic zone for our choices of parameters. Note that because we initialize by setting the numerical initial condition at the nodes equal to an analytical initial condition with a jump and the jump does not necessarily occur at element boundaries, the assigned numerical initial condition contains oscillations. The number of oscillations increases with the polynomial degree but is present at all orders. Also, the straightness of the interface is affected by the grid resolution. The third case examines the behavior of the scheme in the presence of a horizontal jump and horizontal gradient, which is a numerical test for frontal dynamics. Unlike the second case, the assigned initial condition for the third case does not contain oscillations because the jump occurs at element boundaries. All cases evolve in a periodic domain. In all cases, the flow is left to right, and the duration of the simulation is determined by the time the mean inlet velocity would take to travel through the domain. One of the objectives of these tests is to compare schemes at the same overall cost, for example, a lower-order scheme is used on a higher-resolution grid.

The results for the reference solution, $(g1, p5)$; low-order solution, $(g4, p1)$; and high-order solutions, $(g2, p5)$ and $(g1, p6)$ on curved meshes, are shown in

Fig. 9. In the top plot (case 1), we find that oscillations exceeding 10^{-4} are less localized for the low-order schemes than for the high-order schemes. From the middle and bottom plots (cases 2–3), we see that the higher-order schemes have larger magnitude oscillations around the jumps (as expected from the initialization). However, the sharpness of the interfaces are comparable between ($g1$, $p5$) and ($g2$, $p5$) and also between ($g4$, $p1$) and ($g1$, $p6$). Running a lower-order case ($g3$, $p1$) with 16,800 DOFs (not shown) resulted in a much more diffuse solution compared to ($g1$, $p6$), even though ($g1$, $p6$) is as under-resolved. The higher-order discretization with fewer degrees of freedom is therefore less numerically dissipative, illustrating one of the advantages of a high-order scheme.

The locations where the solution is outside of the intervals [0.9999, 1.0001], [0, 1], and [−2, 3] for cases 1, 2, and 3, respectively, are plotted in Fig. 10. From this figure, we see that the smallest magnitude and most localized errors for case 1 happen with high-order schemes using curved boundaries. The largest magnitude errors occur for high-order schemes using straight boundaries, and the least localized errors occur for low-order schemes. From case 2, we see that for both high- and low-order schemes, the initial oscillations do not remain close to the jump but spread through the domain, although the amplitude of the radiated oscillations are at least an order of magnitude smaller than the initialized oscillations. In case 3, we see that the high- and low-order schemes have similar performance, both developing numerical oscillations with the same order of magnitude around the jump. The absolute magnitude of the oscillations is approximately half the size with low order compared to high order and can be explained by the greater number of degrees of freedom with the low-order case. From these plots, we note that

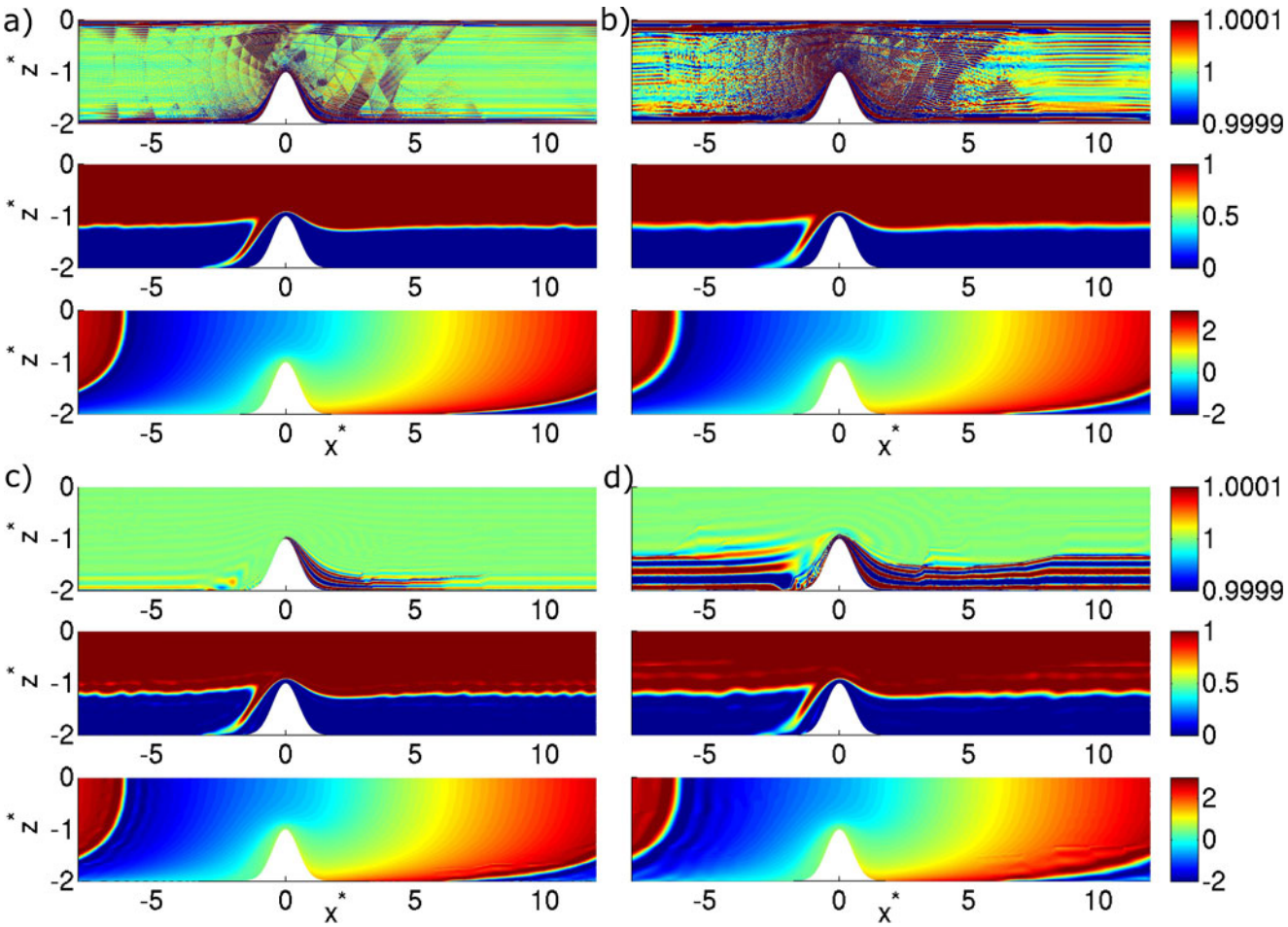

Fig. 9 Solution of three passive tracer cases, on **a** ($g5$, $p1$) with 268,800 DOFs, **b** ($g4$, $p1$) with 67,200 DOFs, **c** ($g2$, $p5$) with 29,400 DOFs, and **d** ($g1$, $p6$) with 9,800 DOFs with curved boundaries. *Top* plots show advection of a constant tracer, *middle* is advection of tracer with vertical jump, and *bottom* is advection with a horizontal jump and horizontal gradient. Note that **b** ($g4$, $p1$) and **c** ($g2$, $p5$) have similar costs while **d** ($g1$, $p6$) is much cheaper and under-resolved (see Table 3 for costs)

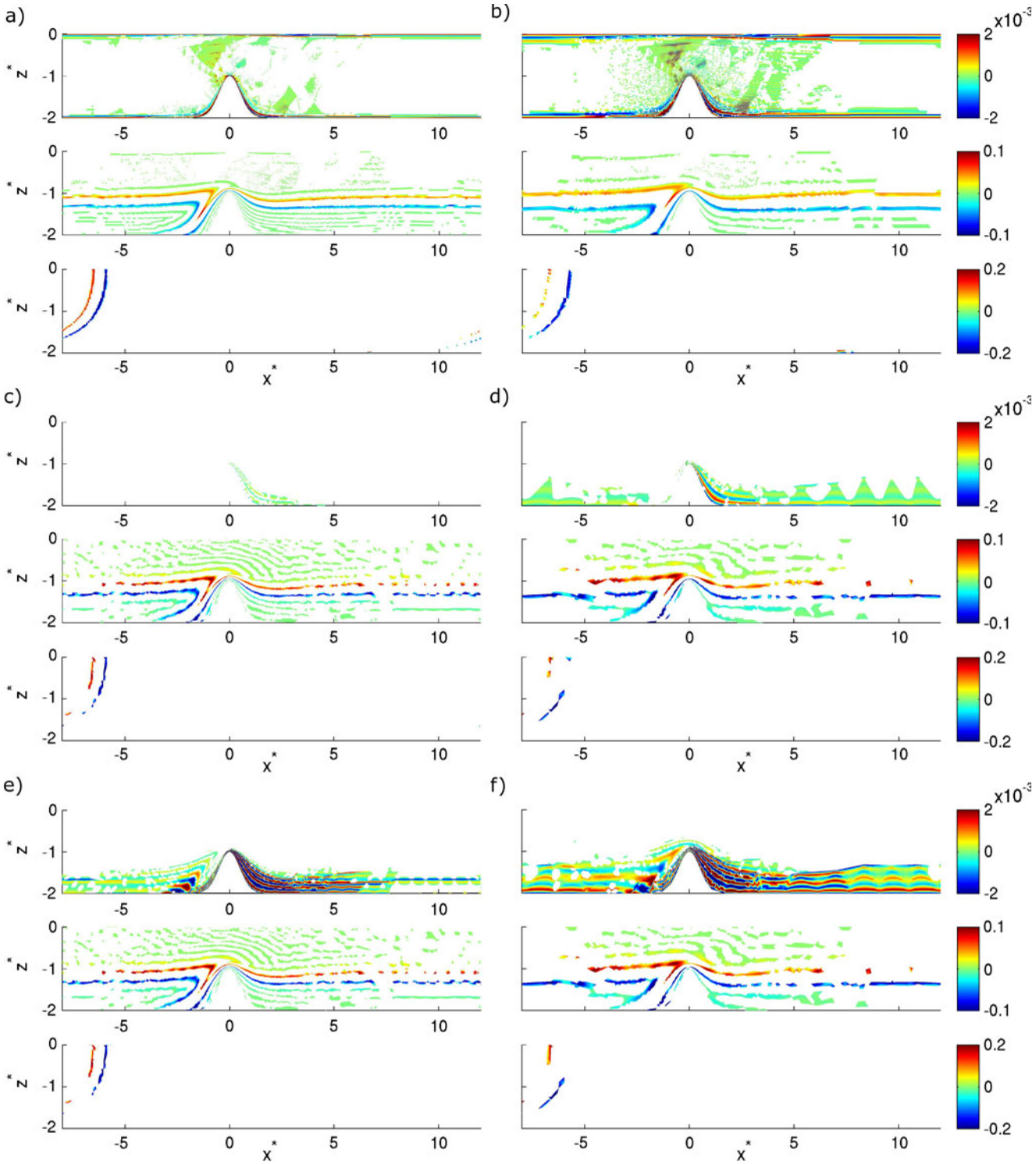

Fig. 10 Field values outside initially specified fields with **a** ($g5$, $p1$), **b** ($g4$, $p1$), **c** ($g2$, $p5$) on curved mesh, **d** ($g1$, $p6$) on curved mesh, **e** ($g2$, $p5$) on straight mesh, and **f** ($g1$, $p6$) on straight mesh. For each case (**a**–**e**), the *top* plot shows advection of a constant tracer with interval [0.9999, 1.0001], *middle* is advection of tracer with vertical jump and interval [0, 1], and *bottom* is advection with a horizontal gradient and horizontal jump with interval [−2, 3]

the magnitude of the oscillations are within 20% of the solution, and this could have a significant effect on the biology. This illustrates that discontinuous or highly nonsmooth functions caused by physics or biology are problematic. In the context of this work, the jump will be smoother for the test cases in Section 4.4 where we examine the case for approximately balanced advection and biological source terms. Alternatively, a shock-capturing (see, for example, Hoteit et al. 2004; Persson and Peraire 2006; Krivodonova 2007), filtering (for example, Hesthaven and Kirby 2008), or postprocessing technique could be used to handle the discontinuity (see, for example, Cockburn et al. 2003; Qiu and Shu 2005). In Section 4.6, we illustrate how a filtering approach can be used to damp oscillations for high-order schemes. Note that special treatment is required for both high- and low-order schemes, with a slope-limiting procedure more often used for the latter.

We note that the large oscillations in case 1 for the high-order scheme happen only down-stream from the peak. This is because we do not ensure that the definition for the discrete divergence is the same in the equations calculating the potential flow field and the tracer/biological ADR equations (see, for example, Dawson et al. 2004). Therefore, we expect to see discretization errors appearing in the ADR discrete divergence operator, which are advected downstream. Examining the normalized discrete divergence as defined in the ADR equations, $\left(\frac{\nabla\cdot\mathbf{u}}{|\mathbf{u}|}\right)$, of curved and straight meshes with $(g1, p6)$ and $(g4, p1)$, we found that the largest divergence errors occurs for the higher-order scheme on the straight mesh. The normalized divergence, in this case, was of $\mathcal{O}(1)$ near the peak, and $\left\|\frac{\nabla\cdot\mathbf{u}}{|\mathbf{u}|}\right\|_2 \approx 0.56$ for the domain. The normalized divergence for the low-order and high-order curved schemes were of $\mathcal{O}(10^{-2})$ near the peak, with the maximum error for the low-order scheme approximately half the size of the high-order scheme. However, the error for the high-order scheme was more localized, resulting in a lower volume-averaged error $\left\|\frac{\nabla\cdot\mathbf{u}}{|\mathbf{u}|}\right\|_2 \approx 0.044$, than that of the low-order scheme which was $\left\|\frac{\nabla\cdot\mathbf{u}}{|\mathbf{u}|}\right\|_2 \approx 0.11$. This reveals that it is important to use a curved mesh for higher-order schemes: The resultant flow field is then less numerically divergent than a flow field solved with a low-order finely resolved scheme. Without using a curved mesh, the divergence can be of the same order as the velocity near the geometry, which may excite nonphysical biological dynamics downstream of the peak. In our case (in Section 4.4), the peak was beneath the euphotic zone and thus using a straight boundary representation would not cause problems; however, in the general case, this result cannot be overlooked. Of course, our advection scheme is conservative, but to also be constancy preserving, the numerical flow field needs to be discretely divergence-free.

Since the $(g1, p6)$ simulations took less computational time than $(g4, p1)$ and because the $(g2, p5)$ simulation had a sharper interface, these results suggest that higher-order schemes performs better than the lower-order schemes for the advection dominated case, as long as curved boundaries are used.

4.4 Full NPZ equations

In this section, we explore the case where the advection and biological source terms are approximately balanced. We examine effects of low-order and high-order temporal discretizations in Section 4.4.1. In Section 4.4.2, we illustrate the difference between using a quadrature-based and quadrature-free scheme to discretize the nonlinear biological source terms. Finally, in Section 4.4.3, we study effects of spatial resolution, through both grid resolution and polynomial degree.

We still study the three biological parameter sets: single stable points, stable limit cycles at the bottom of the euphotic zone, and stable limit cycles for the entire euphotic zone, as given in Table 2. Since the timescale of biology varies in depth, the advection and biological source terms can only be balanced for one depth. While this results in many choices of approximately balanced parameter sets, we focus on one where $\bar{\tau}_a = 12.5$ days. For these tests, the inlet is specified as the steady-state solution with a smoothed discontinuity. The discontinuity is smoothed by fitting it with a cubic polynomial which can be resolved on $(g1, p6)$. The fit is biased such that 3/4 of the polynomial is below the euphotic zone. For the outlet boundary, we use $\frac{\partial\Phi}{\partial\hat{\mathbf{n}}} = 0$. Also these results are compared to a $(g5, p1)$ simulation, which is taken as the true solution.

The final solution fields for the three different regimes of biological dynamics (from Table 2) and using quadrature-based source terms for $(g5, p1)$ (the reference solution) is plotted in Fig. 11. The results show that idealized strait bathymetry effectively perturbs the biology away from the inlet conditions. The case with single stable points (bio case 1) adjusts back to the stable equilibrium, whereas the two cases with limit cycles show complex structures in the vertical. In all cases, a phytoplankton bloom over the bump is observed.

To qualitatively evaluate the effect of refining the grid or polynomial degree, we show the solution field for phytoplankton for $(g3, p1)$, $(g3, p2)$, $(g4, p1)$, and $(g4, p2)$ in Fig. 12, and these discretizations have

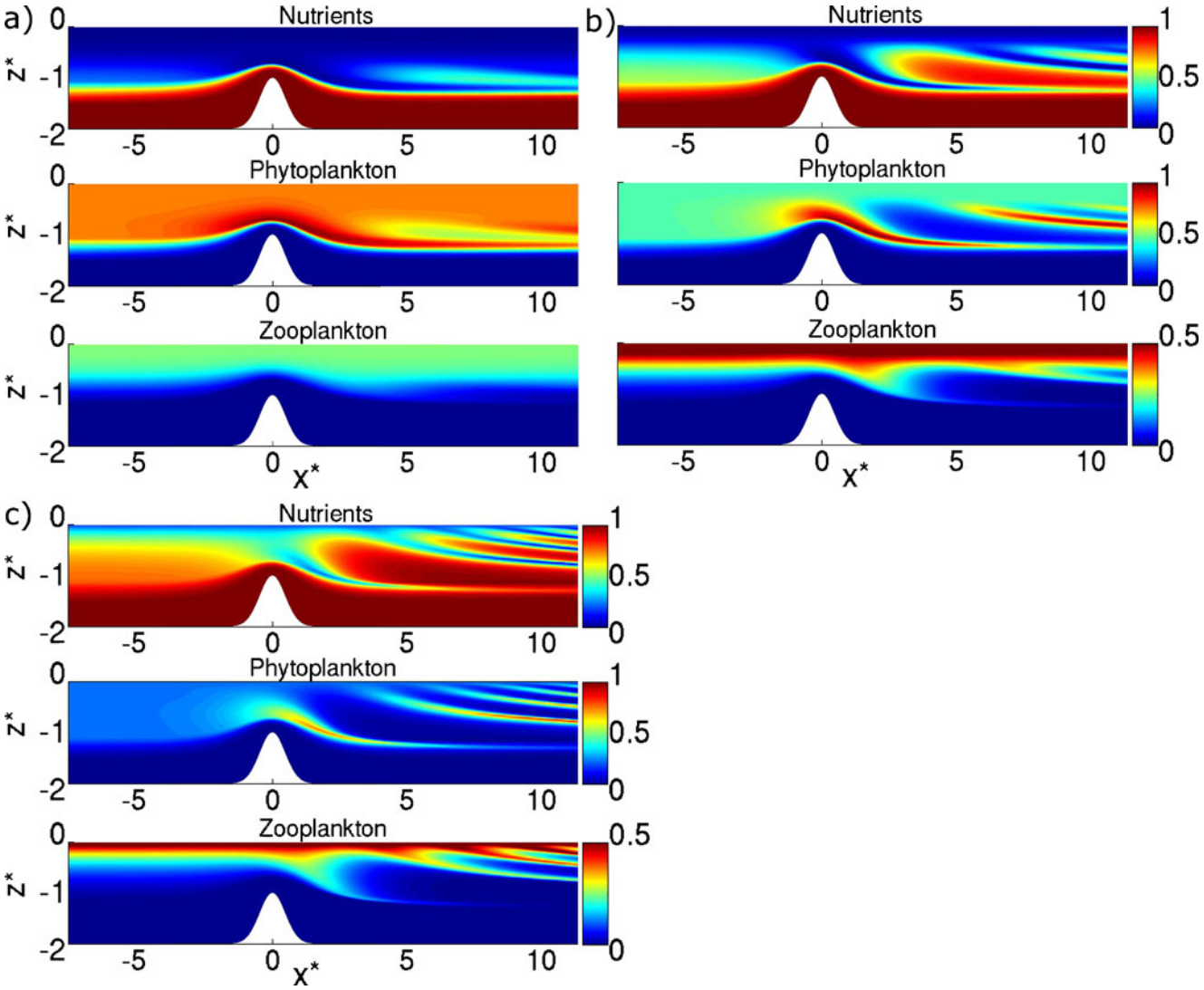

Fig. 11 Biological dynamics at $t^* = 20$ (with $\bar{\tau}_a = 12.5$ days) using ($g5$, $p1$). Biolocial dynamics with **a** single stable points, **b** stable limit cycles for depths $z^* = 0.4$–0.9, and **c** stable limit cycles in whole euphotic zone. This is the reference solution against which all other solutions are compared

16,800, 33,600, 67,200, and 134,400 DOFs, respectively. This figure shows that the solution is converging with increased resolution. More quantitative comparisons are completed next.

4.4.1 Comparing low-order and high-order temporal discretizations

We compare the solutions using fourth-order Runge–Kutta, second-order Runge–Kutta, and first-order Explicit Euler on ($g2$, $p4$) for the biology with stable limit cycles in the euphotic zone (bio case 3). The differences of the lower-order schemes compared to fourth-order Runge–Kutta at $t^* = 40$ is plotted in Fig. 13 for the phytoplankton field. Note that the timestep size for the first-order scheme is half of the second-order scheme, such that the cost of the two are the same. For this test case, we used periodic boundary conditions. From the figure, we note that the major differences occur within the euphotic zone. The stable explicit timestep for the second-order scheme is set by the Courant condition for the advection discretization, and since the largest velocity occurs in the smallest element for this discretization, the timestep size is approximately four orders of magnitude smaller than the biological timescale. Therefore, it is expected that temporal errors in the source term should be small even for the low-order scheme. Nonetheless, we still observe differences between the first-, second-, and fourth-order schemes. We

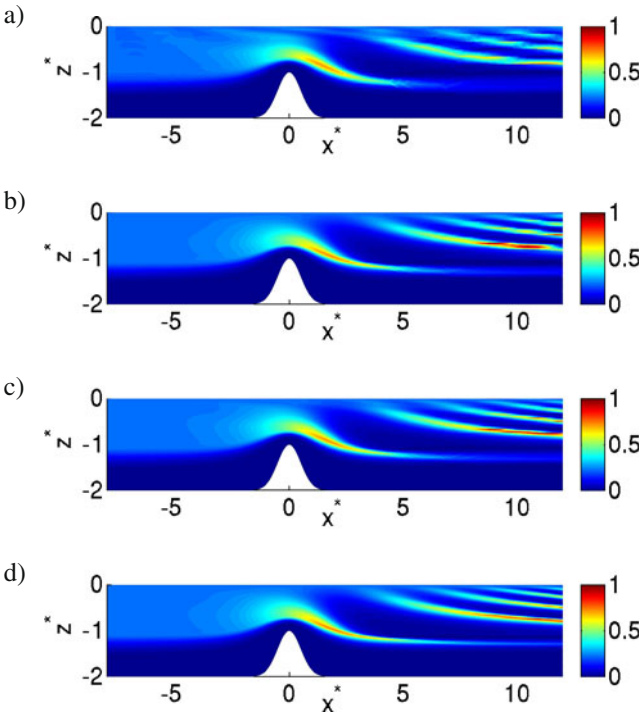

Fig. 12 Phytoplankton fields at time $t^* = 20$ (with $\bar{\tau}_a = 12.5$ days), as computed using four different spatial resolutions and order of the FE scheme: **a** ($g3$, $p1$), (16,800 DOFs) **b** ($g3$, $p2$) (33,600 DOFs), **c** ($g4$, $p1$) (67,200 DOFs), and **d** ($g4$, $p2$) (134,400 DOFs). All fields are for biological dynamics with stable limit cycles in the euphotic zone (bio case 3 in Table 2)

found that the difference at $t^* = 40$ is approximately two orders of magnitude larger than at $t^* = 20$, which indicates that the errors are growing quickly. For the first-order scheme, the maximum error is of $\mathcal{O}(1)$ at $t^* = 40$. This suggests that a low-order time discretization may result in significant errors when long integration times or fast biological timescales are involved. For example, the latter occurs in coastal applications. As

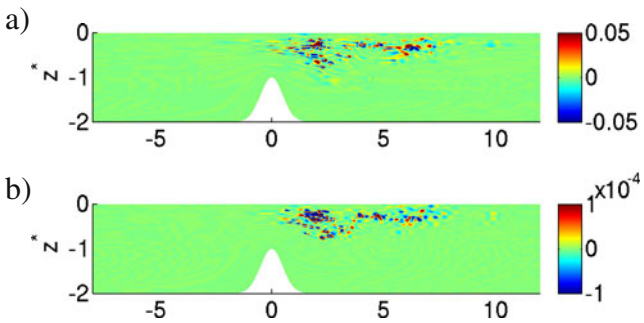

Fig. 13 Temporal discretization differences for phytoplankton field with stable limit cycles in euphotic zone at $t^* = 40$ using periodic boundary conditions and on spatial grid ($g4$, $p2$). **a** "First-order Euler" minus "fourth-order Runge–Kutta" and **b** "second-order Runge–Kutta" minus "fourth-order Runge–Kutta"

another example, for stiff biogeochemical source terms, Burchard et al. (2005) found that even fourth Runge–Kutta integration is insufficient to maintain the nonnegativity of the biological components. They suggest that positivity preserving Patankar–Runge–Kutta schemes should be used to obtain a nonnegative, conservative solution.

4.4.2 Comparing quadrature-based and quadrature-free source terms

In Section 4.1, we found that the greatest difference between the quadrature and quadrature-free treatment of the source terms occurred for the biological parameter set with stable limit cycles in the euphotic zone (i.e., bio case 3). Here we examine this case for full ADR dynamics using the ($g1$, $p6$) discretization. Note that we obtained the same results and conclusions with the ($g4$, $p1$) and ($g5$, $p2$) discretizations (not shown). Plotting the difference (quadrature-free minus quadrature-based) of the solution in Fig. 14 for ($g1$, $p6$), we see that the largest differences occur near the outlet of the domain where the mesh solution is under-resolved. The quadrature-based solution is more accurate in the under-resolved region because the source-term integral is more accurately evaluated, and this was verified by comparing the errors of the two schemes. However, where the solution is sufficiently resolved, the quadrature-free and quadrature-based treatments of

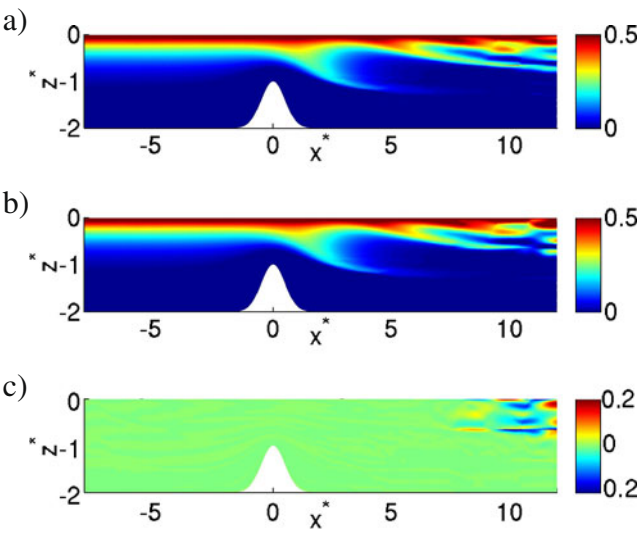

Fig. 14 Zooplankton fields at $t^* = 20$ computed using ($g1$, $p6$) and **a** quadrature-based source terms and **b** quadrature-free source terms. **c** The difference between the quadrature-free and quadrature-based source-term simulations. The biological dynamics used has stable limit cycles within the euphotic zone (bio case 3)

the source terms have similar accuracy, that is, they differ by approximately 0.1%. From the 1D studies, we found that the quadrature-free algorithm was less oscillatory at element interfaces than the quadrature-based algorithm, and we observed the same effect in these 2D simulations for $p > 7$ on $g1$, although the difference between quadrature-free and quadrature-based was less drastic. The largest differences between the quadrature-based and quadrature-free schemes did occur at element boundaries, and the quadrature-based algorithm was more accurate when under-resolved.

Using Eq. 30, we verify the conservation of the scheme. The results for the quadrature-free and quadrature-based source terms were similar up to floating point precision. Also, we find that the conservation error is dominated by the flow field divergence error. Therefore, the conservation properties of the source term discretization does not affect the choice between quadrature-free and quadrature-based algorithms.

Because the quadrature-free and quadrature-based algorithms had similar accuracy in well-resolved regions, we recommend using the quadrature-free treatment in these regions because of the improved efficiency. However, when the solution is poorly resolved, the quadrature-based treatment of the source terms is more accurate. Now, depending on the total solution cost of a particular numerical scheme, a finer resolution quadrature-free scheme may be more efficient for the same accuracy than a quadrature-based scheme.

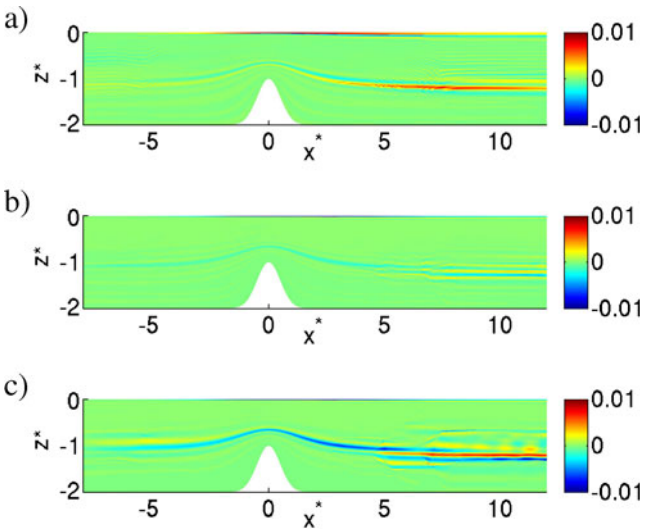

Fig. 15 Difference between zooplankton fields at $t^* = 20$ (with $\bar{\tau}_a = 12.5$ days) computed using ($g5$, $p1$) and **a** ($g4$, $p1$), **b** ($g2$, $p5$), and **c** ($g1$, $p6$). This shows the locations of the largest numerical errors for the high-order and low-order schemes. The biological dynamics used have single stable points at all depths (bio case 1)

4.4.3 Comparing low-order and high-order spatial discretizations

Figure 15 shows the differences between the reference solution and the solutions using other grids and polynomial degrees in Fig. 15 for zooplankton. We see that the ($g2$, $p5$) simulation has the smallest differences and is therefore the most accurate. This is a key result since it indicates that when the solution is resolved, for the same cost/efficiency, a higher-order scheme on a coarser grid performs better than a lower-order scheme on a finer grid. Results for the biological dynamics with limit cycles at the bottom of the euphotic zone (bio case 2) are similar, but for the biology with limit cycles in the entire euphotic zone (bio case 3), both high- and low-order schemes are under-resolved for $x^* > 7$, especially for the ($g1$, $p6$) scheme. The differences between a fine grid solution ($g5$, $p1$) and the low-order and high-order schemes are plotted in Fig. 16 for zooplankton. From Fig. 16, we note that the errors in the low-order scheme are more localized in the $x^* > 7$ region. However, the differences for ($g4$, $p1$) and ($g2$, $p5$) are similar in the $x^* > 7$ region. The ($g1$, $p6$) scheme has the least localized and largest magnitude errors in the $x^* > 7$ region. However, as plotted in Fig. 17 where the solution is smooth and the biology has less structure in the vertical, both the high-order schemes are more accurate than the low-order scheme. Particularly, note the solution near the surface for $x^* < 7$ in Fig. 17.

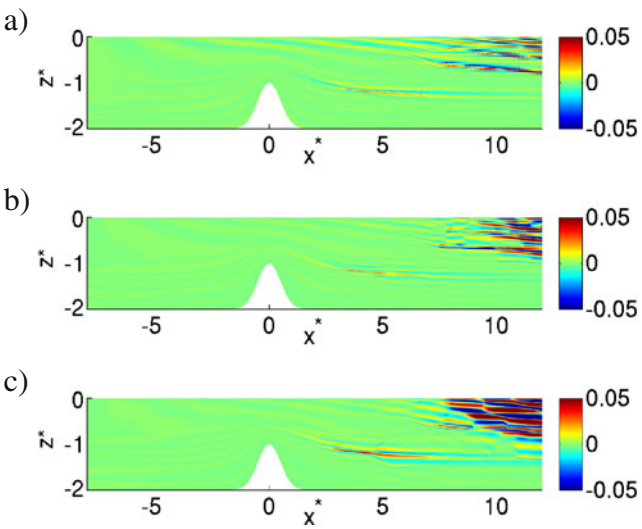

Fig. 16 As Fig. 15, but for the biological dynamics with stable limit cycles within the euphotic zone (bio case 3)

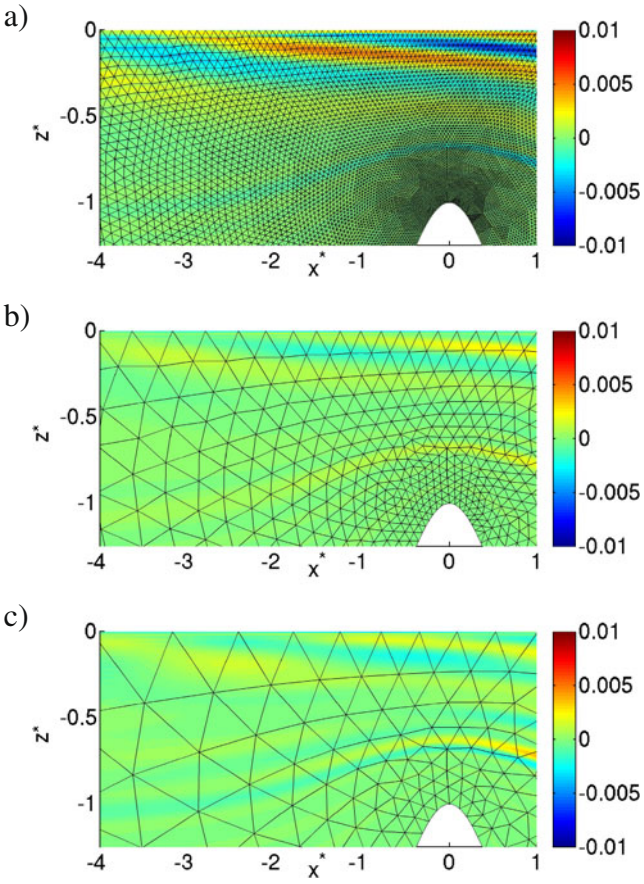

Fig. 17 As Fig. 16, but zoomed in the region above the bathymetry. The difference between zooplankton fields using ($g5$, $p1$) and **a** ($g4$, $p1$), **b** ($g2$, $p5$), and **c** ($g1$, $p6$)

We examine the error characteristics of these fields more closely by considering the truncated Taylor expansions of the true solution. By the mean value theorem, the truncation error for ($g4$, $p1$) is $\frac{h^2}{2!}\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^2\phi(\mathbf{x}_\eta)$ for some unknown point $\mathbf{x}_\eta$, and for ($g2$, $p5$) and ($g1$, $p6$) these terms are $\frac{h^7}{6!}\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^6\phi(\mathbf{x}_\zeta)$ and $\frac{h^7}{7!}\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^7\phi(\mathbf{x}_\xi)$ for unknown points $\mathbf{x}_\zeta$ and $\mathbf{x}_\xi$, where h is the characteristic discretization length of an element. Now, we can examine the approximate truncation error by running simulations ($g4$, $p2$), ($g2$, $p6$), and ($g1$, $p7$) and evaluating the highest-order nonzero derivatives of the approximate solution ϕ_h. To evaluate the derivatives, we interpolate the solution onto an orthogonal modal polynomial basis, that is, we find the coefficients a_{ij} such that $\phi_h = \sum_i \phi_i\theta_i = \sum_{ij} a_{ij}P_{ij}$, where P_{ij} is a modal orthogonal polynomial with maximum degree of i on x and j on y, for a maximum total degree of $i+j$. The derivatives then evaluate as $\frac{1}{2!}\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^2\phi_h(\mathbf{x}_\eta) = \sum_{i+j=2} a_{ij}$, $\frac{1}{6!}\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^6\phi_h(\mathbf{x}_\xi) = \sum_{i+j=6} a_{ij}$, and $\frac{1}{7!}\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^7\phi_h(\mathbf{x}_\xi) = \sum_{i+j=7} a_{ij}$, that is, we simply need to sum the coefficients of the modal orthogonal polynomial basis which correspond to terms with total degree of 2, 6, and 7, respectively. Since the coefficients of the polynomials are evaluated on the reference element, $h \approx 1$ will be the same for all elements. Also, while this approach gives an estimate of the leading order truncated term, it does not give an exact value. In our case, we are not interested in a rigorous error estimator, but instead we only require an estimate of the error to aid the discussion.

Our approach is similar to that followed by Mavriplis (1989), where Legendre polynomials were used instead. The author proposed a smoothness estimator, where the coefficients a_{ij} are fit to the exponentially decaying function $a(i+j) = Ce^{\sigma(i+j)}$. There the author claims that $\sigma < -1$ indicates good resolution or smooth functions and $\sigma > -1$ indicates poor resolution or nonsmooth functions. The adaptive strategy used was to increase the polynomial degree for elements with $\sigma < -1$ and to refine the mesh for elements with $\sigma > -1$, if the error level in that element was insufficient. We evaluate this smoothness indicator σ on ($g1$, $p7$) by doing a least squares fit of the coefficients to $Ce^{\sigma(i+j)}$. In regions where the magnitude of the solution is close to 0, that is below the euphotic zone for the zooplankton field, $a_{ij}\ \forall i,\ j$ will be small, and the smoothness indicator σ will not be accurate. The approximate size of the truncated derivative terms along with the smoothness indicator are plotted in Fig. 18. Only the smoothness indicators calculated on ($g2$, $p5$) and ($g1$, $p7$) are plotted since the accuracy of the smoothness indicator improves with the number of terms in the polynomial expansion and is not accurately represented on ($g4$, $p2$) (Mavriplis 1989).

From Fig. 18, we note that the largest differences in Fig. 16 correspond to the regions with the largest approximate truncations errors in Fig. 18. Also, in the region $x^* > 7$ where the low-order solution is more accurate than ($g1$, $p6$), we have $\sigma > -1$, which suggests that refining the elements instead of the order of accuracy is more appropriate. After one level of refinement on ($g2$, $p5$), we see that the smoothness indicator shows a smaller region of nonsmooth elements. This illustrates that the smoothness is defined in terms of the numerical discretization and is not solely a function of the solution field. Also note in the region where the high-order solution is more accurate (see Fig. 17), the approximate derivative of the truncation term is small in both fields and $\sigma < -1$, suggesting that a higher degree polynomial basis is more appropriate in this region. This shows

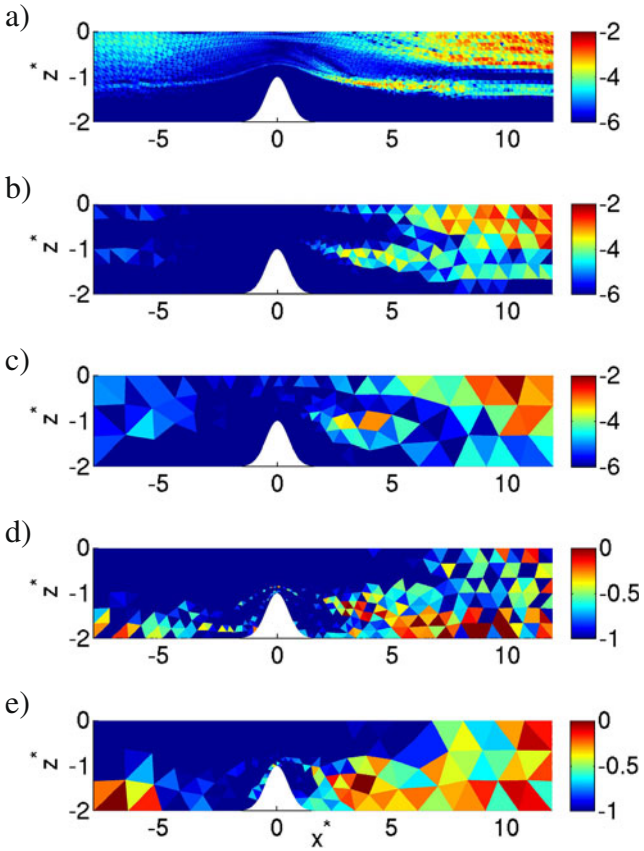

Fig. 18 Approximate truncation errors for zooplankton fields at $t^* = 20$ (with $\bar{\tau}_a = 12.5$ days). Calculated on **a** ($g4$, $p2$) using $log_{10}(\sum_{i+j=2} a_{ij})$, on **b** $g2$, $p6$ using $log_{10}(\sum_{i+j=6} a_{ij})$, and on **c** $g1$, $p7$ using $log_{10}(\sum_{i+j=7} a_{ij})$. **d**, **e** Smoothness indicator σ calculated on **d** ($g2$, $p6$) and **e** ($g1$, $p7$) . The biological dynamics used has stable limit cycles within the euphotic zone (bio case 3)

that our mesh is not optimized in terms of the solution field and highlights the importance of using both mesh refinement and polynomial basis adaptation to generate an optimal discretization for complex biological ocean dynamics. Also, this shows that whether a coarsely discretized higher-order scheme is better than a finely discretized lower-order scheme depends on the smoothness of the solution and can vary spatially across the solution. The benefit from a higher-order solution is as follows: When the solution is smooth, increasing the polynomial degree causes the error to decrease exponentially, whereas the error would only decrease algebraically if decreasing the element size. The cost of increasing the polynomial degree also scales algebraically, and because of this, a higher-order scheme performs better for smooth or well-resolved fields. Using our implementation, the ($g1$, $p6$) simulation took approximately 0.34 of the time taken by the ($g4$, $p1$) simulation. We also ran ($g3$, $p1$), which was approximately three times more efficient than ($g1$, $p6$), but this solution (not shown) was less than 1% accurate for the majority of the domain. When the solution is not resolved (i.e., not smooth for the grid resolution or polynomial degree chosen), higher-order schemes will lead to Gibbs oscillations and filtering is required (see Section 4.6), while lower-order schemes may "look good" but will be very dissipative. When the solution is resolved (i.e., smooth enough for the grid resolution or polynomial degree chosen), higher-order discretizations perform better than lower-order ones: They are more accurate and less dissipative for the same cost.

Finally, we note that the approximate truncation error and smoothness metrics were different for the different biological components. Therefore, the optimal discretization for one component is not the same as the optimal discretization for another component. Ueckermann (2009) proposed a scheme that uses a different order basis function for different biological components, but also cautions that an incurred interpolation cost needs to be considered for adaptation strategies.

4.5 Evolution of biological patch

In this section, we demonstrate how biological activity can enhance the differences between low-order and high-order discretizations beyond the effect of numerical dissipation alone. For this example, we modify bio case 1 (single stable points at all depths) from Section 4.4 by introducing a vertical column, or "patch", of biology that uses the parameters from bio case 2 (stable limit cycles at depths $z^* = 0.4$–0.9 and single stable points elsewhere). This is easily done in the dimensional form of the equations by increasing the value of $\mathcal{N}_T$ locally in the patch. Such situations occur frequently in nature, e.g., an eddy or front upwelling additional nutrients locally toward the surface. The initial condition and boundary condition is the same as in bio case 1 (the steady-state solution with smoothed discontinuity), except inside the patch where the initial conditions for bio case 2 are used instead, that is:

$$\phi^*_{(\text{patch})} = \phi^*_{(\text{bio case 1})} + \left[\phi^*_{(\text{bio case 2})} - \phi^*_{(\text{bio case 1})}\right] \cdot e^{-\frac{(x^*+6.4)^4}{2\cdot(8^4)}}, \qquad (31)$$

where $\phi^*_{(\text{patch})}$ is the initial condition used for this example, $\phi^*_{(\text{bio case 1})}$ is the steady state with smoothed discontinuity for bio case 1, and $\phi^*_{(\text{bio case 2})}$ is the steady state with smoothed discontinuity for bio case 2. Note

that, for this example, we nondimensionalize $\phi^* = \frac{\phi}{\mathcal{N}_{T,\text{bio case 1}}}$ by the total biomass for bio case 1. In addition, we superimpose a periodic velocity onto the mean velocity,

$$\mathbf{u}^* = \mathbf{u}^*_{\text{mean}}[1 + 5 \cdot \text{sign}\{\cos(0.16\pi t^*)\}], \tag{32}$$

where $\mathbf{u}^*$ is now the velocity used for this example and $\mathbf{u}^*_{\text{mean}}$ is the potential flowfield solved from Section 4.2. The superimposed velocity increases the distance traveled, as well as the number of time integration steps (due to the CFL condition), and therefore has the effect of increasing the numerical dissipation.

Figure 19 plots the phytoplankton fields and total biomass for ($g2$, $p5$) and ($g4$, $p1$) around the patch, as well as the difference of the solution, ($g2$, $p5$) minus ($g4$, $p1$), at $t^* = 14.4$. We do not use the ($g5$, $p1$) solution (as was done in Section 4.4) for the difference plots here because we found that even ($g5$, $p1$) is more dissipated than ($g2$, $p5$) and therefore ($g2$, $p5$) is more accurate inside the biological patch where our calculations are performed. These results show that the total biomass peak is not maintained by the low-order scheme, ($g4$, $p1$), and the details in the phytoplankton fields are also dissipated. Since these simulations do not contain physical diffusion, any diffusion is due to the numerical scheme, and therefore, the ($g2$, $p5$) solution is more accurate than ($g4$, $p1$) because it does maintain the total biomass peak. Apart from the effects of the periodic velocity, the solution inside the patch should resemble that of Fig. 11b), and ($g2$, $p5$) resembles this solution more closely than ($g4$, $p1$).

While some of the differences between the ($g2$, $p5$) and ($g4$, $p1$) simulations can be accredited solely to the numerical dissipation, the error due to numerical dissipation is amplified by the change in biological activity. To illustrate this point, we show in Fig. 20 the relative normed difference between the total biomass of the two solutions (Q_1, Eq. 33), the sum of relative normed differences between the biological components (Q_2, Eq. 34), the relative normed difference in production

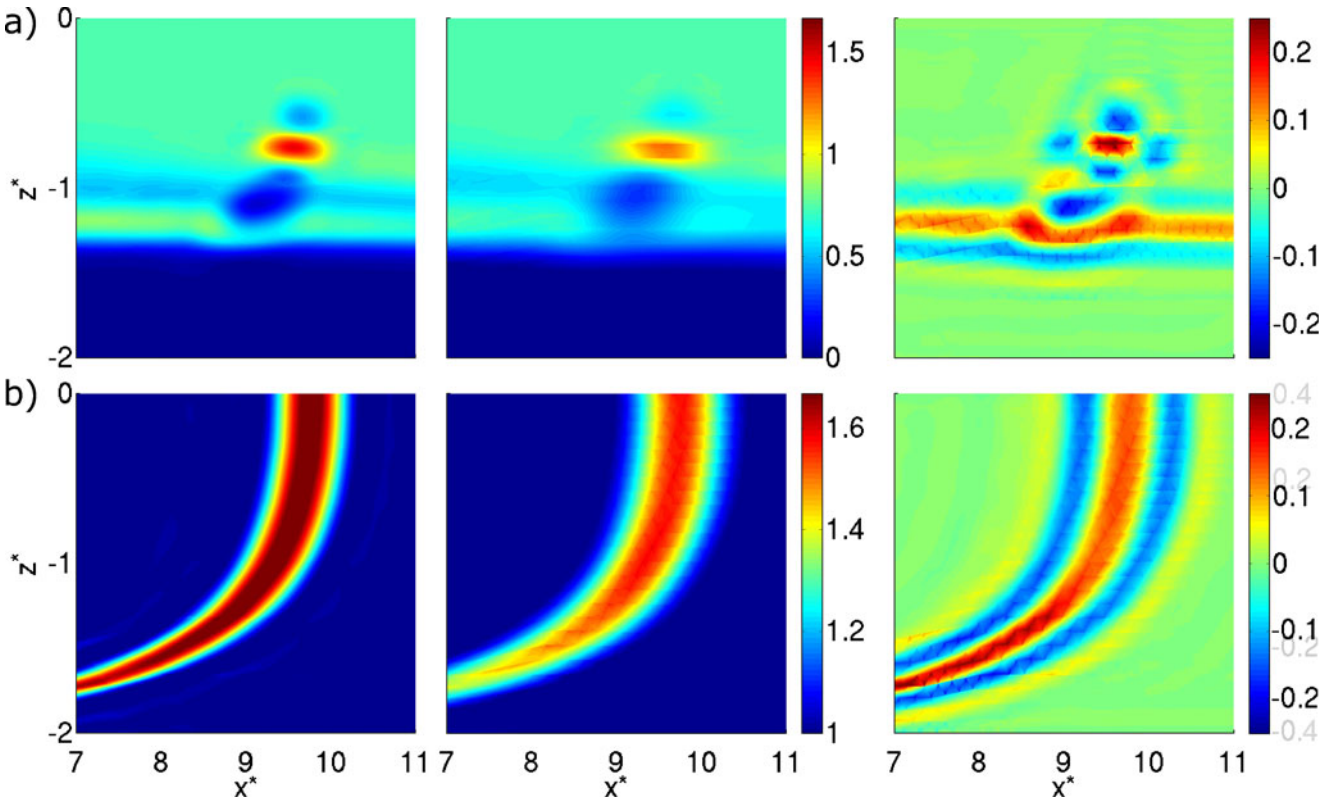

Fig. 19 Detail around the biological patch with stable limit cycles at the bottom of the euphotic zone at time $t^* = 14.4$ for **a** the phytoplankton fields and **b** the total biomass. The solution for ($g2$, $p5$) is plotted on the *left*, ($g4$, $p1$) in the *middle*, and the difference between the solutions, [($g2$, $p5$) − ($g4$, $p1$)], is plotted on the *right*. This shows that ($g2$, $p5$) correctly maintains the full peak of the biological patch, while ($g4$, $p1$) does not, leading to large differences in the phytoplankton fields

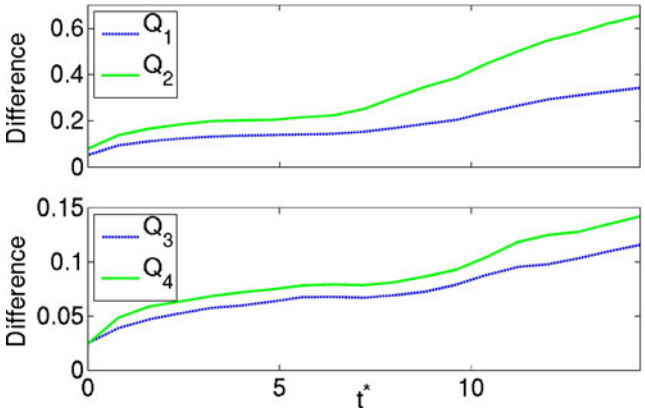

Fig. 20 The relative normed difference between the total biomass of the two solutions (Q_1, Eq. 33), the sum of relative normed differences between the biological components (Q_2, Eq. 34), the relative normed difference in production (Q_3, Eq. 35), and the relative normed difference in grazing (Q_4, Eq. 36) over time from $t^* = 0$ to $t^* = 14.4$. This shows that the difference in biological components is amplified beyond the effect of numerical dissipation due to differences in the source terms such as the production and grazing

(Q_3, Eq. 35), and the relative normed difference in grazing (Q_4, Eq. 36)

$$Q_1 = \frac{\left\| \left\{\phi_N^*+\phi_P^*+\phi_Z^*\right\}_{(g2,p5)} - \left\{\phi_N^*+\phi_P^*+\phi_Z^*\right\}_{(g1,p4)} \right\|_2^{\text{patch}}}{\left\| \left\{\phi_N^*+\phi_P^*+\phi_Z^*-1\right\} \right\|_2^{\text{patch}}} \tag{33}$$

$$Q_2 = \frac{\sum_{I=(N,P,Z)} \left\| \phi_{I,(g2,p5)}^* - \phi_{I,(g1,p4)}^* \right\|_2^{\text{patch}}}{\left\| \left\{\phi_N^* + \phi_P^* + \phi_Z^* - 1\right\}_{(g4,p1)} \right\|_2^{\text{patch}}} \tag{34}$$

$$Q_3 = \frac{\left\| \left\{ \mathcal{U}^* e^{z^*/h^*} \frac{\phi_P^* \phi_N^*}{\phi_N^* + k_s^*} \right\}_{g2p5} - \left\{ \mathcal{U}^* e^{z^*/h^*} \frac{\phi_P^* \phi_N^*}{\phi_N^* + k_s^*} \right\}_{g4p1} \right\|_2^{\text{patch}}}{\left\| \left\{ \mathcal{U}^* e^{z^*/h^*} \frac{\phi_P^* \phi_N^*}{\phi_N^* + k_s^*} \right\}_{g4p1} \right\|_2^{\text{patch}}} \tag{35}$$

$$Q_4 = \frac{\left\| \left\{ a g_\nu^* \phi_Z^* (1 - e^{-\nu^* \phi_P^*}) \right\}_{g2p5} - \left\{ a g_\nu^* \phi_Z^* (1 - e^{-\nu^* \phi_P^*}) \right\}_{g4p1} \right\|_2^{\text{patch}}}{\left\| \left\{ a g_\nu^* \phi_Z^* (1 - e^{-\nu^* \phi_P^*}) \right\}_{g4p1} \right\|_2^{\text{patch}}}, \tag{36}$$

where $\|e\|_2^{\text{patch}} = \left(\int_{\text{patch}} e^2 dx^* dz^* \right)^{\frac{1}{2}}$ with the patch area determined from $(g4, p1)$ and the quantity $\left\| \{\phi_N^* + \phi_P^* + \phi_Z^* - 1\} \right\|_2^{\text{patch}}$ gives the size of the difference between the base solution and the solution inside the patch since the base number of nutrients (nondimensionalized to 1) is subtracted out.

Since our numerical scheme conserves the total biomass, the first quantity, Q_1, gives a quantitative estimate of the numerical dissipation error only. The second quantity, Q_2, should be the same as Q_1 if the only difference is due to numerical dissipation. However, from Fig. 20, we note that $Q_2 > Q_1$, which means the differences in dissipation is amplified by nonlinearities and the biology. This is explained by the differences in biological terms in the two simulations, for example, in the production and grazing terms, Q_3 and Q_4. Also, note that these differences are growing over time, and for longer integration periods, the differences will be even greater. As a final note, the initial differences between the two solutions are due to interpolation errors, since the polynomial representation and number of degrees of freedom are not the same for the two simulations.

The example shows that the numerical dissipation due to a lower-order numerical scheme can be amplified by the biological reaction terms. This is significant since for accurate biological ocean science through numerical simulations, it is important to maintain the amplitudes of biological patches. This is particularly true for biology with multiple attractors, where relatively small perturbations can lead to vastly different solutions. The conclusion is that for the same cost, higher-order schemes on coarser grids are more accurate than lower-order schemes on finer grids.

4.6 Filtering based on smoothness index

Based on the results in Section 4.4 and on the paper by Hesthaven and Kirby (2008), we created a selective exponential filter. We found that the default filter was described in Hesthaven and Kirby (2008), which is applied at every timestep, to result in a diffuse interface, even for high-order filters. This prompted us to develop the following filter. Consider a modal representation of the numerical solution, $\phi_h = \sum_{ij} a_{ij} P_{ij}$ (e.g. Ueckermann 2009). After each time integration step, the solution on every element is modified as follows:

$$\phi_h^F = \sum_{ij} \sigma(\eta_{ij}) a_{ij} P_{ij}, \tag{37}$$

where

$$\sigma = \begin{cases} \exp(-\alpha \eta_{ij}^s), & \text{if SI} > 0 \\ 1, & \text{if SI} < 0 \end{cases} \tag{38}$$

$$\text{SI} = \frac{a_{ij}}{\max_{i+j \geq 1}(a_{ij})} - \exp(-(i+j)+1) \tag{39}$$

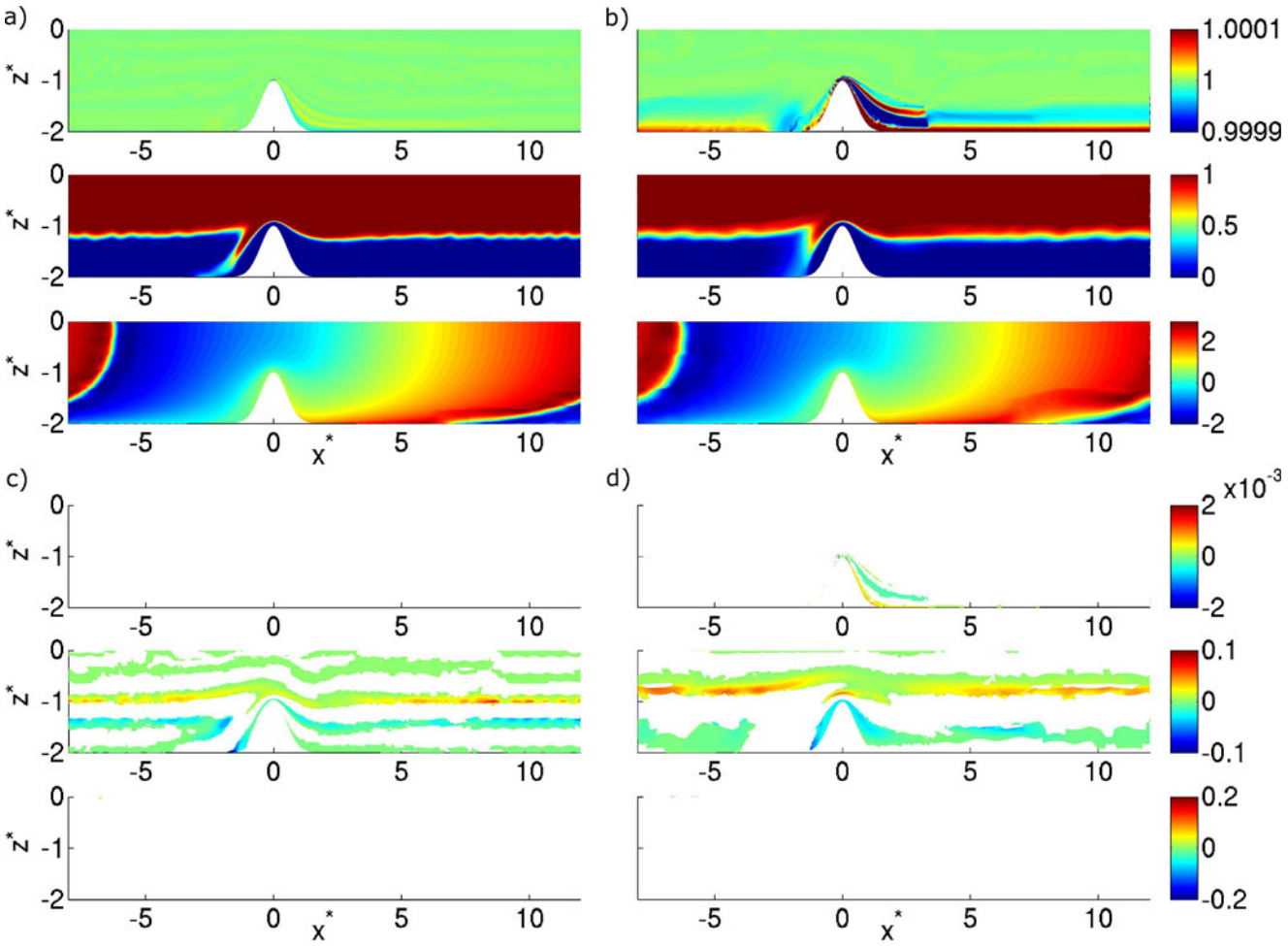

Fig. 21 Solution of three passive tracer cases as Fig. 9, for **a** ($g2$, $p5$) and **b** ($g1$, $p6$) but using our selective exponential filter (37) with $s = 5$. Field values outside initially specified fields as Fig. 10, with **c** ($g2$, $p5$) and **d** ($g1$, $p6$) using the same filter. For each case **a**–**d**, the filter damps the initialized oscillations of cases 1 and 2 and no oscillations are created in case 3

$$\eta = \frac{i+j}{p+1} \tag{40}$$

$$\alpha = -\log(0.01), \tag{41}$$

with s being the order of the filter and SI the smoothness index.

This results in a filter that is applied only when the smoothness index indicates that the solution is not smooth. The smoothness index relies on information about the decay rate of the modal coefficients, so it cannot be used for lower-order schemes. Using this new filter, we obtain a solution where spurious numerical oscillations are no longer generated by the numerical scheme and the interfaces are not significantly diffused. The new errors plots are shown in Fig. 21. We also verified that this filter does not affect smooth regions of the domain with other idealized test cases (not shown).

While this filter looks promising, additional verification is needed, and a number of improvements are also possible. In particular, determining the strength or order, s, of the filter in an optimal manner is an open question. Also, selectively applying the filter only in directions where the solution is nonsmooth is a topic of future research.

5 Conclusions

We completed a set of computational studies for the modeling of multiscale biogeochemical dynamics in coastal ocean regions with complex bathymetric features, utilizing recent advances in computational fluid

dynamics. Specifically, we compared low- to high-order discretization schemes, both in time and space, employing standard and hybrid discontinuous Galerkin finite element methods, on both straight and curved elements. We studied the effects of a varied set of numerical properties including quadrature-free and quadrature-based discretizations of the source terms, order of the spatial discretizations of advection and diffusion operators, order of the temporal discretization in explicit schemes, and resolution of the spatial mesh, with and without our new curved elements. We verified the convergence of our numerical schemes for both the biology and flow fields, validated the codes on analytical solutions, and completed a rigorous truncation error analysis.

Our numerical analyses concentrated on the nonlinear nutrient–phytoplankton–zooplankton dynamics under advection and diffusion within an ocean strait or sill, in an idealized 2D geometry. We first nondimensionalized the PDEs, evaluated stability regions, and selected three biological dynamical regimes: single stable points at all depths, stable limit cycles at the bottom of the euphotic zone, and stable limit cycles within the whole euphotic zone (the latter two cases have limit cycles that are depth and light dependent). We evaluated the effects of numerical parameters on the three biological regimes but illustrated only the most relevant results. In addition, for each of these biological regimes, we examined three types of coupled physics–biology interactions: biological terms dominating, advection terms dominating, and advection and biological terms balancing. For the advection-dominating case, we studied the advection over a strait of a uniform tracer, horizontal front and vertical front. For the balanced situation, relatively common in the real ocean, we considered biological dynamics that were either as fast as (e.g., coastal ocean) or slower than advection time scales.

In the regime where biological terms dominate, we found that both the quadrature-based and quadrature-free treatment of the source terms give accurate, convergent results, although the quadrature-based algorithm had slightly smaller errors. We also showed that oscillations can occur solely due to numerics (Gibbs-like phenomena) for a high-order discretizations. A key result is that, for any numerical scheme, careful 1D studies should be performed to understand the potential errors from the nonlinear source-term discretization.

For the advection-dominating regime, we confirmed the flow field convergence and using passive tracers studied numerical advection artifacts that would also affect the biology. We found that for discretizations that do not resolve the solution, oscillations due to discontinuities in the tracers could be large for both low-order and high-order discretizations but can be damped using a filtering approach for the high-order case. However, our results suggested that when the solution is resolved enough, higher-order schemes on coarser grids perform better (higher accuracy, less dissipative) for the same cost than lower-order scheme on finer grids, as long as curved boundaries were used.

For the case of approximately balanced advection and biological terms, we compared low- and high-order temporal and spatial discretizations and studied quadrature-based and quadrature-free discretizations of the source terms. We found that for lower-order temporal discretizations, the errors grew rapidly and would lead to inaccurate solutions for applications with faster biological timescales or longer integration times. We also showed that the quadrature-based source-term discretization was more accurate in regions where the solution was under-resolved, but in well-resolved regions, there was only a 0.1% discrepancy, and the quadrature-free algorithm could be used for efficiency purposes. By quantitatively evaluating the truncation error and smoothness of the solution fields, we confirmed that higher-order spatial discretizations were more accurate in regions where the solution was smooth (i.e., resolved enough) but less accurate where nonsmooth (un-resolved) due to Gibbs-like oscillations. To reduce these oscillations, we developed a new numerical filter that is active only when and where the solution is not smooth locally, using a smoothness indicator. Finally, we demonstrated the importance of nondissipative numerical schemes when biological patches are present which is common in the real ocean. First, we found that effects of numerical dissipation were amplified by biological activity, causing dissipation errors to increase faster with integration time. Higher-order spatial discretizations were more accurate when modeling biological patches because they maintained the patches while lower-order schemes did not. For resolved biology (e.g., as in Fig. 15), higher-order schemes on coarser grids were for the same cost more accurate than lower-order schemes on finer grids. This conclusion is most important for longer-term simulations. It has major implications for fundamental studies of biological blooms, patchiness, and other nonlinear dynamics in coastal regions with complex bathymetric features such as straits, sills, ridges, and shelfbreaks. One can expect similar implications for longer-term eddy-resolving ecosystem studies or climate applications.

Based on our results, future research directions are to further develop schemes to reduce Gibbs-like oscillations without significant loss of accuracy and efficiency (e.g., Persson and Peraire 2006). Without oscillation

limiters or filtering, the optimal performance could be obtained by using different polynomial degree basis functions in the domain, where low-order elements could be used in nonsmooth regions while high-order elements could be used in smooth regions. Because the smoothness can be determined from the discretization, an adaptive grid and polynomial degree scheme could be developed. Another possibility in this case could be to increase the grid resolution and decrease the order of schemes (e.g., medium-order schemes, i.e., ($g3$, $p3$) or ($g3$, $p4$)) up to the point when numerical oscillations reach the size of other errors. Another research direction is to develop and evaluate schemes that would preserve the nonnegativity of the biological solution. Our results can now be utilized for idealized studies of biological dynamics in straits or sills. Uncertainty quantifications (Lermusiaux 2006; Sapsis and Lermusiaux 2009) as well as adaptive model learning (Lermusiaux 2007) for biological predictions would also be useful. Finally, we are now well positioned to implement these new methods in 3D ocean modeling systems (e.g., MSEAS Group 2010) for realistic coupled biogeochemical–physical ocean science and applications.

Acknowledgements We are grateful to Dr. Patrick Haley Jr. and Ms. Lisa Burton for their collaboration on numerics and biology. We thank Dr. N.C. Nguyen for discussions on DG and HDG and Dr. J. Peraire discussions on numerics. We thank the Office of Naval Research for sustained research support under grants N00014-07-1-1061, N00014-08-1-1097, and N00014-07-1-0473 to the Massachusetts Institute of Technology. We are grateful to the Natural Sciences and Engineering Research Council (NSERC) of Canada for the Postgraduate Scholarship to the first author and to Mr. and Mrs. Pappalardo for partially supporting this research through a Pappalardo Fellowhip. We thank Dr. M. Iskandarani and the reviewers for the detailed and helpful comments and suggestions.

References

Agarwal A (2009) Statistical field estimation and scale estimation for complex coastal regions and archipelagos. S.M. thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering

Agarwal A, Lermusiaux PFJ (2010) Statistical field estimation for complex coastal regions and archipelagos. Ocean Model, Sub Judice (in press)

Anderson D, McGillicuddy D, Townsend D, Turner J (2005) The ecology and oceanography of toxic Alexandrium Fundyense blooms in the Gulf of Maine—preface. Deep-sea Res Part 2 Top Stud Oceanogr 52(19–21):2365–2368

Bassi F, Rebay S (1997) A high-order accurate discontinuous Galerkin finite element method for the numerical solution of the compressible Navier–Stokes equations. J Comput Phys 131:267–279

Bernard PE, Remacle JF, Legat V (2009) Boundary discretization for high-order discontinuous Galerkin computations of tidal flows around shallow water islands. Int J Numer Methods Fluids 59(5):535–557

Besiktepe ST, Lermusiaux PFJ, Robinson AR (2002) Coupled physical and biogeochemical data-driven simulations of Massachusetts Bay in late summer: real-time and postcruise data assimilation. J Mar Syst 40:171–212

Bourgault D, Kelley DE (2004) A laterally averaged nonhydrostatic ocean model. J Atmos Ocean Technol 21:1910–1924

Budgell WP, Oliveira A, Skogen MD (2007) Scalar advection schemes for ocean modelling on unstructured triangular grids. Ocean Dyn 57(4–5):339–361

Burchard H, Deleersnijder E, Meister A (2005) Application of modified Patankar schemes to stiff biogeochemical models for the water column. Ocean Dyn 55:326–337

Burton LJ (2009) Modeling coupled physics and biology in ocean straits with application to the San Bernardino strait in the Philippine Archipelago. S.M. thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering

Cebeci T, Shao JP, Kafyeke F, Laurendeau E (2005) Computational fluid dynamics for engineers. Springer, New York

Chapra SC, Canale RP (2006) Numerical methods for engineers, 5th edn. McGraw-Hill Higher Education, Boston

Chung TJ (2002) Computational fluid dynamics. Cambridge University Press, New York

Cockburn B (1998) An introduction to the discontinuous Galerkin method for convection dominated flows. In: Quarteroni A (ed) Advanced numerical approximation of nonlinear hyperbolic equations. Springer, New York

Cockburn B, Shu CW (1989) TVB Runge–Kutta local projection discontinuous Galerkin finite-element method for conservation-laws. 2. General framework. Math Comput 52(186):411–435

Cockburn B, Shu CW (1998a) The local discontinuous Galerkin method for time-dependent convection–diffusion systems. SIAM J Numer Anal 35(6):2440–2463

Cockburn B, Shu CW (1998b) The Runge–Kutta discontinuous Galerkin method for conservation laws V—multidimensional systems. J Comput Phys 141(2):199–224

Cockburn B, Lin SY, Shu CW (1989) TVB Runge–Kutta local projection discontinuous Galerkin finite-element method for conservation-laws. 3. One-dimensional systems. J Comput Phys 84(1):90–113

Cockburn B, Hou SC, Shu CW (1990) The Runge–Kutta local projection discontinuous Galerkin finite-element method for conservation-laws. 4. The multidimensional case. Math Comput 54(190):545–581

Cockburn B, Luskin M, Shu CW, Suli E (2003) Enhanced accuracy by post-processing for finite element methods for hyperbolic equations. Math Comput 72(242):577–606

Cockburn B, Gopalakrishnan J, Lazarov R (2009) Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems. SIAM J Numer Anal 47(2):1319–1365

Cushman-Roisin B (1987) Introduction to geophysical fluid dynamics. Prentice Hall, Upper Saddle River

Dawson C, Sun S, Wheeler MF (2004) Compatible algorithms for coupled flow and transport. Comput Methods Appl Mech Eng 193:2565–2580

Deleersnijder E, Lermusiaux PFJ (2008) Multi-scale modeling: nested grid and unstructured grid approaches. Ocean Dyn 58:335–336

Doney SC, Lima I, Moore JK, Lindsay K, Behrenfeld MJ, Mahowald TKWN, Glovera DM, Takahashi T (2009) Skill

metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data. J Mar Syst 76(1–2):95–112
Fasham MJR, Ducklow HW, McKelvie SM (1990) A nitrogen-based model of plankton dynamics in the oceanic mixed layer. J Mar Res 48(49):591–639
Fennel W, Neumann T (2004) Introduction to the modelling of marine ecosystems, Elsevier oceanography series, vol 72. Elsevier, New York
Ferziger JH, Peric M (2002) Computational methods for fluid dynamics, 3rd edn. Springer, New York
Flierl G, McGillicuddy DJ (2002) Mesoscale and submesoscale physical–biological interactions. The Sea 12:1–74
Geuzaine C, Remacle JF (2009) Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities. Int J Numer Methods Eng 79(11):1309–1331
Haidvogel DB, Beckmann A (1999) Numerical ocean circulation modeling. Imperial College, London. Distributed by World Scientific, River Edge
Haley PJ, Lermusiaux P, Robinson A, Leslie W, Logutov O, Cossarini G, Liang X, Moreno P, Ramp S, Doyle J, Bellingham J, Chavez F, Johnston S (2009) Forecasting and reanalysis in the Monterey Bay/California current region for the autonomous ocean sampling network-II experiment. Special issue on AOSN-II. Deep-sea Res Part II Top Stud Oceanogr 56:127–148
Haley PJ Jr, Lermusiaux PFJ (2010) Multiscale two-way embedding schemes for free-surface primitive-equations in the Multidisciplinary Simulation, Estimation and Assimilation System. Ocean Dyn (in this special issue). doi:10.1007/s10236-010-0349-4
Hanert E, Le Roux DY, Legat V, Deleersnijder E (2004) Advection schemes for unstructured grid ocean modelling. Ocean Model 7(1–2):39–58
Hecht MW, Holland WR, Rasch PJ (1995) Upwind-weighted advection schemes for ocean tracer transport—an evaluation in a passive tracer context. J Geophys Res Oceans 100(C10):20, 763–720, 778
Hesthaven JS, Kirby RM (2008) Filtering in Legendre spectral methods. Math Comput 77(263):1425–1452
Hesthaven JS, Warburton T (2008) Nodal discontinuous Galerkin methods. Texts in applied mathematics, vol 54. Springer, New York
Hofmann EE, Friedrichs MAM (2002) Predictive modeling for marine ecosystems. The Sea 12:537–565
Hofmann EE, Lascara CM (1998) Overview of interdisciplinary modelling for marine ecosystems. In: Brink KH, Robinson AR (eds) The sea, vol 10. Wiley, New York, pp 507–540
Hoteit H, Ackerer P, Mosé R, Erhel J, Philippe B (2004) New two-dimensional slope limiters for discontinuous Galerkin methods on arbitrary meshes. Int J Numer Methods Eng 61(14):2566–2593
Iskandarani M, Levin JC, Choi BJ, Haidvogel DB (2005) Comparison of advection schemes for high-order h-p finite element and finite volume methods. Ocean Model 10(1–2):233–252
Ji R, Davis C, Chen C, Beardsley R (2008) Influence of local and external processes on the annual nitrogen cycle and primary productivity on Georges Bank: a 3-D biological–physical modeling study. J Mar Syst 73(1–2):31–47
Karniadakis GE, Sherwin SJ (2005) Spectral/hp element methods for CFD, 2nd edn. Oxford University Press, New York
Krivodonova L (2007) Limiters for high-order discontinuous Galerkin methods. J Comput Phys 226(1):276–296
Kubatko EJ, Bunya S, Dawson C, Westerink JJ, Mirabito C (2009) A performance comparison of continuous and discontinuous finite element shallow water models. J Sci Comput 40:315–339
Lalli C, Parsons T (1997) Biological oceanography: an introduction, 2nd edn. Butterworth-Heinemann, Oxford
Lambrechts J, Remacle1 JF, Hillewaert K (2010) Efficient assembly of high order continuous and discontinuous finite element operators. In: IMUM2010—9th international workshop on multiscale (un)-structured mesh numerical modeling for coastal, shelf, and global ocean dynamics. http://mseas.mit.edu/IMUM2010/
Lermusiaux P (1999) Data assimilation via error subspace statistical estimation, part II: Mid-Atlantic bight shelfbreak front simulations, and ESSE validation. Mon Weather Rev 127(8):1408–1432
Lermusiaux PFJ, Xu J (2010) Coupled ocean-acoustic prediction of transmission loss in a continental shelfbreak region: predictive skill, uncertainty quantification and dynamical sensitivities. IEEE Trans J Ocean Eng. doi:10.1109/JOE.2010.2068611
Lermusiaux PFJ (2006) Uncertainty estimation and prediction for interdisciplinary ocean dynamics. J Comput Phys 176–199 (Special issue on "Uncertainty Quantification." J. Glimm and G. Karniadakis, Eds.)
Lermusiaux PFJ (2007) Adaptive modeling, adaptive data assimilation and adaptive sampling. Physica D 230:172–196 (Special issue on "Mathematical Issues and Challenges in Data Assimilation for Geophysical Systems: Interdisciplinary Perspectives." C.K.R.T. Jones and K. Ide, Eds.)
Lermusiaux PFJ, Haley PJ Jr, Leslie WG, Logutov OG (2009) Philippines strait dynamics experiment—MSEAS home page. http://mseas.mit.edu/Research/Straits/
Lévy M, Estublier A, Madec G (2001) Choice of an advection scheme for biogeochemical models. Geophys Res Lett 28(19):3725–3728
Logutov OG, Lermusiaux PFJ (2008) Inverse barotropic tidal estimation for regional ocean applications. Ocean Model 25:17–34
Lomax H, Pulliam TH, Zingg DW (2003) Fundamentals of computational fluid dynamics (scientific computation). Springer, New York
Lynch DR, McGillicuddy DJ, Werner FE (2009) Skill assessment for coupled biological/physical models of marine systems. J Mar Syst 76(1–2):1–3
Mavriplis CA (1989) Nonconforming discretization and a posteriori error estimators for adaptive spectral element techniques. Ph.D. thesis, MIT
MSEAS-Group (2010) Multidisciplinary simulation, estimation, and assimilation systems (http://mseas.mit.edu/, http://mseas.mit.edu/codes). Reports in ocean science and engineering 6, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts
Nguyen NC, Peraire J, Cockburn B (2009) An implicit high-order hybridizable discontinuous Galerkin method for linear convection–diffusion equations. J Comput Phys 228(9):3232–3254
Oschlies A, Garcon V (1998) Eddy-induced enhancement of primary production in a model of the North Atlantic Ocean. Nature 394:266–269

Pain CC, Piggott MD, Goddard AJH, Fang F, Gorman GJ, Marshall DP, Eaton MD, Power PW, de Oliveira CRE (2005) Three-dimensional unstructured mesh ocean modelling. Ocean Model 10(1–2):5–33

Pedlosky J (1987) Geophysical fluid dynamics, 2nd edn. Springer, New York

Persson PO, Peraire J (2006) Sub-cell shock capturing for discontinuous Galerkin methods. AIAA-2006-112

Persson PO, Strang G (2004) A simple mesh generator in MATLAB. SIAM Rev 46(2):329–345

Qiu J, Shu CW (2005) Runge–Kutta discontinuous Galerkin method using WENO limiters. SIAM J Sci Comput 26(3):907–929

Reed WH, Hill TR (1973) Triangular mesh methods for the neutron transport equations. Los Alamos Scientific Laboratory Report LA-UR-73-479, Fanstord University

Robinson AR, Lermusiaux PFJ (1999) Report of a workshop on the assimilation of biological data in coupled physical/ecosystem models. GLOBEC Special Contribution 3, Bologna, Italy

Robinson AR, McCarthy JJ, Rothschild BJ (2002) Biological–physical interactions in the sea. The Sea, vol 12. Wiley, New York

Rothstein LM, Cullen JJ, Abbott M, Chassignet EP, Denman K, Doney SC, Ducklow H, Fennel K, Follows M, Haidvogel D, Hoffman E, Karl DM, Kindle J, Lima I, Maltrud M, McClain C, McGillicuddy DJ, Olascoaga MJ, Spitz Y, Wiggert J, Yoder J (2006) Modeling ocean ecosystems: the paradigm program. Oceanography 19(1):22–51

Sapsis T, Lermusiaux P (2009) Dynamically orthogonal field equations for continuous stochastic dynamical systems. Physica D 238:2347–2360

Signell R (1989) Tidal dynamics and dispersion around coastal headlands. Ph.D. thesis, WHOI/MIT Joint Program

Slingo J, Bates K, Nikiforakis N, Piggott M, Roberts M, Shaffrey L, Stevens I, Vidale PL, Weller H (2009) Developing the next-generation climate system models: challenges and achievements. Philos Trans Royal Soc Math Phys Eng Sci 367(1890):815–831

Solin P, Segeth K, Dolezel I (2003) Higher-order finite element methods. Chapman & Hall/CRC, Boca Raton

Spitz Y, Allen J, Gan J (2005) Modeling of ecosystem processes on the Oregon shelf during the 2001 summer upwelling. J Geophys Res 110:C10S17

Stow CA, Jolliff J, McGillicuddy DJ, Doney SC, Allen JI, Friedrichs MAM, Rose KA, Wallhead P (2009) Skill assessment for coupled biological/physical models of marine systems. J Mar Syst 76(1–2):4–15

Strang G, Fix GJ (1973) An analysis of the finite element method. Prentice Hall, Englewood Cliffs

Ueckermann MP (2009) Towards next generation ocean models: novel discontinuous Galerkin schemes for 2D unsteady biogeochemical models. SM thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering

Ocean Dynamics (2010) 60:1447–1461
DOI 10.1007/s10236-010-0350-y

An accurate momentum advection scheme for a z-level coordinate models

Olga Kleptsova · Guus S. Stelling · Julie D. Pietrzak

Received: 30 April 2010 / Accepted: 18 October 2010 / Published online: 6 November 2010

Abstract In this paper, we focus on a conservative momentum advection discretisation in the presence of z-layers. While in the 2D case conservation of momentum is achieved automatically for an Eulerian advection scheme, special attention is required in the multi-layer case. We show here that an artificial vertical structure of the flow can be introduced solely by the presence of the z-layers, which we refer to as the staircase problem. To avoid this staircase problem, the z-layers have to be remapped in a specific way. The remapping procedure also deals with the case of an uneven number of layers adjacent to a column side, thus allowing one to simulate flooding and drying phenomena in a 3D model.

Keywords Shallow water equations · Advection · Flooding and drying · Momentum conservation · z-layer · Staircase problem

1 Introduction

An important aspect in 3D ocean modelling is the choice of the vertical coordinate system. Three distinct types of vertical coordinates are, in general use, geopotential z-level coordinates, terrain-following σ coordinates and isopycnal coordinates. There is no single coordinate system that is suitable for all applications, as each of these coordinate systems has its own advantages and disadvantages. Here we are interested in coastal ocean modelling and, in particular, in accurately simulating flows in the shallow near shore region, including flooding and drying problems. It is in this region that exchange between the land and ocean occurs; it is a region of increasing importance to oceanographers.

Isopycnal coordinates have been successfully used in large-scale ocean models, such as Miami Isopycnic Coordinate Ocean Model (Bleck et al. 1992) and Hallberg Isopycnic Model (Hallberg 1997). These are capable of retaining sharp interfaces and fronts. In the coastal regions, however, the layers of predefined constant density must be able to collapse into one layer under well-mixed conditions and to inflate into a number of layers under stratified conditions. Together with weak ability to simulate surface and bottom boundary layers, this limits the use of isopycnal coordinates in coastal models.

The main advantage of the σ-coordinate system is the fact that it is fitted to both the moving free surface and bottom topography. This allows one to accurately approximate the vertical flow distribution without a large number of vertical grid points. The terrain-following coordinates allow an efficient grid refinement near the free surface and the bed, which makes it easy to resolve the boundary layers. However, a systematic error in the calculation of the baroclinic pressure gradient terms can arise in regions with steep topography and sharp density gradients when using σ-coordinates. A number of methods exist to reduce the pressure gradient error; some of them are summarised in Stelling and Van Kester (1994) and Kliem and Pietrzak (1999).

Responsible Editor: Vincent Legat

O. Kleptsova (✉) · G. S. Stelling · J. D. Pietrzak
Faculty of Civil Engineering and Geosciences,
Delft University of Technology, Stevinweg 1,
2628CN Delft, The Netherlands
e-mail: O.Kleptsova@TUDelft.nl

Efforts are now underway to create generalised and hybrid coordinate models, such as Hybrid Coordinate Ocean Model (Bleck 2002; Burchard and Petersen 1997), that can employ appropriate coordinates in different regions. Alternative vertical coordinate systems are also being explored. Adcroft and Campin (2004) propose a rescaled height coordinate system which is essentially height-based but shares some similarity with σ-coordinates. Halyer and Lermusiaux (2010) employ time-dependent, terrain-following coordinates. They first define a set of terrain-following depths for the mean sea level and then a set of time variable model depths such that the change in cell thickness is proportional to the relative thickness of the original (undisturbed) cell.

Geopotential z-level coordinates do not suffer from the pressure gradient error associated with σ-coordinates. The main problem with z-level models are connected with flow along a sloping bottom and surface. In this case, the stepwise discontinuous representation of the topography and free surface can generate false flow structures, which we refer to here as the staircase problem. The treatment of the bottom topography can be improved by using a finite volume discretisation which allows one to use variable bottom layer thickness (partial cells) or shaped volumes ("shaved" cells) as proposed by Adcroft et al. (1997). However, as shown here, this is not the complete solution.

The z-coordinates are referenced to a time mean water level. The free-surface displacement moves relative to this coordinates system and is to be implemented as a moving boundary. In Griffies et al. (2001) and Campin et al. (2004), this is done by allowing the top model layer to vary in thickness. The free-surface variation, however, must be smaller than that of the top layer thickness. This becomes a serious limitation with increasing vertical resolution and in shallow regions where extensive flooding and drying can take place. A model can be coded to allow the top layer to vanish as it is done in Delfin (Ham et al. 2005) or to become dry as in SUNTANS (Fringer et al. 2006). The second layer then takes on the role of the surface layer with variable thickness. The major difficulty here is to make the transition of a vanishing layer smooth enough to avoid the staircase problem which leads to the generation of false currents, stability and conservation problems.

In a large-scale hydrostatic model, a vertical structure of the flow must be created only due to physical mechanisms such as sheer stress or complex bottom profiles yielding flow separation. Without that, the results of 2D and 3D models should be identical. Therefore, special attention is required for the discretisation of the momentum equation in the 3D case.

The staircase problem can be avoided by using a semi-Lagrangian advection scheme, such as the schemes by Casulli and Walters (2000) and Ham et al. (2005). This approach is, however, unable to provide momentum conservation and, thus, it is not suitable for the modelling of rapidly varied shallow water flows as typically found in flooding situations, for example, dam break problems or tsunamis. Other schemes, such as the schemes of Fringer et al. (2006) and Stuhne and Peltier (2009) both employing the Eulerian advection scheme by Perot (2000), have the staircase problem. Indeed, the model of Fringer et al. (2006) does not conserve momentum in the cells that contain the free surface. SUNTANS is, however, mainly used for internal gravity wave simulations, and the free surface dynamics are not their main concern. Stuhne and Peltier (2009) in their 3D global M_2 tide simulation note deterioration of results in the coastal regions compared to their 2D simulation.

In this paper, we show how we can design an Eulerian advective approximation which solves the staircase problem and prevents the model from creating an artificial vertical structure. To show the strength of this approach, we even apply it to dam break problems. Although the staircase problem is not related to the choice of the horizontal discretisation as such, in the C-grid class of models, an artificial vertical structure can also be created due to the Coriolis force, since in this class of models a tangential velocity reconstruction has to be used.

The paper is organised as follows: the shallow water equations are described in Section 2 and the general model description is given in Section 3. Section 4 describes the two discretisations of the advection term, as proposed by Perot (2000) and Kramer and Stelling (2008). The scheme of Kramer and Stelling (2008) is generalised for the multi-layer case. In Section 5, we compare the behaviour of the advection schemes using a number of 2D test cases. In Section 6, we show how an artificial vertical structure in the flow can be created. Next we propose a local layer remapping procedure which allows one to remove the staircase problem thus preventing the model from unphysical behaviour. The proposed procedure also deals with the vanishing layer case, thus allowing one to simulate flooding and drying phenomena in the presence of multiple z-layers. Conclusions are given in Section 7.

2 Basic equations

Coastal flows may be modelled using the shallow water equations, a well-known simplification of the

Navier–Stokes equations. The shallow water equations are derived under the following assumptions about the flow:

- Vertical accelerations are small compared with gravity, that is, the pressure p is assumed to be hydrostatic.
- Density differences are small compared to the reference density ρ_0, and hence, the density ρ may be assumed to be constant except in the pressure gradient term.

In this paper, we do not consider the effects of density differences so the density cancels out of the pressure gradient term. In common with similar models (Casulli and Walters 2000; Baptista et al. 2005), we also disregard horizontal mixing.

Let $\mathbf{d}$ be any unit vector in the horizontal plane and let $u_{\mathbf{d}} = \mathbf{u} \cdot \mathbf{d}$. Then the shallow water equations may be written as follows:

$$\nabla \cdot \mathbf{u} = 0$$

$$\frac{\partial u_{\mathbf{d}}}{\partial t} + \mathbf{u} \cdot \nabla u_{\mathbf{d}} + g \nabla_{xy} \eta \cdot \mathbf{d} + \frac{\partial}{\partial z} \nu^v \frac{\partial u_{\mathbf{d}}}{\partial z} + (2\Omega \times \mathbf{u}) \cdot \mathbf{d} = 0 \tag{1}$$

where $\mathbf{u}$ denotes velocity vector, η is the free surface elevation, ν^v is the vertical viscosity and Ω is the Earth's rotation vector.

Assuming bottom impermeability, the normal component of the velocity at the sea bed must vanish. This is expressed by the following kinematic condition:

$$w = \frac{Db}{Dt} = \mathbf{u}^{xy} \cdot \nabla_{xy} b \qquad \text{at } z = b \tag{2}$$

where $b(x, y)$ is the bottom height. Here we assume that the vertical coordinate z is pointing upwards and has its origin at the mean sea level. Thus, the total water depth $h(x, y, t)$ is defined as $h(x, y, t) = \eta(x, y, t) - b(x, y)$.

The kinematic condition at the free-surface is given by

$$w = \frac{D\eta}{Dt} = \frac{\partial \eta}{\partial t} + \mathbf{u}^{xy} \cdot \nabla_{xy} \eta \qquad \text{at } z = \eta \tag{3}$$

Here $\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla$ is the Lagrangian or material derivative with ∇ being the gradient operator. In this case, a 2D field is encountered; we write ∇_{xy} to indicate the 2D gradient operator.

The kinematic boundary conditions 2–3 may be used in combination with the continuity equation integrated over the water column to produce the following expression for the free surface:

$$\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x} \left[\int_b^\eta u dz \right] + \frac{\partial}{\partial y} \left[\int_b^\eta v dz \right] = 0 \tag{4}$$

3 Model description

We consider a discretisation on triangular meshes for which the component of velocity normal to each mesh edge is stored at the centre of the edge and the surface elevation is stored at the circumcentre of each triangle. The continuity equation is discretised using a finite-volume scheme with a single triangle as the control volume, while the pressure gradient term is discretised using a form of finite-difference scheme.

The following grid notations are adopted (see Fig. 1). We use index c to refer to a column of the grid and index f to refer to a column side. The index k is reserved to refer to the layer of a given cell. The layer of a horizontal face above level k is $k + \frac{1}{2}$, and the layer below is naturally $k - \frac{1}{2}$. This numbering is physically consistent in that $(k + 1) - \frac{1}{2} = k + \frac{1}{2}$. The length of the column side f is referred to as l_f. A face belonging to a column side f and lying on a vertical level k we denote as f_k. Similarly, c_k denotes the cell at a vertical level k of the column c. Height of the face f_k is denoted as $h_{f,k}$, whereas h_f refers to the total water depth at the column side f. Similarly, h_c and $h_{c,k}$ refer to the water depth at the centre of the column c and to the height of the cell c_k in this column. A_c denotes the water column cross-section area, that is, the area of the (triangular) base of the column c.

Following Casulli and Walters (2000), the model treats the barotropic pressure gradient, the vertical viscosity in the momentum equations and the divergence term in the continuity equation implicitly and all other terms explicitly. The explicit treatment of the advection and Coriolis terms makes the velocity submatrix block tridiagonal which can be inverted using fast direct methods. This allows efficient elimination of the velocity variables from the continuity equation resulting in a linear implicit system for the free surface elevation.

We choose a semi-implicit θ-scheme as the temporal discretisation of the free surface equation. The momentum equation is solved for the velocity component normal to each cell face. The advection and Coriolis operators are dealt explicitly relative to the pressure term. For the time evolution of the pressure gradient, we adopt the same θ-method chosen for the free surface

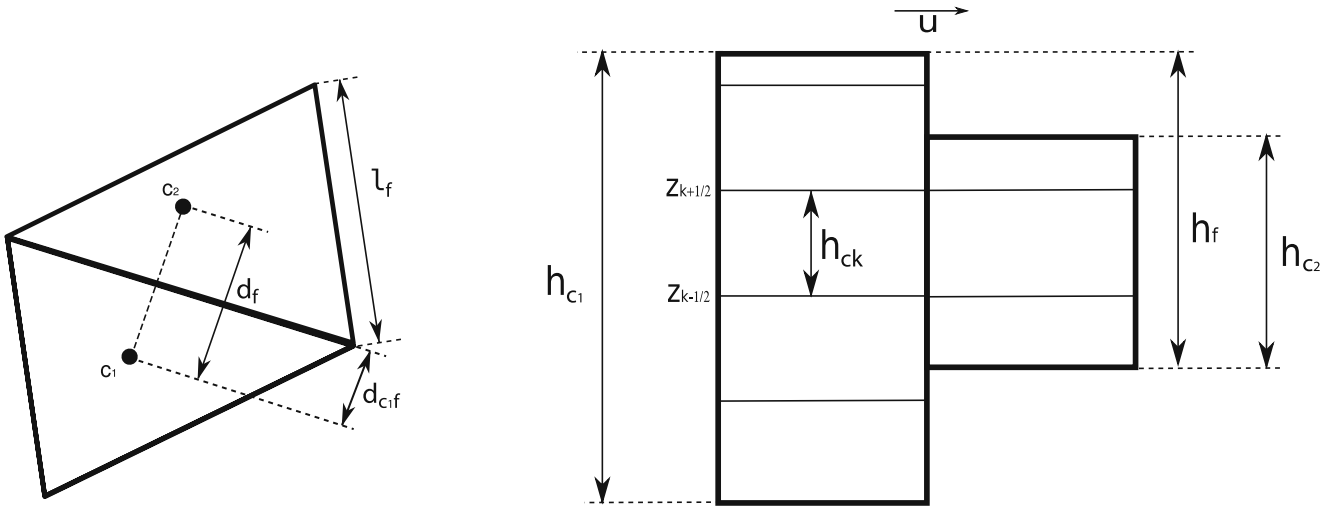

Fig. 1 Grid fragment together with the grid notations

equation. A detailed description of the model can be found in Ham et al. (2005).

Friction terms are currently disregarded. Thus, we arrive at the following set of discrete equations:

$$A_c\eta_c^{n+1} = A_c\eta_c^n - \theta\Delta t\sum_f\left(s_{f,c}l_f\sum_k h_{f,k}^n u_{f,k}^{n+1}\right) - (1-\theta)\Delta t\sum_f\left(s_{f,c}l_f\sum_k h_{f,k}^n u_{f,k}^n\right) = 0$$

$$u_{f,k}^{n+1} = u_{f,k}^n + \Delta t a_{f,k}^n + F(\mathbf{u})_{f,k}^n - g\Delta t\left(\theta G_f^{n+1} + (1-\theta)G_f^n\right) \quad (5)$$

where $u_{f,k}^n$ is the horizontal velocity component in the direction $\mathbf{n}_f$ normal to the f-th column side of the grid, at the time level n.

Function $s_{f,c}$ is defined for each column side and column such that it is equal to 0 if column c does not contain column side f, 1 if it does and the normal $\mathbf{n}_f$ at f is the outer one with respect to c, and it is -1 if the normal is the inner one.

Here G is a linear operator for pressure gradient term. The operators a and F are explicit operators which account for the contribution from the discretisation of the momentum advection and Coriolis force, respectively. We use the 3D order Adams–Bashforth scheme for time integration of the Coriolis term; the momentum advection term is integrated using explicit Euler scheme.

The C-grid models only solve for the component of velocity normal to a face. Therefore, in order to evaluate the Coriolis term, it is necessary to interpolate the tangent velocity at the centre of a cell face as a linear combination of normal velocity components of the nearby faces. Following Perot (2000), first the full velocity vector $\mathbf{u}_{c,k}$ in the cell interior is reconstructed. Next the velocity vector $\mathbf{u}_{f,k}$ at the face centre is reconstructed by taking a linear combination of the velocity vectors located at the centres of the two cells adjacent to the face.

$$\mathbf{u}_{c,k} = \sum_{f_k}\delta_{f_k,c_k}d_{f,c}\frac{h_{f,k}l_f}{h_{c,k}A_c}u_{f,k}\mathbf{n}_f \quad (6)$$

$$\mathbf{u}_{f,k} = \sum_{c_k}\delta_{f_k,c_k}\frac{d_{f,c}}{d_f}\mathbf{u}_{c,k} \quad (7)$$

The function δ_{f_k,c_k} is defined for each face and cell such that it is equal to 0 if cell c_k does not contain face f_k and 1 if it does. A detailed description of the spacial discretisation of the Coriolis term we use is given in Kleptsova et al. (2009). In the section below, we focus on the spacial discretisation of the advection operator a.

To implement flooding and drying, it is a common practice to mask a column c as dry if its water depth h_c becomes less than a threshold value. Following Stelling and Duinmeijer (2003), we define water depth h_f at a column side f using the first-order upwinding as follows:

$$h_f = \begin{cases} h_{c1} \text{ if } s_{f,c1}u_f > 0 \\ h_{c2} \text{ if } s_{f,c2}u_f < 0 \\ \max(\eta_{c1},\eta_{c2}) - \max(b_{c1},b_{c2}) \text{ if } u_f = 0 \end{cases} \quad (8)$$

The column side f is then masked as dry and the velocity u_f at the column side is set to zero once the water depth h_f becomes less than a prescribed threshold value $h_{\min}$. Thus, the water depth used to calculate

outgoing fluxes of a column c is that of the column itself. Therefore, there will never be a flux out of a dry column. Stelling and Duinmeijer (2003) show that non-negative water depth is ensured if:

$$\frac{\Delta t u_f^n}{\Delta x} \leq 1 \tag{9}$$

4 Advection

In this section, we consider two Eulerian advection schemes. The first scheme was proposed by Perot (2000) for Navier–Stokes equations. The second scheme based on the scheme by Perot (2000) was derived by Kramer and Stelling (2008) for the depth-integrated and depth-averaged shallow water equations.

4.1 Advection discretisation by Perot (2000)

Following Perot (2000), first a finite-volume discretisation $\mathbf{a}_{c,k}$ of the cell-based advection term is derived by integrating the vector $\nabla \cdot (\mathbf{uu}) = \mathbf{u} \cdot [\nabla u, \nabla v, \nabla w]$ over a cell using the Gauss' theorem

$$\int_{\Omega^3} \nabla \cdot (\mathbf{uu}) dV = \int_{\partial\Omega} \mathbf{u}(\mathbf{u} \cdot \mathbf{N}) dS \tag{10}$$

Assuming a prismatic cell with the base area A_c and height $h_{c,k}$, this yields

$$\begin{aligned} A_c h_{c,k} \mathbf{a}_{c,k} = & \sum_{f_k} s_{f,c} h_{f,k} l_f u_{f,k} \mathbf{u}_{f,k} \\ & + \left[A_c \mathbf{u}_{c,k+\frac{1}{2}} \left(\mathbf{u}_{c,k+\frac{1}{2}} \cdot \mathbf{n}_{c,k+\frac{1}{2}} \right) \right. \\ & \left. - A_c \mathbf{u}_{c,k-\frac{1}{2}} \left(\mathbf{u}_{c,k-\frac{1}{2}} \cdot \mathbf{n}_{c,k-\frac{1}{2}} \right) \right] \end{aligned} \tag{11}$$

where $\mathbf{n}_{c,k\pm\frac{1}{2}}$ are the upward pointing vectors normal to the "horizontal" faces of the cell and attached at the centres of the faces. The face velocity vectors $\mathbf{u}_{f,k}$ are interpolated as given in Eqs. 6 and 7. The velocity vectors $\mathbf{u}_{c,k\pm\frac{1}{2}}$ attached at the centres of the "horizontal" faces of the cell are interpolated as follows:

$$\begin{aligned} \mathbf{u}_{c,k-\frac{1}{2}} &= \frac{h_{c,k-1}}{h_{c,k-1}+h_{c,k}} \mathbf{u}_{c,k-1} + \frac{h_{c,k}}{h_{c,k-1}+h_{c,k}} \mathbf{u}_{c,k} \\ \mathbf{u}_{c,k+\frac{1}{2}} &= \frac{h_{c,k}}{h_{c,k}+h_{c,k+1}} \mathbf{u}_{c,k} + \frac{h_{c,k+1}}{h_{c,k}+h_{c,k+1}} \mathbf{u}_{c,k+1} \end{aligned} \tag{12}$$

Observe that the velocity component $\mathbf{u} \cdot \mathbf{n}$ normal to the "horizontal" faces of a cell is not, in general, equal to the vertical velocity component w, as can be seen from Fig. 2.

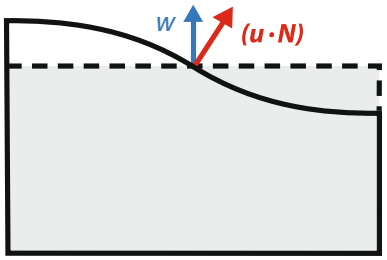

Fig. 2 A water column top cell together with vertical and normal to the surface velocities

Integrating the 3D continuity equation in vertical from bottom to a vertical level $z_{k+\frac{1}{2}}$ and applying kinematic boundary condition 2 at the bed, we arrive at the following expression for the velocity normal to the "horizontal" faces of a cell:

$$\left(\mathbf{u}_{c,k+\frac{1}{2}} \cdot \mathbf{n}_{c,k+\frac{1}{2}} \right) = -\frac{1}{A_c} \sum_{i=k_b}^{k} s_{f,c} h_{f,i} l_f u_{f,i} \tag{13}$$

where k_b is the index of the column's bottom layer.

Next the face normal component of the advection term is reconstructed out of a given set of the cell-based vectors $\mathbf{a}_c$, by taking the following linear combination:

$$a_{f,k} = \sum_{c_k} \delta_{f_k,c_k} \alpha_{f,c} (\mathbf{a}_{c,k} \cdot \mathbf{n}_f) \tag{14}$$

The weighting coefficients $\alpha_{f,c}$ are defined by Perot (2000) as

$$\alpha_{f,c} = \frac{d_{f,c}}{d_f} \tag{15}$$

Other definitions of the weighting factors are possible (see, for example, Wenneker et al. 2002). Influence of the weighting factors on the performance of the depth-integrated scheme is examined in Kramer and Stelling (2008).

Thus, the advection component normal to the face j_k shown in Fig. 3 can be written as

$$\begin{aligned} a_{j,k} = & \sum_{c_k} \delta_{j_k,c_k} \alpha_{j,c} \frac{1}{A_c h_{c,k}} \\ & \cdot \left[\sum_{f_k} s_{f,c} h_{f,k} l_f u_{f,k} (\mathbf{u}_{f,k} \cdot \mathbf{n}_j) \right. \\ & - \left(\left(\mathbf{u}_{c,k+\frac{1}{2}} \cdot \mathbf{n}_j \right) \sum_{i=k_b}^{k} s_{f,c} h_{f,i} l_f u_{f,i} \right. \\ & \left. \left. - \left(\mathbf{u}_{c,k-\frac{1}{2}} \cdot \mathbf{n}_j \right) \sum_{i=k_b}^{k-1} s_{f,c} h_{f,i} l_f u_{f,i} \right) \right] \end{aligned} \tag{16}$$

In single layer case ($k = 1$), the horizontal velocity field is assumed to be constant in the vertical. In this

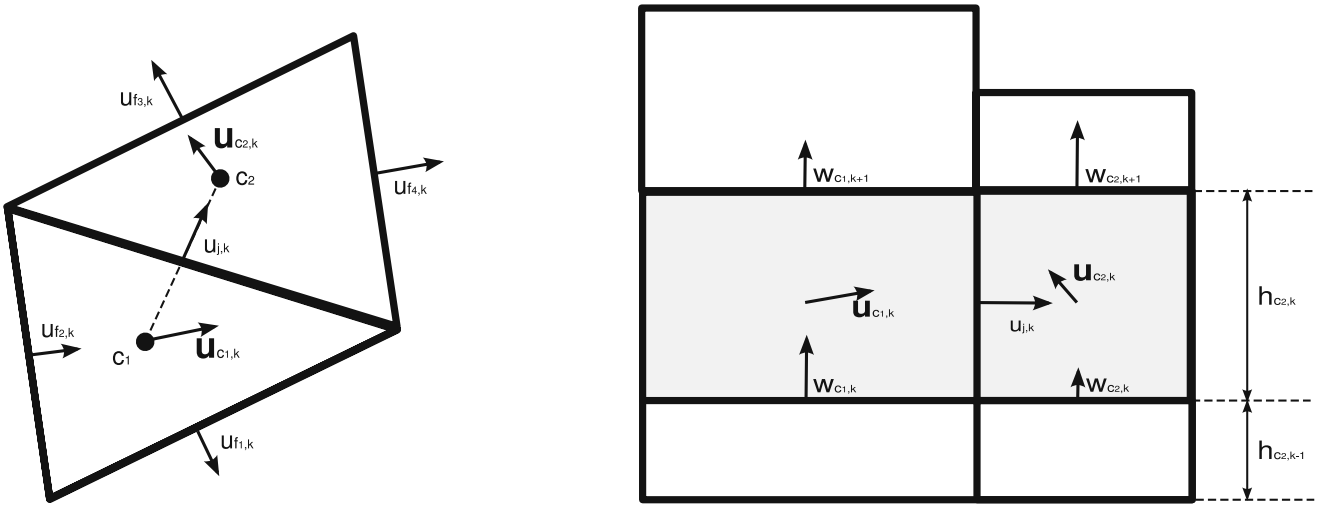

Fig. 3 The control volume (*shaded area*) used to discretise the horizontal momentum equation at face j, showing the plan form (*left*) and side (*right*) views

case, cells c_1 and faces f_1 correspond to columns c and column sides f. Therefore, height of the cells and faces become equal to the water depth at the respective columns and column sides. Thus, the Eq. 16 reduces to

$$a_j = \sum_c \delta_{j,c}\alpha_{j,c} \sum_f s_{f,c}\frac{h_f l_f}{A_c h_c} u_f(\mathbf{u}_f - \mathbf{u}_c)\cdot \mathbf{n}_j \tag{17}$$

4.2 Advection discretisation by Kramer and Stelling (2008)

In Kramer and Stelling (2008), another discretisation of the advection term a_j for the depth-averaged equations (i.e. for one-layer case) is presented. This scheme is a combination of the unstructured grid variant of the advection scheme by Stelling and Duinmeijer (2003) with the advection scheme by Perot (2000) described above.

$$a_j = \sum_c \delta_{j,c}\alpha_{j,c} \sum_f s_{f,c}\frac{h_f l_f}{A_c \bar{h}_j} u_f(\mathbf{u}^*_f \cdot \mathbf{n}_j - u_j) \tag{18}$$

The water depth $\bar{h}_j$ at the column side (face) j is defined as

$$\bar{h}_j = \sum_c \alpha_{j,c} h_c \tag{19}$$

The vector $\mathbf{u}^*_f$ is the full velocity vector at the column side (face) f reconstructed out of the velocity components from the column (cell) c^* upwind of column side (face) f as given in Eq. 6, that is,

$$\mathbf{u}^*_f = \mathbf{u}_{c*} \tag{20}$$

If the upwind column (cell) is the one containing the column side (face) j, $\mathbf{u}^*_f \cdot \mathbf{n}_j$ is approximated as u_j. Thus, the fluxes going out of the column (cell) can be omitted without changing of momentum.

This scheme was originally derived applying the Perot (2000) scheme to the depth-integrated velocities $h\mathbf{u}$ and rewriting it for the depth-averaged velocities. This was done using the equivalence between the depth-integrated and the depth-averaged momentum equations and the fact that time derivative of $h\mathbf{u}$ can be split into a contribution from the change in water volume and a contribution from the change in velocity.

In the case of multiple layers, the contribution from the change in velocity corresponds to the contribution from the vertical faces of the cell; the contribution from the change in water volume corresponds to the contribution from the "horizontal" faces. Thus, to obtain a multi-layer version of scheme by Kramer and Stelling (2008), we can use the following cell-based advection vector $\mathbf{a}_{c,k}(j_k)$ calculated for a face j_k of the cell c_k (the analogue of Eq. 11)

$$\begin{aligned} A_c \bar{h}_{j,k}\mathbf{a}_{c,k}(j_k) = &\sum_{f_k} s_{f,c} h_{f,k} l_f u_{f,k} \mathbf{u}^*_{f,k} \\ &+ A_c \mathbf{u}_{j,k+\frac{1}{2}}\left(\mathbf{u}_{c,k+\frac{1}{2}}\cdot \mathbf{n}_{c,k+\frac{1}{2}}\right) \\ &- A_c \mathbf{u}_{j,k-\frac{1}{2}}\left(\mathbf{u}_{c,k-\frac{1}{2}}\cdot \mathbf{n}_{c,k-\frac{1}{2}}\right) \end{aligned} \tag{21}$$

with

$$\begin{aligned} \mathbf{u}_{j,k-\frac{1}{2}} &= \frac{\bar{h}_{j,k-1}}{\bar{h}_{j,k-1}+\bar{h}_{j,k}}\mathbf{u}_{j,k-1} + \frac{\bar{h}_{j,k}}{\bar{h}_{j,k-1}+\bar{h}_{j,k}}\mathbf{u}_{j,k} \\ \mathbf{u}_{j,k+\frac{1}{2}} &= \frac{\bar{h}_{j,k}}{\bar{h}_{j,k}+\bar{h}_{j,k+1}}\mathbf{u}_{j,k} + \frac{\bar{h}_{j,k+1}}{\bar{h}_{j,k}+\bar{h}_{j,k+1}}\mathbf{u}_{j,k+1} \end{aligned} \tag{22}$$

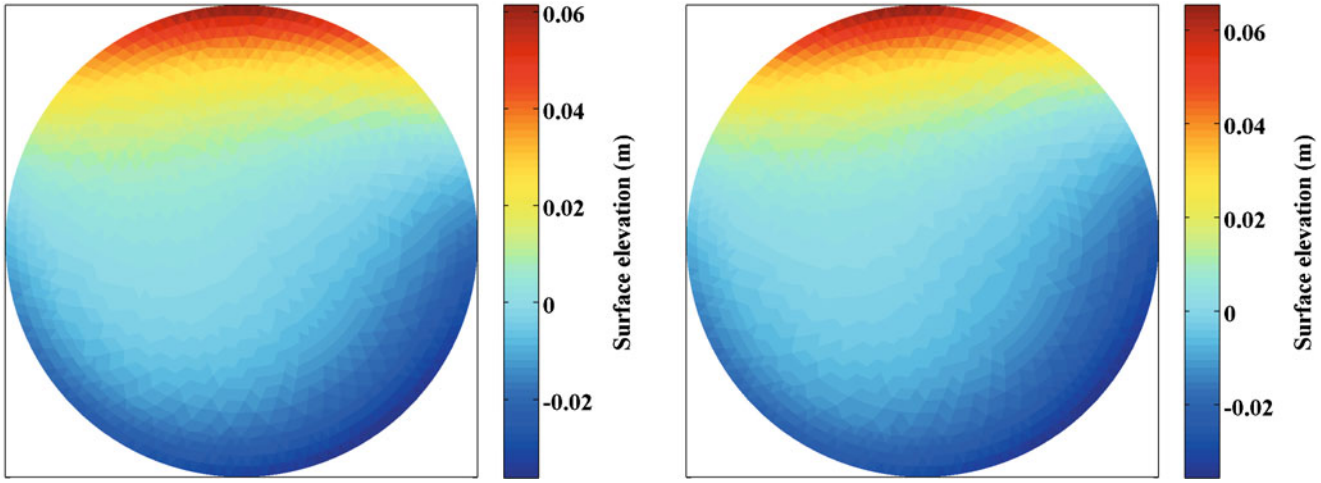

Fig. 4 Kelvin wave in a circular basin: the sea surface elevation after 1,666.6 h calculated using the advection scheme by Perot (2000) (*left*) and Kramer and Stelling (2008) (*right*)

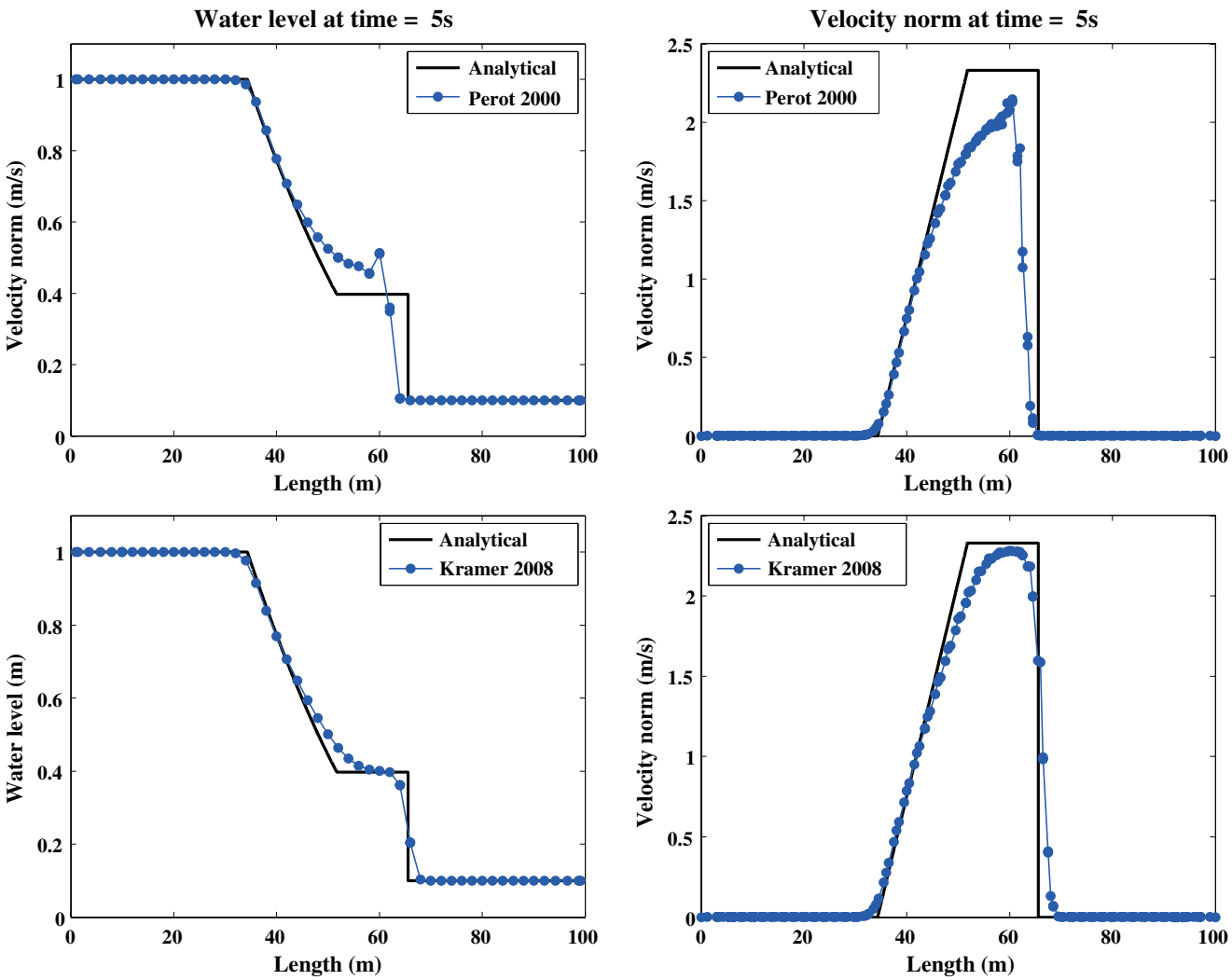

Fig. 5 Dam break over wet bed: surface elevation and velocity calculated using the advection schemes given by Eq. 17 (*top row*) and Eq. 18 (*bottom row*) compared with the analytical solution (*black line*)

Thus, the multi-layer version of the advection scheme 18 can be written as follows:

$$a_{j,k} = \sum_{c_k} \delta_{j_k,c_k} \alpha_{j,c} \sum_{f_k} s_{f,c} \frac{1}{A_c \bar{h}_{j,k}} \left[h_{f,k} l_f u_{f,k} (\mathbf{u}^*_{f,k} \cdot \mathbf{n}_j) - \left(u_{j,k+\frac{1}{2}} \sum_{i=k_b}^{k} h_{f,i} l_f u_{f,i} - u_{j,k-\frac{1}{2}} \sum_{i=k_b}^{k-1} h_{f,i} l_f u_{f,i} \right) \right] \tag{23}$$

with k_b being the index of the column's bottom layer.

The advection schemes described above will be compared using a number of 2D test cases in the section below.

5 Test cases

All of the simulations described in this section are performed with one layer in vertical, that is, the advection term is discretised according to Eqs. 17 and 18.

5.1 Kelvin wave test case

If the water depth is (locally) close to uniform, then the water depth at a column side (height of a face) is approximately equal to that at the neighbouring columns (cells). Similarly, a cell (column)-based velocity vector projected in the direction normal to a face of the cell (normal to a side of the column) is approximately equal

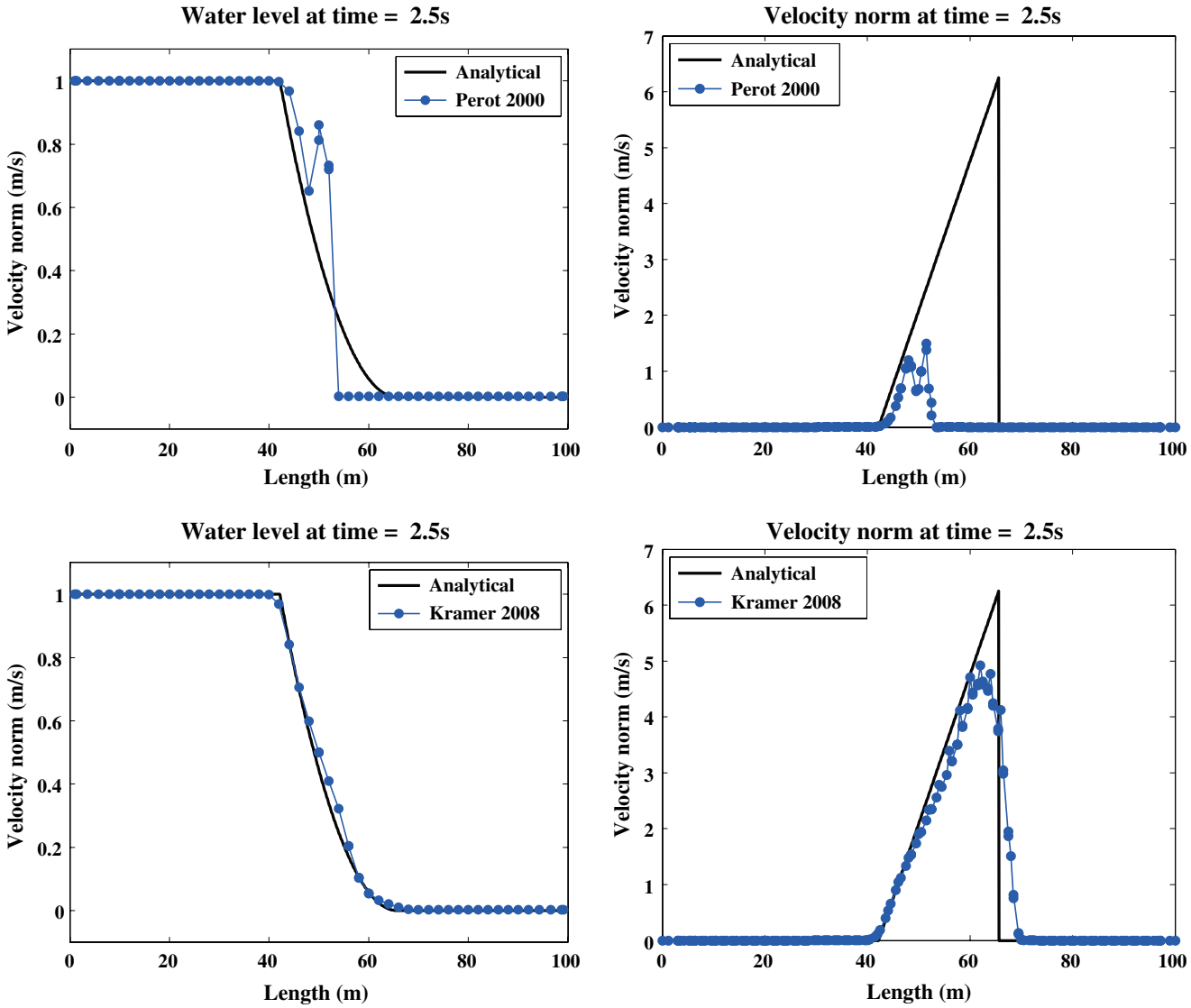

Fig. 6 Dam break over dry bed: surface elevation and velocity calculated using the advection schemes given by Eq. 17 (*top row*) and Eq. 18 (*bottom row*) compared with the analytical solution (*black line*)

to the face (column side) normal velocity component if the flow velocity is (locally) close to uniform. Thus, the advection schemes described above will behave similarly in the test cases with no sudden changes in water depth and flow velocity.

To illustrate this, we use a Kelvin wave in a shallow circular basin test case, described in Ham et al. (2005). The basin was given a uniform depth of 5 m. The initial state chosen was:

$$\eta(r,\theta) = 0.05e^{(r-r_0)/L_D}\cos\theta$$
$$u_\theta(r,\theta) = 0.05\sqrt{g/h}e^{(r-r_0)/L_D}\cos\theta$$
$$u_r(r,\theta) = 0 \tag{24}$$

where L_D is the Rossby radius, in this case approximately 68 km, and $r_0 = 250$ km is the basin radius. In the limiting case of an infinitely large basin, this is the expression for a Kelvin wave of amplitude 5 cm. The simulation was performed at a specified latitude of 45°. Figure 4 shows the sea surface elevation after 1,666.6 h. As expected, the results obtained using different advection schemes are identical up to visible precision.

5.2 Dam break test case

A dam break is calculated in a 100-m-long and 10-m-wide channel. At $t = 0$, the shock starts at $x = 50$ m with zero initial velocity and the upstream water level of 1 m. For the dam break over wet bed case, the water level downstream is 0.1 m. The nominal triangle edge length of the grid used is 2 m. The time step of 0.01 s was used for the simulation of the dam break over wet bed. The numerical solutions for the interval $y \in (4, 6)m$ compared to the analytical solution are shown in Fig. 5. For the simulation of the dam break over dry bed, the time step was set to 0.001 s. Threshold value $h_{\min}$ used to mask columns as dry was set to 2.5 mm. Comparison of the numerical solution for the interval $y \in (4, 6)m$ to the analytical solution is shown in Fig. 6.

In both cases, the results obtained using the advection scheme similar to Kramer and Stelling (2008) given by Eqs. 18 and 23 show better agreement with the analytical solution than the scheme by Perot (2000) given by Eqs. 16 and 17. Besides that, the advection scheme by Kramer and Stelling (2008) leads to a smoother solution due to the first-order upwinding.

5.3 Tsunami run-up on a plane beach

Here we examine run-up and run-down motion of a tsunami-type transient wave onto a plane sloping beach with slope $s = 1/10$. The initial free surface shown in Fig. 7 is specified according to

$$\eta = 500\big(0.006e^{-0.4444(\frac{x}{5{,}000}-4.1209)^2} - 0.018e^{-4.0(\frac{x}{5{,}000}-1.6384)^2}\big) \tag{25}$$

which corresponds to the leading depression N-wave shape, typically caused by an offshore submarine landslide. This case corresponds to the case D in Carrier et al. (2003), who derived a general semi-analytic solution for such events based on nonlinear shallow water equations.

Upon the release of the initial wave form, tsunami-type waves propagate in both landward and offshore directions, though only the landward travelling wave runs up the beach. At the offshore boundary, a closed boundary condition is imposed. Since the computational domain is sufficiently long (50 km), this does not affect the run-up process on the other side of the domain.

Simulation was performed on an 8,681-node grid, with resolution varying from 8 m in the shallow region to 318 m in the deep water. The time step was set to 1 s. The columns were masked as dry if their water depth was less then $h_{\min} = 1$ cm. The initial velocity is set to zero everywhere. Figure 8 shows the comparison of the computed surface elevation against the analytical solution for $t = 160$, 175 and 220 s.

Once again, the results obtained using the advection scheme similar to Kramer and Stelling (2008) given by Eqs. 18 and 23 agree with the analytical solution

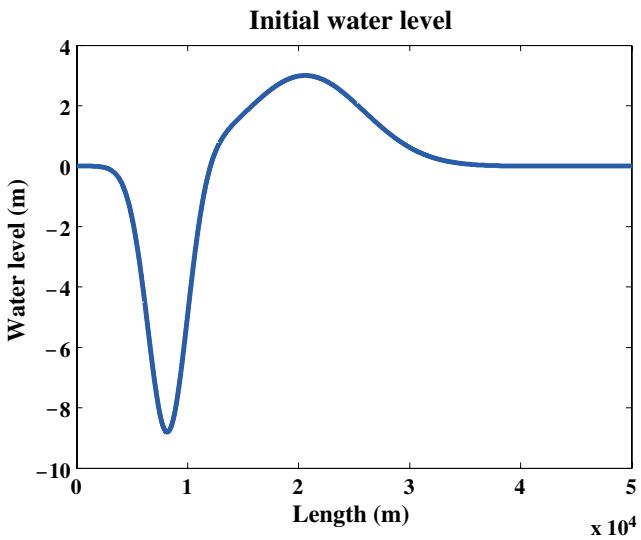

Fig. 7 Tsunami run-up on a plane beach: a portion of the initial surface elevation given by Eq. 25 used in the tsunami run-up simulation

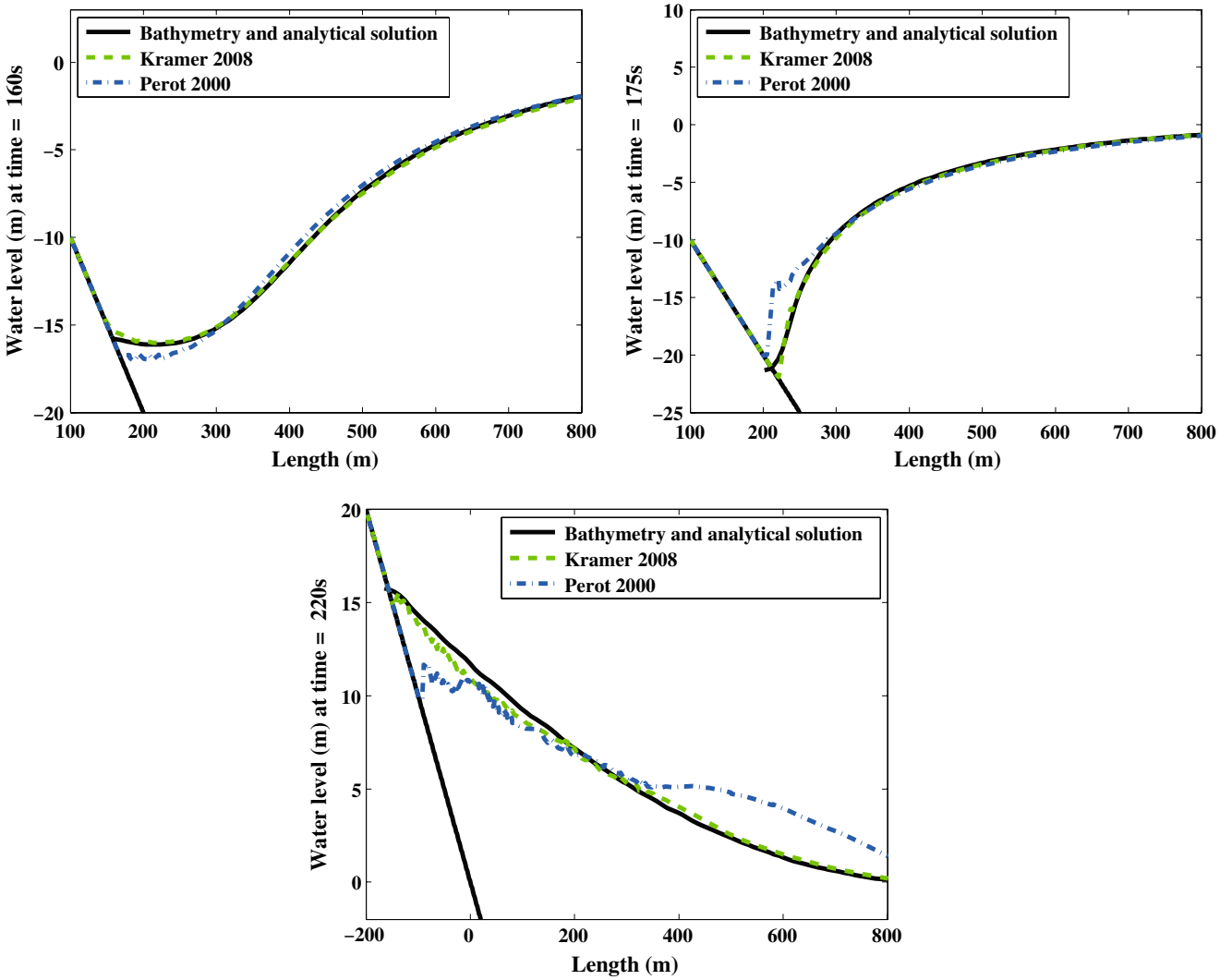

Fig. 8 Tsunami run-up on a plane beach: comparison of the computed surface elevation profiles against the analytical prediction for $t = 160$ s (*top*), $t = 175$ s (*middle*) and $t = 220$ s (*bottom*)

better than the ones obtained using the scheme by Perot (2000) given by Eqs. 16 and 17.

5.4 Parabolic flood wave

A water mass with a parabolic shape given by

$$\eta_0 = h_0 \left(1 - \frac{x^2 + y^2}{R_0^2}\right) \tag{26}$$

is released on a flat bed without friction. Here $R_0 = 50$ km and $h_0 = 2$ km are the initial radius and the initial height of the water mass, respectively. The analytical solution of the test is given by (see Thacker 1981)

$$\eta = h_0 \left[\frac{T^2}{t^2 + T^2} - \frac{x^2 + y^2}{R_0^2}\left(\frac{T^2}{t^2 + T^2}\right)^2\right] \tag{27}$$

Here

$$T = \frac{R_0}{\sqrt{2gh_0}} = 250 \text{ s} \tag{28}$$

is the time after which the initial height h_0 has been halved. Initially, the water mass is set at rest. The time step used is $\Delta t = 2$ s. The numerical solution for the cross section $y = 0$ compared with the analytical solution Eq. 27 for $t = 200$, 600 and 1,000 s is shown in Fig. 9.

In this case, the advection schemes by Perot (2000) and Kramer and Stelling (2008) give similar results which are in a good agreement with the analytical solution for the time $t = 200$ s. For the time $t = 600$ and 1,000 s, results obtained using the scheme by Kramer and Stelling (2008) are much better than that of the scheme by Perot (2000). Moreover, for this test case

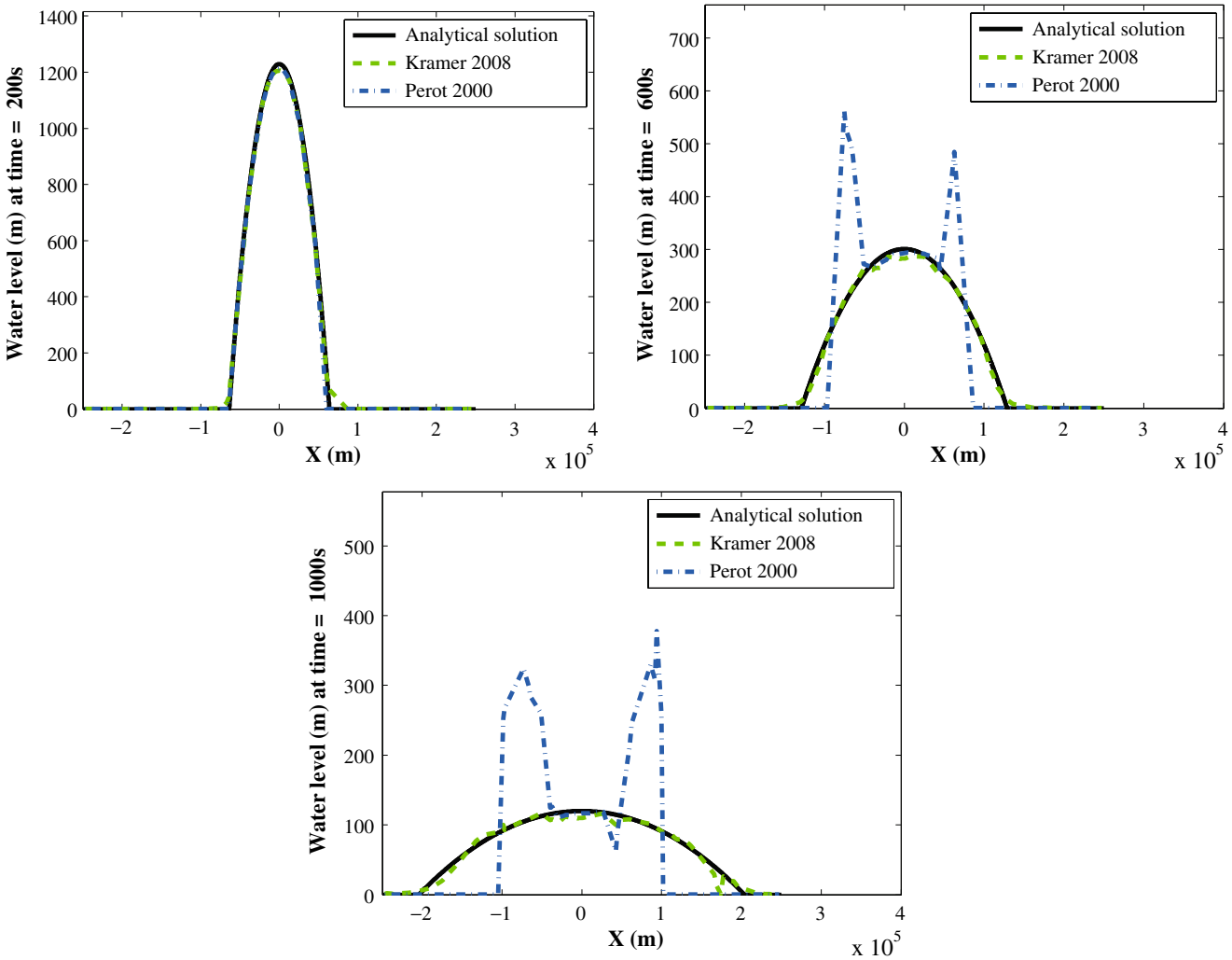

Fig. 9 Parabolic flood wave: comparison of the computed (*dots*) surface elevation profile against the analytical prediction (*solid line*) for $t = 200$, 600 and 1,000 s

using of the advection scheme by Perot (2000) led to the model instability. Reducing the time step by the factor of 10 did not solve the stability problem. The results shown here were obtained with h_c substituted by $\bar{h}_j$ in the denominator of Eq. 17. Some other experiments also show that substitution of $\bar{h}_j$ in place of h_c in Eq. 18 leads to increased stability and slightly better results.

6 Comments on the advection discretisation in the multi-layer case

The advection scheme by Perot (2000) has been used successfully in a number of unstructured grid models (see, for example, Stuhne and Peltier 2009 and Fringer et al. 2006). However, Stuhne and Peltier (2009) noticed that the results of their multi-layer simulation of the global M_2 tide were much worse in the coastal region than the same results but from a 2D simulation. Whereas the deep ocean amphidromic patterns are similarly resolved in both the 2D and 3D cases. Fringer et al. (2006) successfully use the advection scheme by Perot (2000) for internal wave simulations. They, however, claim that the advection scheme does not conserve momentum in the cells containing the free surfaces. This is quite a surprising statement given the good deep ocean results of Stuhne and Peltier (2009). A possible reason for the lack of conservation could be the use of the vertical velocity w in place of the velocity normal to the "horizontal" faces (see Fig. 2). If that is so, we should not see the deterioration of the results in the case of multiple layer simulation, since the normal velocity interpolation is used.

To assess this, we use a two-layer simulation of the dam break over a wet bed as described in Section 5.

Simulations were performed using both the momentum advection schemes 16 and 23; however, only the results for the scheme 23 are shown. Figure 10 shows the surface elevation and the face velocity norm for the two layers at $z_1 = 0.08$ m and $z_2 = 1.0$ m above the bed.

Both the free surface and velocity are represented much worse here than in the one-layer simulation, shown in Fig. 5. Moreover, the calculated flow velocity differs across the layers, whereas it should be uniform. As can be seen, difference in the velocity norm between the layers for the faces on one column side is as high as 3.31 m/s.

Our conjecture is that the poor-free surface representation is caused by the artificial vertical structure in the flow, which is created solely due to the presence of vertical z-layers. For the z-layer models, the face heights are usually defined as $h_{f,k} = h_{c,k}$, and all of the cell heights $h_{c,k}$ are equal to each other everywhere except for the cells containing the free surface (and bed). Due to this, the contribution to the change of momentum from the change of velocity (contribution from the vertical faces in Eqs. 11 and 21 for the internal layers) is not the same as for the top and bottom layer. Similarly, the velocity reconstruction procedure (Eq. 6) for the internal layers is different from that for the top and bottom layer.

In the next section, using the momentum advection scheme 23 as an example, we show how one can prevent the model from creating artificial vertical structure in the flow.

6.1 Improved implementation of the advection term

In absence of bottom friction, the flow velocity should be constant in depth. That is, the momentum equation 5 should be identical for all of the layers. This means that advection and Coriolis operators for a particular layer should be the same as for the whole water column. This is possible if the ratio of a cell height to the height of its face is the same as the ratio of the respective column water depth to the column side water depth this face belong to. That is, the identity

$$\frac{h_{f,k}}{h_{c,k}} = \frac{h_f}{h_c} \tag{29}$$

should be valid for all cells and faces.

Assume the situation shown in the left panel of Fig. 11: The column upwind of the column side j has three layers with heights $h_{u,k}$, whereas the downwind column has only two layers with heights $h_{d,k}$. Define for the downwind column adjacent to the column side j exactly three sub-layers whose thickness $h'_{d,k}$ is determined by

$$h'_{d,k} = \frac{h_{j,k}}{h_j} h_d \tag{30}$$

Since the water depth at the column side j is equal to that of the upwind column and $h_{j,k} = h_{u,k}$, the relation 29 holds automatically for the column side j and the upwind column. The heights of the other faces need

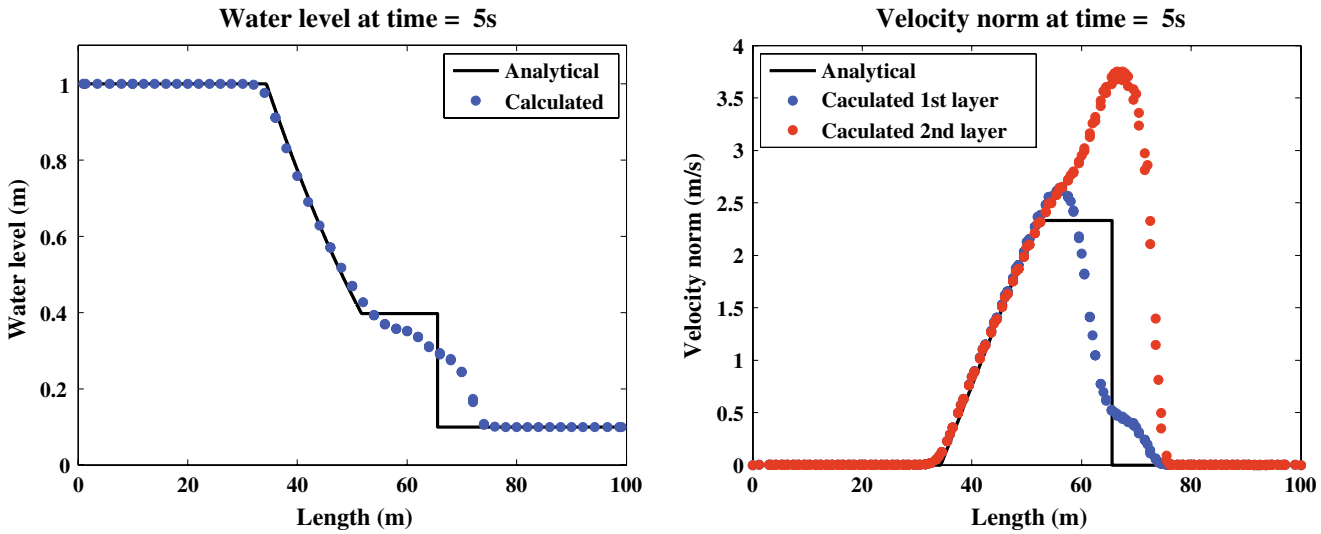

Fig. 10 Dam break over wet bed: surface elevation and velocity norm for the two-layer calculation compared with the analytical solution

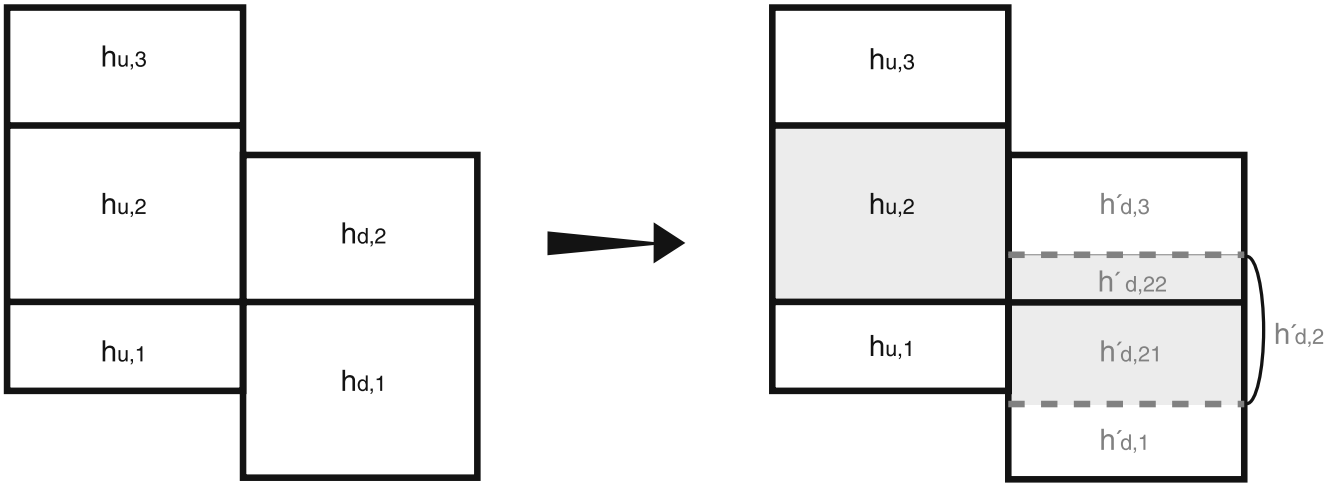

Fig. 11 The side view of the control volume (*shaded area*) used to discretise the horizontal momentum equation at face

to be redefined according to Eq. 29 for both of the columns.

With the above definition, the advection term $a_{j,k'}$ given by Eq. 23 becomes

$$a_{j,k'} = \sum_{c'_k} \delta_{j_{k'},c_{k'}} \alpha_{j,c} \sum_{f'_k} s_{f,c} \frac{1}{A_c \bar{h}_j} \left[h_f l_f u_{f,k'} (\mathbf{u}^*_{f,k'} \cdot \mathbf{n}_j) - \left(u_{j,k'+\frac{1}{2}} \sum_{i=k'_b}^{k'} h_f l_f u_{f,i} - u_{j,k'-\frac{1}{2}} \sum_{i=k'_b}^{k'-1} h_f l_f u_{f,i} \right) \right] \tag{31}$$

and the velocity reconstruction procedure (Eq. 6) reads as

$$\mathbf{u}_{c,k'} = \sum_f \delta_{f,c} d^c_f \frac{h_f l_f}{A_c h_c} u_{f,k'} \mathbf{n}_f \tag{32}$$

Here k' refers to the index of the sub-layer, not the actual layer.

Note that the sub-layer 2 crosses two actual layers in the downwind column. In that case, the normal velocity component $u_{f,k'}$ is approximated as

$$u_{f,2'} = \frac{h'_{d,21}}{h'_{d,2}} u_{f,1} + \frac{h'_{d,22}}{h'_{d,2}} u_{f,2} \tag{33}$$

where $h'_{d,21}$ and $h'_{d,22}$ are the heights of the parts of the sub-layer 2 belonging to the actual layers 1 and 2, respectively (see Fig. 11). Moreover, the same principle should be applied to the velocity reconstruction procedure (Eqs. 6–7) also used in the discretisation of the Coriolis term.

The two-layer simulation of the dam break over wet bed described above was repeated using the advection

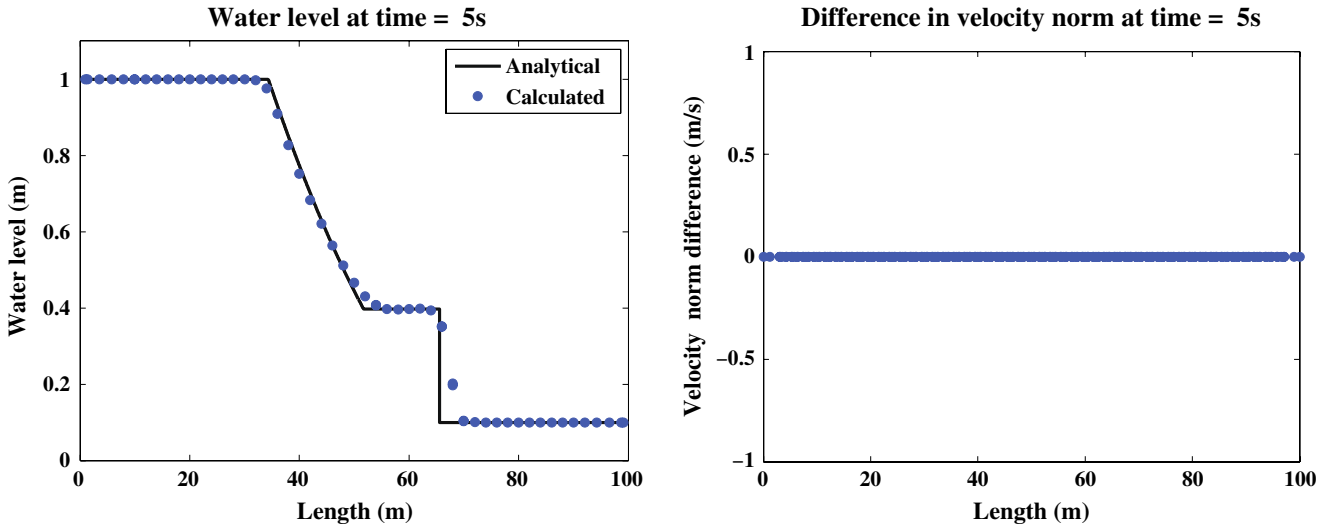

Fig. 12 Dam break over wet bed: surface elevation (*left*) compared with the analytical solution (*black line*) and the difference in velocity norm between the two layers

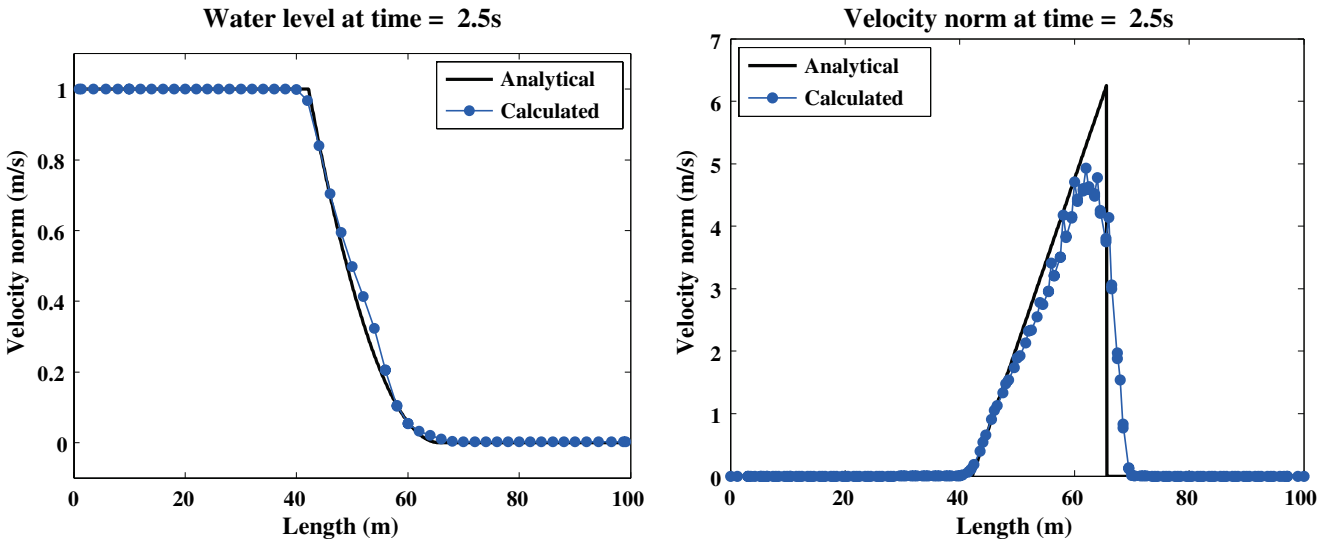

Fig. 13 Dam break over dry bed: surface elevation (*left*) and the velocity norm for the bottom layer (*right*) compared with the analytical solution (*black line*)

discretisation given by Eq. 31; the results are shown in Fig. 12. As can be seen, the free surface elevation is represented at least as well as in the one-layer case shown in Fig. 5. The difference in the velocity norm between the two layers for the faces on one column side is exactly 0.

Simulation of a dam break over a dry bed as described in Section 5 was repeated using ten vertical layers located at $z = 0.1, 0.2, 0.3, ..., 0.9$ and 1.0 m above the bed. The resulting surface elevation and velocity norm for the bottom layer (shown in Fig. 13) are identical to that of the one-layer simulation (Fig. 6). The velocity norm for the other layers are equal to the velocity norm of the bottom layer in the locations where they are defined.

All the other experiments from Section 5 were also repeated with multiple layers using the advection scheme 31. The results are not shown here since they are, as expected, identical to the one-layer case.

7 Discussion

Staggered C-grids has been used in a variety of unstructured grid models for large-scale ocean applications. The unstructured finite-volume scheme of Casulli and Walters (2000) combines a semi-implicit time integration of the equations with a semi-Lagrangian approach for the advection term. Such a scheme can be shown to be stable at any flow Courant number. This approach was also adopted in Delfin (Ham et al. 2005). The drawback of such schemes is that most implementations do not provide conservation of momentum and hence are not suitable for simulation of such phenomena as flooding and drying. Accurate simulations of flooding and drying are important for dam break problems and tsunami simulations.

Here we compare two Eulerian advection schemes, namely the scheme by Perot (2000) and Kramer and Stelling (2008) which we generalise to the 3D case. We show that the scheme of Kramer and Stelling (2008) gives better results for dam break problems.

We show that special attention is required to the discretisation of the momentum equation in the presence of multiple z-layers. It is a common practice to vary only the thickness of the top and bottom layers to represent the free surface and bathymetry variation. The thickness of the internal layers which do not contain the free surface and bed is kept constant. The heights of the faces for these layers are usually defined to be equal to the heights of the layers, thus replicating the flat bed case. We show that this generates a staircase problem which leads to inaccurate solutions and may erroneously introduce vertical structure in the flow.

A model must not create a vertical structure in the flow if there is no physical reason for that. Therefore, the discretised momentum equation for a particular layer should be identical to that of any other layer if phenomena such as bottom friction, viscosity and diffusion are disregarded. This means that heights of any two cells sharing a face should have the same ratio as the respective column heights. In addition, the ratio of the cells' face heights should be the same as the ratio of the heights of the column sides. Here we

propose to locally redefine/remap cells and faces in such a way that the above-mentioned conditions are fulfilled. The remapping procedure allows us to always have equal number of (sub)cells to the left and to the right of a particular column side, thus removing the discontinuities associated with the representation of the free surface and bathymetry as a series of steps. It also allows us to simulate flooding and drying phenomena in the presence of multiple z-layers.

In C-grid models, a discretisation of the Coriolis force may become an additional source for the staircase problem, since this class of models only solves for the component of velocity normal to a face and the tangential component is interpolated. Therefore, the same layer remapping procedure should be applied while reconstructing the tangential velocity in order to prevent the model from the creating artificial vertical structure due to the discretisation of the Coriolis term in the presence of multiple z-layers.

References

Adcroft A, Campin J-M (2004) Rescaled height coordinates for accurate representation of free-surface flows in ocean circulation models. Ocean Model 7(3–4):269–284

Adcroft A, Hill C, Marshall J (1997) Representation of topography by shaved cells in a height coordinate ocean model. Mon Weather Rev 125(9):2293–2315

Baptista A, Zhang Y, Chawla A, Zulauf M, Seaton C, Myers E III, Kindle J, Wilkin M, Burla M, Turner P (2005) A cross scale model for 3D baroclinic circulation in estuary-plume-shelf systems: II. Application to the Columbia River. Cont Shelf Res 25:935–972

Bleck R (2002) An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Model 37: 55–88

Bleck R, Rooth C, Hu D, Smith L (1992) Ventilation patterns and mode water formation in a wind- and thermodynamically driven isopycnic coordinate model of the North Atlantic. J Phys Oceanogr 22:1486–1505

Burchard H, Petersen O (1997) Hybridization between σ- and z-co-ordinates for improving the internal pressure gradient calculation in marine models with steep bottom slopes. Int J Numer Methods Fluids 25(1):1003–1023

Campin J-M, Adcroft A, Hill C, Marshall J (2004) Conservation of properties in a free-surface model. Ocean Model 6(3–4): 221–244

Carrier GF, Wu TT, Yen H (2003) Tsunami run-up and drawdown on a plane beach. J Fluid Mech 475(1):79–99

Casulli V, Walters R (2000) An unstructured grid, three-dimensional model based on the shallow water equations. Int J Numer Methods Fluids 32(3):331–348

Fringer O, Gerritsen M, Street R (2006) An unstructured grid, finite volume, nonhydrostatic, parallel coastal ocean simulator. Ocean Model 14:139–173

Griffies SM, Pacanowski RC, Schmidt M, Balaji V (2001) Tracer conservation with an explicit free surface method for z-coordinate ocean models. Mon Weather Rev 129(5): 1081–1098

Hallberg R (1997) Stable split time stepping schemes for large-scale ocean modeling. J Comput Phys 135(1):54–65

Halyer PJ, Lermusiaux PFJ (2010) Multiscale two-way embedding schemes for free-surface primitive-equations in the "multidisciplinary simulation, estimation and assimilation system" (mseas). Ocean Dynamics (this issue)

Ham D, Pietrzak J, Stelling G (2005) A scalable unstructured grid 3-dimensional finite volume model for the shallow water equations. Ocean Model 10:153–169

Kleptsova O, Pietrzak JD, Stelling GS (2009) On the accurate and stable reconstruction of tangential velocities in C-grid ocean models. Ocean Model 28(1–3, Sp. Iss. SI):118–126

Kliem N, Pietrzak J (1999) On the pressure gradient error in sigma coordinate ocean models: A comparison with a laboratory experiment. J Geophys Res-Oceans 104(C12):29781–29799

Kramer SC, Stelling GS (2008) A conservative unstructured scheme for rapidly varied flows. Int J Numer Methods Fluids 58(2):183–212

Perot B (2000) Conservation properties of unstructured staggered mesh schemes. J Comput Phys 159(1):58–89

Stelling G, Duinmeijer S (2003) A staggered conservative scheme for every Froude number in rapidly varied shallow water flows. Int J Numer Methods Fluids 43(12):1329–1354

Stelling G, Van Kester J (1994) On the approximation of horizontal gradients in sigma co-ordinates for bathymetry with steep bottom slopes. Int J Numer Methods Fluids 18:915–935

Stuhne G, Peltier W (2009) An unstructured c-grid based method for 3-D global ocean dynamics: free-surface formulations and tidal test cases. Ocean Model 28(1–3):97–105

Thacker WC (1981) Some exact solutions to the nonlinear shallow-water wave equations. J Fluid Mech 107(1):499–508

Wenneker I, Segal A, Wesseling P (2002) A mach-uniform unstructured staggered grid method. Int J Numer Methods Fluids 40(9):1209–1235

Ocean Dynamics (2010) 60:1463–1477
DOI 10.1007/s10236-010-0347-6

Sensitivities of an adjoint, unstructured mesh, tidal model on the European Continental Shelf

Silvia Maßmann

Received: 15 April 2010 / Accepted: 27 September 2010 / Published online: 7 October 2010

Abstract Unstructured mesh models can resolve the model domain with a variable and very fine mesh resolution. Nevertheless, tuning the model setup is still required (for example because of parametrized subgrid processes). Adjoint models are commonly used to calculate sensitivities of ocean models and optimize their parameters so that better agreement is achieved between model simulations and observations. One major obstacle in developing an adjoint model is the need to update the reverse code after each modification of the forward code, which is not always straightforward. Automatic differentiation is a tool to generate the adjoint model code without user input. So far this method has mainly been used for structured mesh ocean models. We present here an unstructured mesh, adjoint, tidal model using this technique, and discuss the sensitivities of the misfit between simulated and observed elevations with respect to open boundary values, the bottom friction coefficient and the bottom topography. The forward model simulates tides on the European Continental Shelf and we show that the tidal model dynamics in the adjoint simulations can be used to define regions, where parameters or mesh has to be optimized. We analyze the dependence of the sensitivities on the wave type and mesh resolution to specify whether the model misfit originates from physical or numerical model deficiencies. In the sensitivity patterns, it is possible to identify islands not resolved in the mesh. We propose to refine the mesh prior to the parameter optimization.

Keywords Unstructured mesh methods · Tidal models · Adjoint models · Parameter optimization · European Continental Shelf

Responsible Editor: Eric Deleersnijder

S. Maßmann (✉)
Alfred Wegener Institute for Polar and Marine Research,
Postfach 12-01-61, 27515 Bremerhaven, Germany
e-mail: Silvia.Massmann@awi.de

Present Address:
S. Maßmann
Bundesamt für Seeschifffahrt und Hydrography,
Bernhard-Nocht-Str. 78, 20359 Hamburg, Germany
e-mail: Silvia.Massmann@bsh.de

1 Introduction

Unstructured mesh ocean models provide the opportunity to resolve the domain with varying resolution. This technique leads to a domain representation closer to reality than in structured meshes, but tuning of the model setup is still required to increase accuracy. Making such an unstructured mesh is always a compromise between computational cost, availability of bathymetry data with sufficient resolution, realistic coastline shape and representation of coastal features. Parametrization of sub-grid processes depends on resolution. When numerical and physical influences are inseparable, the determination of parameter values is not straightforward.

Manually adjusting the parameters is prevented by the large number of unknowns and, therefore, adjoint models and optimization algorithms are often applied to automate the search for optimal parameters.

Adjoint models solve the so-called Euler–Lagrange equations obtained by using Lagrangian multipliers with the discrete model equations (see, e.g., Evensen et al. 1998; Bennett 1992). They are used to find optimized parameters minimizing a cost function. The cost

function is usually a weighted misfit between observation and model simulation, and a function of the so-called control and state variables. The control variables or parameters are analyzed concerning their influence on the cost function. Often penalty terms are used in the cost function to force the control variables to stay close to some initial guess. The state variables result from the model integration, e.g., velocity and elevation, and some observed counterpart exists for them. In ocean models, the control variables are, for example, the initial state, the governing parameters such as lateral friction, or external forcing such as wind-stress. The choice of the control variables and form of the cost function, especially the penalty terms, determine the performance of the parameter optimization, but there is little information how to define the cost function with optimal properties. Usually for the cost function, the squared difference of the model output to available data is used, in particular the misfit to satellite-based estimates of e.g., the sea surface height in global ocean models (Schröter et al. 1993; Wenzel et al. 2001; Wenzel and Schröter 2002; Ayoub et al. 2001). The harmonic constituents of TOPEX/POSEIDON data have been used for example by Le Provost et al. (1995a), Andersen (1996), Egbert and Erofeeva (2002) and He et al. (2004) for global tidal inversion.

Depending on the complexity of the model, coding an adjoint model is more or less expensive and cumbersome. In manually differentiated (MD) adjoint models these equations are hand-coded, in contrast to the adjoint model generated by automatic differentiation (AD) tools. AD is a software that adds statements to the source code which compute the derivatives (see the website autodiff.org, or e.g., Giering and Kaminski 1998). The advantage of AD compared with MD is that the differentiation of the model can be repeated without much user input once the forward code changes. One example for an adjoint model generated by AD is the MITgcm (see http://mitgcm.org, or, e.g., Marshall et al. 1997; Marotzke et al. 1999; Heimbach et al. 2005). But still, there are few tidal ocean models using AD, since for AD certain restrictions apply on the code structure. It is generally assumed, that unstructured models are to complex. Its usefulness for especially unstructured models have been seldomly addressed.

MD adjoint models on unstructured meshes are more common, for example, the adjoint of the TELEMAC2D model (see Hervouet and Van Haren 1994 or Hervouet 2007), the adjoint of the QUODDY model (Lynch and Hannah 2001), or the adjoints of stationary ocean models by Dobrindt and Schröter (2003) or Sidorenko (2004), to mention just a few of them. Most of them are MD and some of them lag the development of the forward models, which have been steadily improving over the last 10 years.

In this paper, we simulate tides on the European Continental Shelf with a new, non-stationary, non-linear model based on unstructured meshes. The domain and relevant geographical names are shown in Fig. 1.

There already exist many tidal models in this domain (see citations in Davies et al. 1997 or the brief review in Maßmann 2010), and some are even in operational use and predict storm surges (e.g., Dick et al. 2001; Verlaan et al. 2005). We use the NCLF model introduced in Maßmann et al. (2010), because it uses unstructured meshes and has a code structure that can be adjusted to the requirements of AD.

Since the global tidal solutions have become very reliable in most parts of the ocean (Le Provost et al. 1995b), we use them to drive our regional tidal models. We analyze the initial gradient of the cost function with respect to depth, the bottom friction coefficient and the open boundary values resulting from the AD adjoint model.

Energy dissipation of ocean tides is considerably influenced by the parametrization of the bottom friction in shelf seas (Le Provost et al. 1995b). Since measurements of the bottom friction coefficient are rare, it is often necessary to use tidal inversion to determine its values.

For example, Lardner et al. (1993) and Das and Lardner (1992) optimized bottom friction and water depth in tidal simulations with an MD-adjoint finite difference 2D shallow water model on a structured mesh. Ten Brummelhuis and Heemink (1993) extended the parameter estimation to wind stress coefficients and open boundary values and applied it to a finite difference model on the European Continental Shelf.

Taguchi (2004) focused on the so-called shallow water tides, which were generated by non-linear coastal interactions of the main tidal constituent M_2, and estimated parameters inversely by an MD adjoint, structured model in the Irish Sea.

Systematic calibration of tidal open boundary phase and amplitude, bottom friction, and depth was also done by Verlaan et al. (1996) in a European shelf model. In contrast to others, they divided the region into nine rectangles or triangles to reduce the number of control parameters. This was done as the number of observations was considerably smaller than the degree of freedom of the parameter set. This remedy shows the difficulty to define the penalty terms, which serve as a kind of regularization in the (probably) ill-posed inverse problem. Furthermore, they first tuned amplitude and phase of the open boundary forcing, then depth and

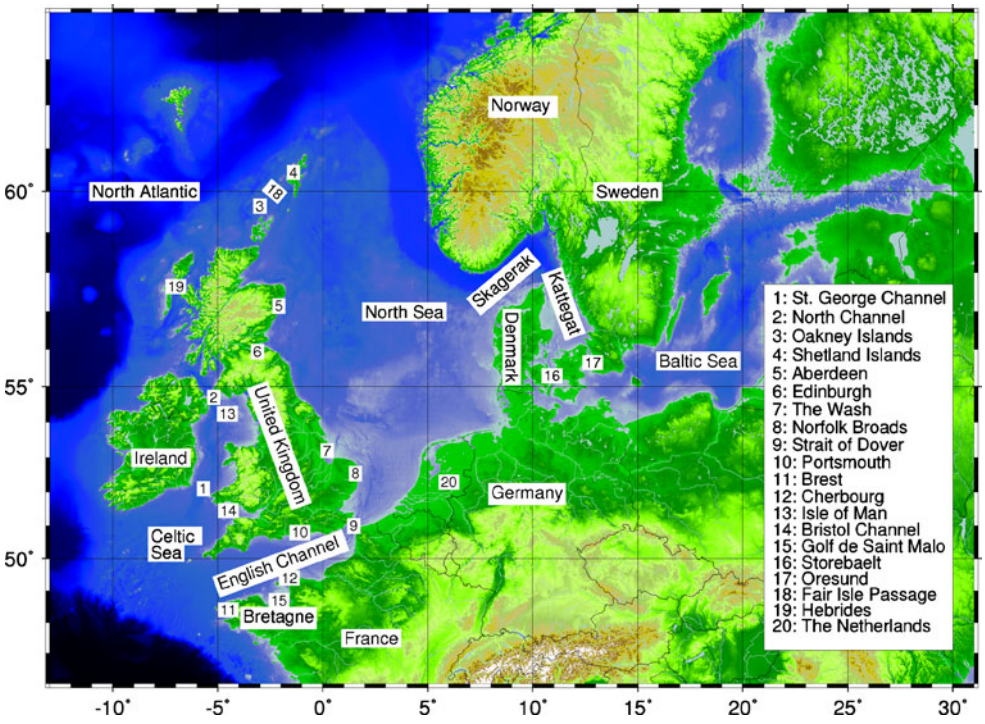

Fig. 1 Geographical names used in the paper

finally the Chezy bottom friction coefficient C (the relationship between C and the bottom friction coefficient r is given by $r = g/C^2$). Optimizing parameters sequentially and not in parallel requires that the parameters are independent of each other. In this study, we show that errors of different origin can be mapped on the adjoint sensitivities and make parameter optimization far from being straightforward.

Furthermore, Heemink et al. (2002) presented results of an adjoint finite difference 3D shallow water model on curvilinear coordinates on the European Continental Shelf. For the gradient computation, they neglected the advection and horizontal viscosity term, which could affect the inverse computations in very non-linear estuarine flow and cause slower convergence of the estimates. Adjoint models based on fully non-linear models are rare especially on unstructured meshes and is now provided by this study.

Since regional models simulate the coastal dynamics in more detail and include more non-linear coastal processes, tuning of the open boundary forcing is often required. For example, Logutov and Lermusiaux (2008) presented the inverse estimation of open boundary values by using the linearized equations in spectral domain and solving a quadratic inverse problem without an adjoint model. In some cases, for example when the optimization algorithms do not need derivatives of the cost function, adjoint models can be avoided to save computational time. But usually the minimization is done iteratively by the Broyden–Fletcher–Goldfarb–Shanno method, which requires the gradients of the cost function. For example, the inverse estimation of open boundary values by adjoint models in the context of the shallow water equations has been studied (amongst others) by Shulman et al. (1998), Zhang et al. (2003), Yang and Hamrick (2005) and He and Wilkin (2006).

The purpose of this study is to identify regions, that have to be represented better in terms of parameters and mesh than the rest of the domain. We compare the sensitivities of the cost function on different meshes (a fine and a coarse) and with two tidal waves. At the open boundaries, the model is forced either by the main semi-diurnal tidal constituent M_2 or the main diurnal one K_1. We discuss (non)-similarities of the sensitivities, give physical and numerical explanations and search for conditions, when the adjoint sensitivities represent parameter errors or insufficient representation by the mesh. In contrast to structured mesh models, the unstructured ones have much more freedom in domain representation due to the varying resolution and therefore, errors due to resolution issues can easily be solved. Our results are very suggestive to use adjoint models on unstructured meshes—apart from the parameter optimization—also for mesh refinements, as we can infer unresolved bathymetric features from the sensitivities. Beyond the scope of this paper are possible implications for adaptive mesh models respectively

mesh generation algorithms, which could use the sensitivities in the mesh update respectively generation as well.

2 The model and its adjoint

The model equations are the non-linear, two dimensional, shallow water equations with leapfrog time stepping, i.e.,

$$\frac{\mathbf{u}^{n+1}-\mathbf{u}^{n-1}}{2\delta t} + f\mathbf{k}\times\mathbf{u}^n + g\nabla\eta^n + \left(\mathbf{u}^n\cdot\nabla\right)\mathbf{u}^n = \nabla\left(A\nabla\mathbf{u}^n\right) - rH^{-1}|\mathbf{u}^n|\mathbf{u}^n, \quad (1)$$

$$\frac{\eta^{n+1}-\eta^{n-1}}{2\delta t} + \nabla(\eta^n + H_0)\mathbf{u}^n = 0, \quad (2)$$

where $\mathbf{u}^n$ denotes the velocity and η^n the elevation at time step t_n. f is the Coriolis parameter, δt the time step size, $\mathbf{k}$ the unit vector perpendicular to the horizontal plane of reference, A the viscosity coefficient, r the bottom friction coefficient, H_0 the depth with respect to the reference plane and $H = \eta + H_0$ the instantaneous total water depth. The model configuration does not allow wetting and drying. Instead a minimum depth of 10 m is imposed on the bathymetry.

The leapfrog time stepping scheme gives second order time accuracy and to suppress its numerical mode a weak filtering similar to the Robert–Asselin time filter discussed in Williams (2009) is applied. The velocity is updated as

$$\begin{aligned}\mathbf{u}^{n-1} &= \mathbf{u}^n + \chi\left(\mathbf{u}^{n+1} - 2\mathbf{u}^n + \mathbf{u}^{n-1}\right),\\ \mathbf{u}^n &= \mathbf{u}^{n+1},\\ \mathbf{u}^{n+1} &= \mathbf{u}^{\text{new}}.\end{aligned} \quad (3)$$

The elevation is treated analogously. Here, χ is a small numerical factor. The leapfrog scheme has the advantage, that neither numerical solvers nor fix-point iterations are encountered. This simplifies the data flow and data dependence analysis of the AD tool.

For the spatial discretization the finite element technique applies on triangular, unstructured meshes with linear, continuous functions (i.e., P_1) for the elevation and linear, non-conforming functions (i.e., P_1^{NC}) for the velocity (see, e.g., Hua and Thomasset 1984). This method naturally filters out the spurious numerical modes which are present in the $P_1 - P_1$ approximation. Tidal simulations in the North Sea have shown accuracy differences between $P_1^{\text{NC}} - P_1$ and $P_1 - P_1$ and have given preference to the non-conforming approach (Maßmann et al. 2010). Some more highlights of the non-conforming method can be found, e.g., in Le Roux (2005) and Hanert et al. (2009).

Basically, the model used in this study is the same as the model, called NCLF, presented in a previous intercomparison study (Maßmann et al. 2010). Among the seven models tested in the intercomparison NCLF gave similar good accuracy, but still had a rather simple code structure. The NCLF model used in Maßmann et al. (2010) has been modified to cope with certain memory allocation requirements of the AD tool TAMC (Giering 1999). These changes do not cause major differences in the model accuracy. Nevertheless, the model results presented in this paper are not comparable to the ones in Maßmann et al. (2010) since the underlying mesh and the position of the open boundary is different. The mesh in Maßmann et al. (2010) was generated by another algorithm resulting in a different concentration of nodes and smoothness of the triangles. Furthermore, the domain is enlarged here to better represent the amphidromic point in the English Channel. The open boundary lies now further away from the region of interest, therefore weakening their influence. Another advantage is that the open boundary forcing data from the global tidal model is now in a region that should be well represented by any global tidal model as well.

Here, the mesh covers the region of the European Continental Shelf and was generated by using TRIANGLE plus some smoothing routines (Shewchuk 1996; Harig et al. 2008). The algorithm first generates a mesh of the wet part of the domain, e.g., regions with negative depth values according to the GEBCO one minute bathymetry (see www.gebco.net or GEBCO user manual by Goodwillie 1997) and then moves the nodes until a smooth coastline is guaranteed. Two meshes are used in this study. The coarse mesh with about 7,000 nodes is shown in Fig. 2. The fine one (not shown here) with about 123,000 nodes stretches over the same domain, but with a much finer resolution of islands and coastal features. The GEBCO one minute bathymetry is interpolated to the mesh nodes and nodal depth is corrected to ensure a minimal depth of 10 m.

The open boundary reaches from the French Atlantic coast to the Norwegian coast along the shelf break. Both meshes use data from OTPS (Egbert and Erofeeva 2002) at the open boundary points (blue points in Fig. 2). In this study, the model is either forced

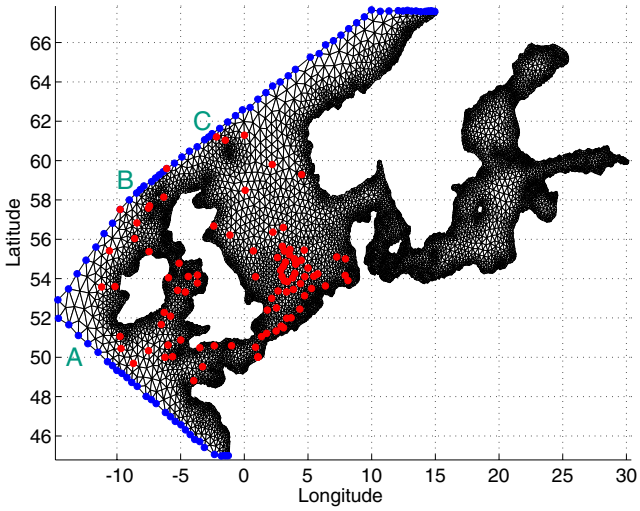

Fig. 2 Coarse mesh. The *red points* show the location of the observation data and *blue points* are the open boundary nodes. On the *x*-coordinate is the longitude in degree W (negative), respectively, E (positive) and on the *y*-coordinate is latitude in degree N

by the main tidal semi-diurnal M_2 or the diurnal K_1 wave only.

The model equations belong to the so-called incomplete parabolic systems (Gustafsson and Sundström 1978), where some energy norm can be used to derive well-posed boundary conditions. It results to a Robin or Neumann type of boundary conditions. In our case, the viscosity term of the shallow water equations, which is the parabolic part of the equations, is very small. Neglecting the viscosity term leads to a hyperbolic system and, for example, the method of characteristics readily gives appropriate boundary conditions (Blayo and Debreu 2005). The number of conditions depends on the Froude number and whether in- or outflow applies at the open boundary. For tidal applications, Flather (1976) proposed such kind of boundary conditions. The co-called Flather boundary condition is widely used (see, e.g., Bourret et al. 2005) for regional tidal models. In previous studies (Maßmann 2010), we experimented with the Flather and the so-called clamped boundary condition. For the clamped boundary condition, no information of the velocity field is required as just the elevation is prescribed at the open boundaries (Androsov et al. 1995). The model accuracy was slightly different, but did not change the results substantially. Therefore, and for keeping the model code simple (and therefore automatic differentiable), we used the clamped boundary condition for the construction of the adjoint model. Amplitude and phase values from OTPS are used to calculate the elevation at the open boundary points for each time step.

In order to assess the difference between observed and simulated state variables, we define a cost function by

$$J = \sum_{m=1}^{M} \left[\left(B_m^{\text{obs}} - B_m^{\text{mod}}\right)^2 + \left(D_m^{\text{obs}} - D_m^{\text{mod}}\right)^2 \right] s_p + \sum_{n=1}^{N} \left[\left(\ln\left(\frac{r_n}{2.6 \cdot 10^{-3}} \right) \right)^2 s_c + \left(\frac{1}{\exp(H_{0n} - A_n^{\text{mod}} - 1)} \right)^2 s_h \right], \tag{4}$$

where M is the number of measurement points and N is the number of nodes. B and D are the cosine and sine amplitude of the tidal surface elevation wave. Analogously the elevation can be represented by a cosine function with amplitude A and phase shift φ. In the third term A is this amplitude. r denotes the bottom friction coefficient and H_0 the bathymetric depth. s_p, s_c, and s_h are scaling coefficients. For the optimization of the parameters (not presented in this paper), the coefficients will be used to facilitate convergence and to normalize the terms. The bottom friction coefficient r is given on each grid node in the whole domain and is initialized by the constant value of $r = 2.6 \cdot 10^{-3}$.

The first term of Eq. 4 measures the misfit between modeled and observed single constituent wave component at the station points. The red dots in Fig. 2 show the location of the data points for the tidal constituents provided by Andersen (2008). This data will be used to compute the error of the model simulations.

The simulated elevation is Fourier-analyzed and decomposed into the constituents before the computation of the cost function. This form of error measure is chosen for several reasons. Firstly, the observation data used in this study only contains the constituents M_2, S_2, K_1, and O_1. But during the model integration also shallow water tides are generated by the non-linear terms in the equations and, therefore, the simulated elevation is a mixture of main and shallow water tides. For the computation of the error it is, therefore, necessary to extract the single constituents. The second reason is that in future applications some weaker constituents can be weighted in favor to account for their smaller amplitudes and absolute errors compared with the major tidal constituents.

The second and the third terms ensure that during optimization (not shown in this paper) the parameters stay within certain limits. The bottom friction coefficient should not deviate from the initial guess

too much. The local depth should not be less than the tidal amplitude to avoid dry falling areas (a wetting and drying scheme is not yet included in the model). These terms do not contribute here in the sensitivities and are just given for completeness.

The model code was differentiated with the AD tool TAMC (Giering 1999) in reverse mode. The forward model code, as it has been used in Maßmann et al. (2010), has been changed to cope with memory allocation restrictions imposed by TAMC. Furthermore, in order to avoid expensive recalculations in the reverse computations, TAMC directives have been used to save the state variables of the forward loops (so-called check-pointing). This is especially important as we have non-linear terms in the model equations.

In the following, we present the initial gradients of the cost function with respect to depth, the bottom friction coefficient and the sine and cosine amplitudes of the open boundary forcing. We perform four simulation runs. In the first two (a and b), we only apply the semi-diurnal M_2 constituent as forcing at the open boundaries using either the coarse (a) or the fine mesh (b). In the other two (c and d), we do the same but with the diurnal K_1 wave as open boundary forcing.

3 Results

Before we present the sensitivities in the tidal M_2 and K_1 single constituent simulations on the coarse and the fine mesh, we want to compare the meshes and its influence on the forward simulations. In the coarse mesh (see Fig. 2), the Dutch coast seems to be not properly located as a close inspection of the mesh boundaries and the location of the coastal tide gage stations suggests. The stations are far more offshore than they are supposed to be. During the meshing procedure the bathymetric depth is cut off at zero. Since Holland partly lies below sea level and in the coarse mesh the resolution is not fine enough to resolve the dikes, the coastline cannot be properly represented although in both cases the same bathymetries are used. The consequence of this misrepresentation is that the misfit between observation and model result increases. In Fig. 3, the observed amplitude respectively phase values at the tide gage stations are plotted over the simulated ones for the M_2 model runs. While for the fine mesh the values are almost perfectly aligned around the diagonal, the correlation in the coarse mesh simulations is strongly degraded. Although the coarse mesh—for the reasons mentioned above—is not a good choice for a realistic model of tides in the North Sea, we keep it to illustrate the influence of the mesh representation on the sensitivities. In a later study, this mesh may be used to test the optimization of both mesh and parameters.

3.1 Sensitivities with respect to depth

Firstly, we compare the initial gradient of the cost function with respect to depth, as shown in Fig. 4. The colorbar uses red and blue, which indicates in red areas (positive values) a decrease of the cost function for decreasing depth and in blue areas (negative values) for increasing depth. The sensitivities are scaled by the inverse of the cost function value to make computations on fine and coarse mesh of comparable order. Note that in order to convert to appropriate scales of dH_0 the number of nodes is additionally needed as a factor (since Eq. 5 is understood locally at each node).

For M_2 on the coarse mesh, the depth sensitivity is negative along the British East Coast, in the English Channel, in the Irish Sea and in the Celtic Sea. Small positive spots are mainly located in the Wash, in the Golfe de St.-Malo, in Bristol Channel, east of the Isle of Man and in the Strait of Dover (see Fig. 1 for a map of the geographical positions). Along the Dutch coast, there is a weaker positive sensitivity, therefore indicating that this area is too deep and should be shallower. The red areas are plausible since in these regions the correction to 10 m minimal depth has been applied.

Comparing the M_2 sensitivities on the coarse mesh (Fig. 4a) with the ones on the fine mesh (Fig. 4b), some regions have even opposite signs. The English Channel has positive sensitivities on depth between the Strait of Dover and Cherbourg and negative sensitivities west of Cherbourg. On the coarse mesh, the whole English Channel is dominantly negative. West of Brest is another region of high-positive sensitivity, which is not present in the coarse mesh. This suggests that on the coarse mesh, its interpolated depths at the nodes and the too coarse coastline had a very big influence especially in the English Channel and the Irish Sea and masked real errors in bathymetry away. Similarities in the sensitivities on the coarse and the fine mesh are in the North Sea at the British East coast. Both sensitivities propose that the depth should be presumably deeper than proposed by the bathymetric data.

For both meshes, the sensitivities of M_2 are smaller near the Hebrides and the northern opening of the North Sea than in the rest of the domain; they are much stronger in the Irish Sea and the English Channel. For K_1 (Figs. 4c, d) the patterns are in some regions vice versa than they are for M_2. While there is still a sensitivity signal in the Strait of Dover and for the coarse mesh

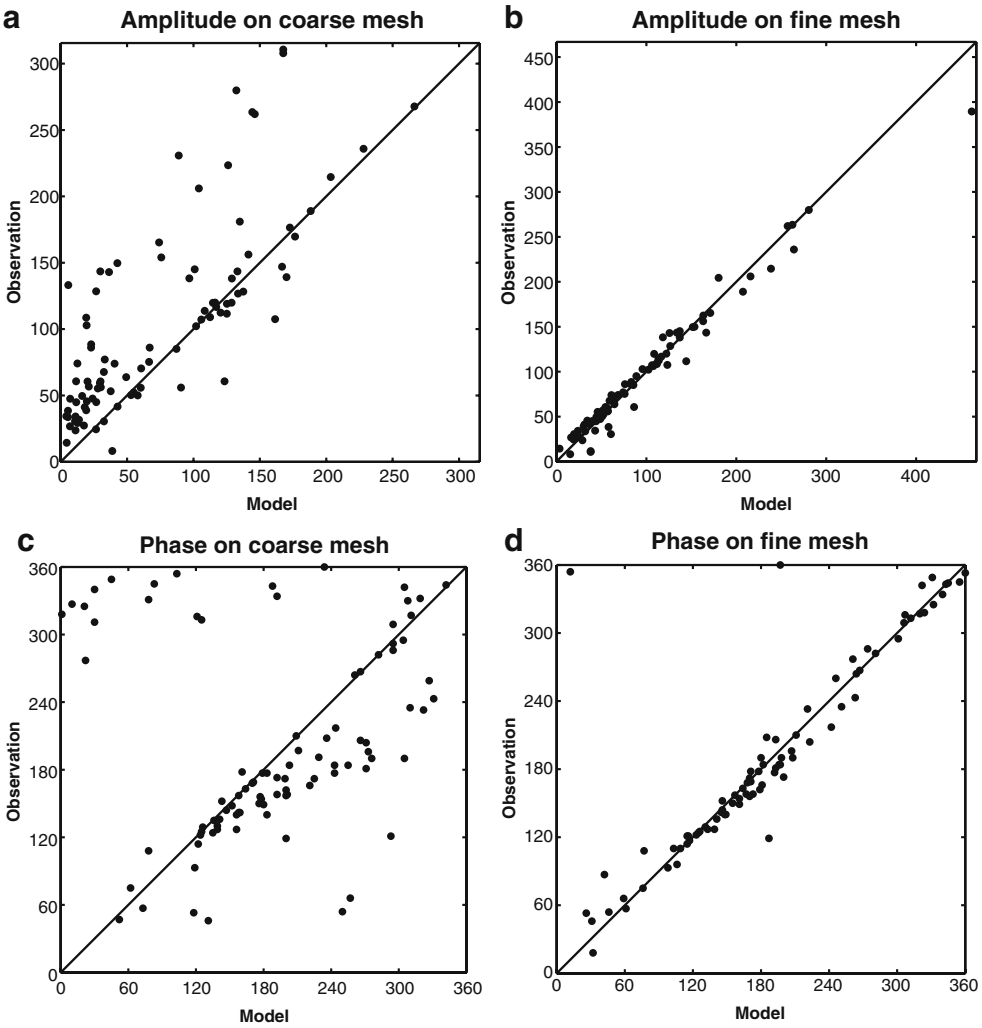

Fig. 3 Correlation between observed and simulated amplitude respectively phase values of the M_2 tidal wave on the coarse (*left figures*) and the fine mesh (*right figures*)

also in the Irish Sea, there are additionally some near the Outer Hebrides and in the North Channel. This may indicate that for K_1 the north-western region of Ireland and Scotland is more important than for M_2.

Analogously to the comparison of M_2 on the coarse and the fine mesh, one can identify many differences between the K_1 sensitivities on coarse and fine mesh by comparing Fig. 4c, d. While on the coarse mesh there are strong patterns almost everywhere around the United Kingdom, the big regions of positive sensitivities on the fine mesh are in the western part of the English Channel and in the Strait of Dover. The predominantly negative region is along the eastern coast of the UK. Amongst other fine structures there are two positive spots (indicating that the depth should be decreased) south-west of the Shetland Islands close to the islands The Sneug and Fair Isle. These islands are too small to be resolved even by the fine mesh. This coincidence is remarkable as no tide gage station is located on the islands. The same spots are also in the sensitivities to the bottom friction coefficient (see Fig. 5d), but with a negative sign implying an increase in the bottom friction. Decreasing depth and increasing the bottom friction coefficient agrees with the existence of unresolved islands.

3.2 Sensitivities with respect to the bottom friction coefficient

The default value of the bottom friction coefficient $r = 2.6 \cdot 10^{-3}$ found in Taylor (1918) has been determined by balancing the energy budget in the Irish Sea. Although already Proudman (1952) found different values for Bristol ($r = 1.4 \cdot 10^{-3}$) and the English Channel ($r = 2.13 \cdot 10^{-2}$) a constant bottom friction coefficient is commonly used.

Since measurements of the energy loss by bottom friction are rare, its coefficient has to be approximated

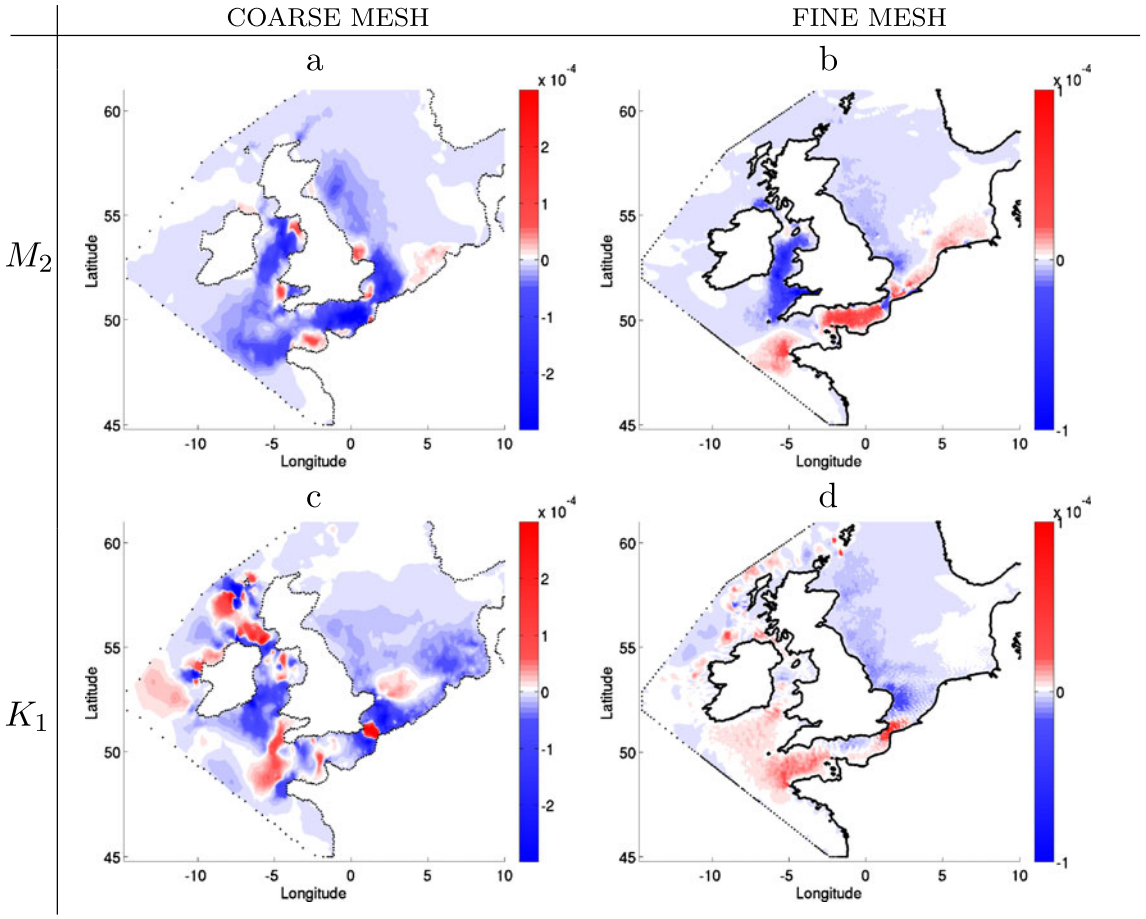

Fig. 4 Initial gradient of the cost function with respect to the bathymetric depth for four separate runs: M_2 on the fine and coarse mesh (*upper figures*) and K_1 on the same meshes (*lower figures*). The values are scaled by the inverse of the cost function value to make the colorbar of comparable order. The unit is m^{-1}

from another basis. In our case, the bottom friction term (see Eq. 1) is an energy sink over the whole water column because of the vertically integrated equations. The amount of energy dissipated by bottom friction depends not only on the bed composition (e.g., sand or mud) but also on the bed form. Therefore, solely bed composition charts do not suffice to determine frictional energy loss and this motivates to inversely estimate the parameters.

In the momentum equations, the depth values and the bottom friction coefficient control the size of the bottom friction terms, and consequently the strength of the fluid velocities. But the bathymetric depth is also used in the continuity equation (see Eq. 2) and influences the surface elevation. This means that, although the bottom friction term contains the reciprocal of the depth, the sensitivities with respect to depth do not have to be inversely to the sensitivities with respect to the bottom friction coefficient. In the following, we investigate the latter and compare them with the sensitivities with respect to depth.

For the M_2 simulation on the coarse mesh the sensitivities of the cost function with respect to the bottom friction coefficient (Fig. 5a) is more complex than the analogous sensitivities with respect to depth. The initial derivative has a positive sign mainly in the Bristol Channel, the Golf de St.-Malo, the eastern part of the English Channel, around the Isle of Man and in a large part of the region between the United Kingdom and the Netherlands. In these regions, the reason for this signal may be a lower amplitude compared with the observation (the underestimation of the amplitude is also confirmed by the correlation in Fig. 3). In the rest of the domain, the sensitivity is dominantly negative, which means more energy is supposed to be dissipated.

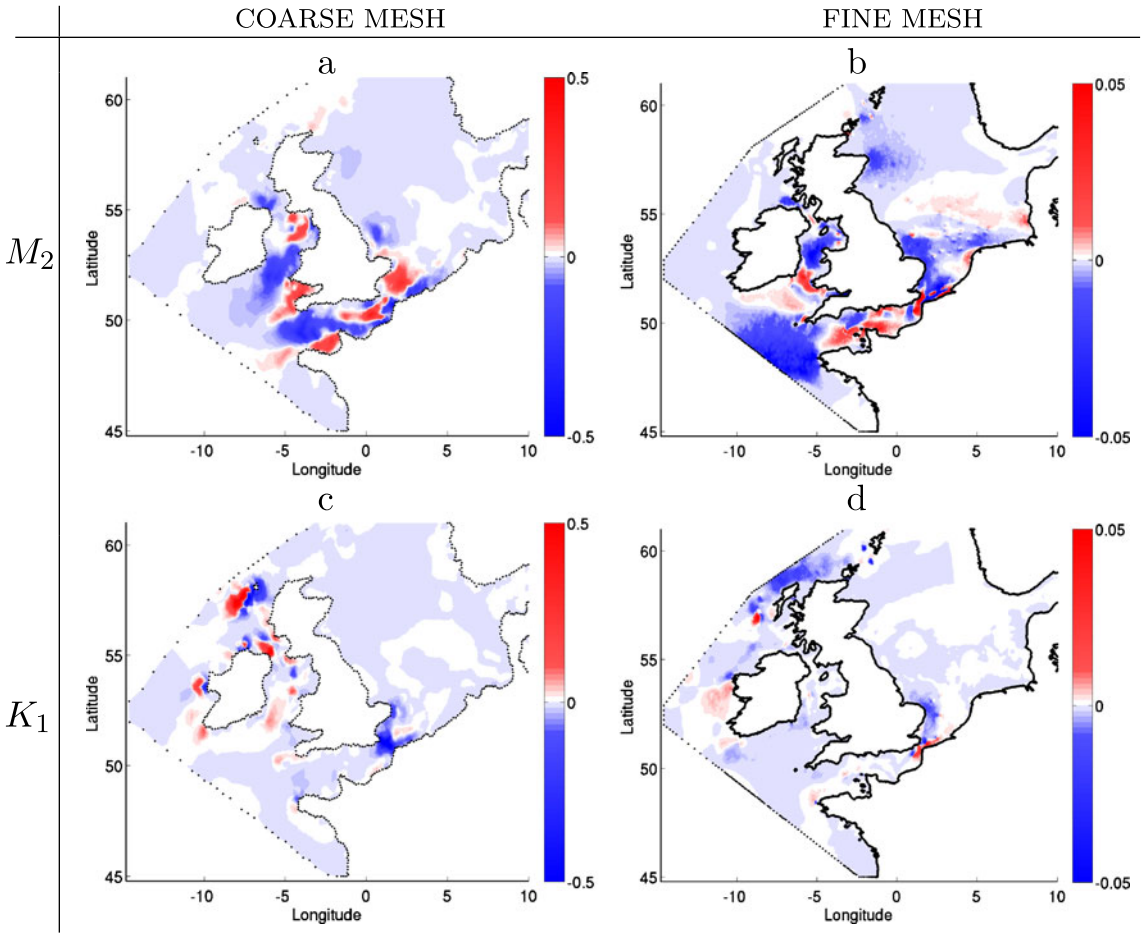

Fig. 5 Initial gradient of the cost function with respect to the bottom friction coefficient for four separate runs: M_2 on the fine and coarse mesh (*upper figures*) and K_1 on the same meshes (*lower figures*). The values are scaled by the inverse of the cost function value to make the colorbar of comparable order. The unit is 1

On the fine mesh (Fig. 5b), the M_2 sensitivities have more details than those on the coarse grid (Fig. 5a). Fine structures are close to the Strait of Dover, near islands, and in the Irish Sea. It may appear a bit patchy, but one should keep in mind that the colorbar is set in a way to highlight positive and negative regions, but does not reflect real amplitudes. The main signal is in the Strait of Dover.

In West of Brest, there is another negative sensitivity. In Fig. 4b, there is a positive depth sensitivity at approximately the same location. This means that the bottom friction term has to be increased and consequently the velocities have to be smaller. More energy is supposed to be dissipated and a decreasing depth indicates that the wave speed should be reduced. On the other hand this region is close to the open boundary and phase and amplitude error may also be introduced and corrected by adjusting the open boundary values.

In Section 3.1, we already observed that K_1 is more sensitive to the bathymetric depth north-west of Scotland and Ireland than M_2 is. As we can see from Fig. 5 this is also true for the bottom friction coefficient. Since we also argued that the sensitivities on the coarse mesh are very much influenced by the mesh misrepresentation, we concentrate the analysis of K_1 sensitivities with respect to the bottom friction coefficient on the fine mesh. As we can see in Fig. 5d there is a strong negative sensitivity north-west and a positive one west of the Hebrides. Additionally, we have the blue spots near the Sneug and Fair Isle as mentioned in Section 3.1 due to the mesh resolution and a significant negative amplitude east of Norfolk and a mixture of negative and positive ones in the Strait of Dover. From other model efforts (Janssen 2009), it is suspected that the parameter east of the Norfolk are quite crucial for the tidal simulations in the North Sea. Our results confirm this experience.

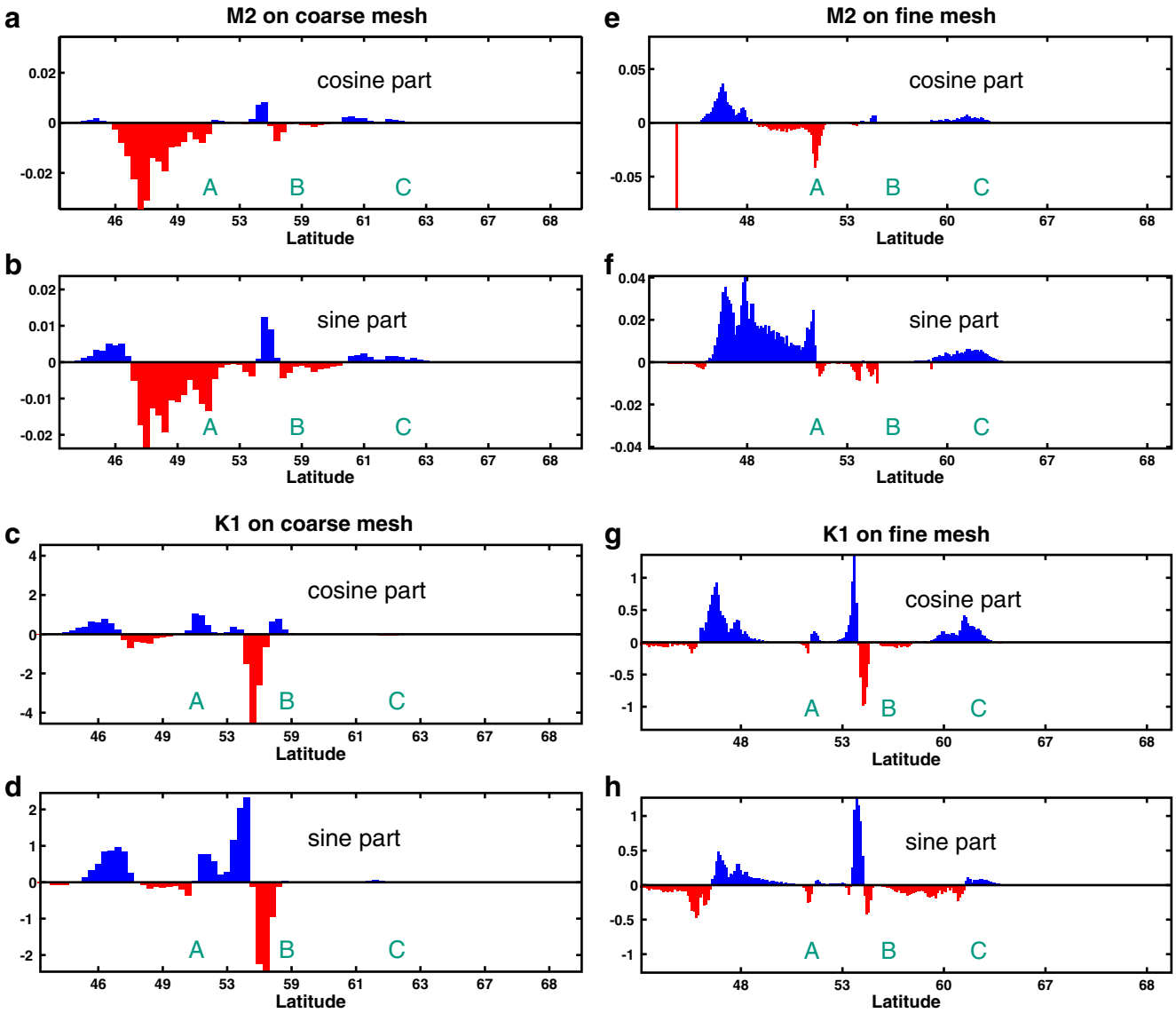

Fig. 6 Initial gradient of the cost function with respect to open boundary sine and cosine amplitude of the tidal oscillation along the open boundary from south to north for four separate runs: M_2 on the fine and coarse mesh (*upper two rows of the figures*) and K_1 on the same meshes (*lower two rows of the figures*). The values are scaled by the inverse of the cost function value to make the colorbar of comparable order. The unit is m^{-1}

3.3 Sensitivities with respect to the open boundary values

The open boundary values are very important as they force the system. We present in Fig. 6a, c, e, and g the cosine amplitudes along the open boundary from south to north for the four different runs (M_2 coarse, K_1 coarse, M_2 fine, and K_1 fine). The sine amplitudes are analogously shown in Fig. 6b, d, f, and h. On the abscissa, the latitudes of the open boundary nodes are given. As the nodes are not equidistant, we plotted three positions (A, B, and C) in the map of the mesh (Fig. 2) and in Fig. 6 to refer to certain parts of the open boundary.

The sensitivities for M_2 on the coarse mesh (Fig. 6a) suggest that there are two main areas of influence of the open boundary on the cost function. One is located west of the Breton coast in the Celtic Sea (south of point A). It is probably a result of the larger error in the English Channel. The wave entering here is the major player in the English Channel and in the Irish Sea, where M_2 has largest amplitudes. Also the M_2 sensitivities on the fine mesh show high-sensitivity values at this part of the open boundary but with partly different signs.

Another region of high sensitivity is north-west of Ireland (between A and B), but with a smaller amplitude. It is located in an area, where the continental shelf break is part of the model mesh. This means a rather sharp change of depth, when a wave entering from the open ocean travels into the model domain. We suspect that some part of the wave may be reflected either by the bathymetry or by the open boundary. Another possibility is that the OTPS data at the open boundary does not properly account for some specific coastal features.

The open boundary part near the Shetland and Orkney Islands (around and south of point C) can be regarded as another example for the possibility that the islands might impose dynamics which are not represented in the open boundary values. Here, a positive sensitivity is present in the coarse and fine M_2 sensitivities and even in the fine K_1 simulations.

For K_1 the sensitivities in the region between points A and B is stronger than the one south of point A on the fine and the coarse mesh (Fig. 6c, d, g, and h). For M_2 this is vice versa. This fits with the results in Sections 3.1 and 3.2, where the region north-west of Scotland and Ireland is more important for K_1 than for M_2.

Remarkable are the similarities of the open boundary sensitivities for M_2 and K_1 on the same mesh especially south of A and around C (compare Fig. 6a, b with c and d or analogously the left figures for the fine meshes). Since the open boundary forcing in the M_2 simulations is completely independent of the one in the K_1 cases, there are two possible explanations for this behavior. Either the OTPS boundary data has this discrepancy and should be corrected in order to be consistent with the tide gage data. Another reason, which we think is more likely for the coarse mesh, is that interior meshing error is mapped on the open boundaries. The second explanation is supported by the fact, that the open boundary sensitivities differ between the fine and coarse mesh especially in the region south of A.

Comparing the sine and cosine sensitivities at the open boundary allows to specify, whether potential error is connected to the phase or the amplitude employed at the open boundary. For M_2 on the fine mesh (compare Fig. 6e, f) the signs of the sine and cosine sensitivities are equal except in a region just south of A. This indicates that mainly changes of the tidal amplitude at the open boundaries suffices to decrease the value of the cost function. Opposite sign indicates that also the tidal phase contains error, which is for example the case south of point A. Similar conclusions can be drawn for K_1 on the fine mesh. The sign of the sine and cosine part of the sensitivities match almost everywhere with one exception just south of point C.

3.4 Magnitudes of the sensitivities

In this section, we want to discuss which one of the chosen parameters may have strongest impact on the simulations. In Table 1 we give the minimum, maximum and mean of the initial derivatives in the M_2 simulations. The first four lines of Table 1 contain the derivative of the cost function with respect to depth $\partial J/\partial H_0$, to the bottom friction coefficient $\partial J/\partial r$ and to the cosine and sine part of the open boundary forcing, $\partial J/\partial D$ and $\partial J/\partial B$. The values are scaled by the number of (open boundary) nodes and the inverse of the cost function value in order to make the values comparable.

$\partial J/\partial H_0$, $\partial J/\partial D$ and $\partial J/\partial B$ have values between -3 and 1 for the coarse mesh, but in the mean $\partial J/\partial H_0$ is

Table 1 Minimum, maximum and mean of the cost function derivatives with respect to depth H_0, the bottom friction coefficient r and the cosine and sine amplitude of the open boundary forcing B and D for the M_2 simulations on coarse and fine mesh

	Coarse Mesh			Fine Mesh		
	Min	Max	Mean	Min	Max	Mean
$\frac{\partial J}{\partial H_0}$	-3.07	1.00	$7.08 \cdot 10^{-2}$	$-4.24 \cdot 10^{+1}$	$1.30 \cdot 10^{+1}$	$1.84 \cdot 10^{-1}$
$\frac{\partial J}{\partial r}$	$-3.80 \cdot 10^{+3}$	$2.68 \cdot 10^{+3}$	$8.24 \cdot 10^{+1}$	$-4.59 \cdot 10^{+4}$	$4.78 \cdot 10^{+4}$	$1.70 \cdot 10^{+2}$
$\frac{\partial J}{\partial B}$	-2.96	0.71	$2.70 \cdot 10^{-1}$	$-2.10 \cdot 10^{+1}$	$0.95 \cdot 10^{+1}$	$9.23 \cdot 10^{-1}$
$\frac{\partial J}{\partial D}$	-2.04	1.06	$2.57 \cdot 10^{-1}$	$-0.26 \cdot 10^{+1}$	$1.07 \cdot 10^{+1}$	1.15
$\frac{\partial J}{\partial H_0}\,\mathrm{d}H_0$	-5.82	4.40	$1.02 \cdot 10^{-1}$	$-4.98 \cdot 10^{+1}$	$2.67 \cdot 10^{+1}$	$3.06 \cdot 10^{-1}$
$\frac{\partial J}{\partial r}\,\mathrm{d}r$	$-4.94 \cdot 10^{-1}$	$3.49 \cdot 10^{-1}$	$1.07 \cdot 10^{-2}$	-5.96	6.21	$2.21 \cdot 10^{-2}$
$\frac{\partial J}{\partial B}\,\mathrm{d}B$	$-3.26 \cdot 10^{-2}$	$3.66 \cdot 10^{-2}$	$5.32 \cdot 10^{-3}$	$-0.94 \cdot 10^{-1}$	$2.83 \cdot 10^{-1}$	$1.57 \cdot 10^{-2}$
$\frac{\partial J}{\partial D}\,\mathrm{d}D$	$-1.30 \cdot 10^{-1}$	$0.29 \cdot 10^{-1}$	$1.25 \cdot 10^{-2}$	$-0.75 \cdot 10^{-1}$	$6.59 \cdot 10^{-1}$	$6.02 \cdot 10^{-2}$

In the first four lines, the sensitivities are scaled by the inverse of cost function values and the number of nodes (n2d) or open boundary nodes (nob). The last four lines are multiplied by 5% of the original parameter value to properly account for the order of $\mathrm{d}H_0$, $\mathrm{d}r$, $\mathrm{d}B$, and $\mathrm{d}D$. The units are the inverse of the variable units over which the differentiation is performed, e.g., m^{-1} or 1

one order of magnitude smaller than the sensitivities with respect to the open boundary values. The smaller the mean value is the less influence has the parameter on the change of the cost function during the optimization. The derivative with respect to the bottom friction coefficient is about three orders of magnitude larger than the others. Similar observations can be made on the fine mesh. Compared with the coarse mesh the derivatives are about one order of magnitude larger. This means that on the fine mesh the parameters will change less during the optimization for decreasing the cost function relatively the same amount. (For the coarse mesh the cost function value is about 78.11 and for the fine mesh it is about 6.05). Since the cost function employs the squared difference (see Eq. 4), its derivative is weighted by the cost function value. Consequently the ratio between the coarse and fine mesh sensitivities, when they are not weighted by the cost function value, is about 13.

In the considerations above, we have not considered the size of the typical error, i.e. the values of $\mathrm{d}H_0$, $\mathrm{d}r$, $\mathrm{d}B$, and $\mathrm{d}D$ in

$$\mathrm{d}J = \frac{\partial J}{\partial H_0}\mathrm{d}H_0 + \frac{\partial J}{\partial r}\mathrm{d}r + \frac{\partial J}{\partial B}\mathrm{d}B + \frac{\partial J}{\partial D}\mathrm{d}D. \tag{5}$$

In order to quantify the influence of each parameter on the change of the cost function, the sensitivities have to be scaled by the order of the parameters and their typical errors. In this case, we assume all parameters have an error of 5%, $\mathrm{d}H_0$, $\mathrm{d}r$, $\mathrm{d}B$, and $\mathrm{d}D$ are consequently 0.05 times the initial parameter values. These new sensitivities are summarized in the last four lines of Table 1. Now depth sensitivities are about one order of magnitude stronger than the bottom friction and open boundary sensitivities. This means that, when 5% error is true for all parameters, a nice representation of the bathymetry is the most important for the tidal simulations.

Assuming 5% error for the bottom friction coefficient and for depth is not always reasonable. For example in areas, where a minimal depth of 10 m has been applied, it can be much higher. As we have illustrated in Section 3.2 the bottom friction coefficient might even be about seven times higher than the default value of $r = 2.6 \cdot 10^{-3}$. In this case, the absolute values of $\partial J/\partial r\,\mathrm{d}r$ would be almost the same order of magnitude as the depth sensitivity. This means that the divergence of the optimized bottom friction coefficient from the default value may be larger and decrease the cost function value as strong as or even stronger than optimized depth and open boundary values. It also means that optimizing the bottom friction coefficient can always be used (or misused) to tune the model results closer to observation. In the previous sections we have illustrated that the mesh may contain some error in the representation of the domain or specific depth features. Then the optimization of frictional parameters may result in frictional energy dissipation not connected to physical reality. Furthermore, the optimized parameters are model and mesh specific. Consequently, resolution issues should be taken care of, especially when one wants to exchange optimized parameters between different models.

4 Discussion

During the analysis of the sensitivities, the resolution is identified as an important factor for faithful simulation results next to the parameters. Parameters are usually defined on the mesh nodes and, therefore, the mesh resolution influences the simulated tidal wave and consequently the calculated sensitivities. The derivatives of the cost function with respect to depth and the bottom friction coefficient even identifies under-resolved islands. This study shows a high correlation between mesh representation of specific bathymetric features and the sensitivities. It may create the possibility to use sensitivities for mesh refinement or in adaptive mesh models.

Although the computation of the adjoint sensitivities and of a new mesh imposes some computational burden (as it has been mentioned by Dobrindt and Schröter 2003), the mesh optimization only has to be done once when setting up the model configuration. Later when the model is used, for example, in an operational sense or for experiments in climate or ocean science, the adjoint model or mesh adaption algorithm do not have to be invoked any more, which saves computational time considerably. After the optimization of the mesh the total number of nodes of the optimized unstructured mesh may even be the same as before or smaller, as the resolution may also be relaxed in regions of smaller sensitivity. This additionally may save computer time.

According to this study the sensitivities may reflect the representation error of the mesh and this highlights the weakness of structured grid models. These models are not able to change resolution smoothly as required by the bathymetry and wave propagation without increasing the computational burden considerably.

Furthermore, the results above indicate, that there are cases where transferring parameter estimates obtained by adjoint sensitivities from the (adjoint) model to another model with a different mesh is not feasible. In our experiments, we even do not change the numerical model, but only the underlying unstructured mesh

and get different initial derivatives of the cost function because of the domain misrepresentation introduced by the coarse mesh.

We recommend to perform a detailed analysis of the sensitivities prior to the parameter optimization to decide whether the initial model misfit can be explained by the uncertainty of the proposed parameter values.

Often the adjoint data assimilation schemes force the model trajectory closer to the observation, but resulting in parameters only suitable for this model. In such cases, the adjoint method solely addresses the model specific errors, instead of giving more realistic new parameters for a general setup. This study illustrates, too, that such situations are easily encountered.

As long as the sensitivities are specific to the model and mesh resolution, the parameter estimates cannot be used for different setups. In particular, when we compare our sensitivities with the ones in Heemink et al. (2002), which were calculated with respect to the error in surface elevation in three-dimensional model on a structured mesh, we see only marginal agreement.

From analyzing the sensitivities from diurnal K_1 and semi-diurnal M_2 tidal simulations we can infer several findings. Firstly, the single-wave sensitivities with respect to the same parameters but on different meshes can contradict when meshing error is mapped on the sensitivities. This implies, that after a possible mesh adaption the sensitivities should differ less and could identify a "perfect" mesh. Since this study solely uses single constituent simulations on only two meshes, we admit that it still has be shown that the sensitivities converge while increasing resolution or optimizing the mesh. This is left for future studies.

Secondly, our results for K_1 and M_2 sensitivities differ sometimes in the same regions. For example, near the Hebrides and the Shetland Islands for K_1 and the Irish and English Channel for M_2, the sensitivities were stronger for one constituent and less for the other. The two waves, and consequently their errors, have spatially different patterns and this explains these differences. This may imply, that a certain wave length can be used to optimize certain parts of the domain, while it is insensitive to other regions and scales. Of course on a "perfect" mesh the parameters obtained by the optimization using one single wave would also give better results for simulations of another single wave, since physically more realistic parameters are found. Unfortunately, our results do not indicate that this situation is encountered here, since the sensitivities in some regions even contradict for several reasons. In general, the superposition of all tidal constituents, the so-called summary tide, would not simply result in sensitivities as a superposition of single constituent runs. The summary tide changes depth, therefore influences the wave characteristics in the forward run and also the sensitivities. Accordingly, the next step is to extent the preceding sensitivity analysis to the summary tide on different meshes.

The sensitivity analysis is also used to determine parameters. Although the bottom friction coefficient may be seven times the default value, this parameter is often taken constant since measurements are rare. By comparing magnitudes of the sensitivities we show that in this two dimensional model the bottom friction coefficient may be as important as the depth and the open boundary forcing. Regions, where the parameter should be changed, can be identified. But as discussed above the sensitivities on different meshes are not completely the same and, consequently, different parameter values may result during an optimization. Furthermore, different parametrization may be needed, when either element scale or depth become small enough. Changes in the sub-grid-scale and mixing parametrization depending on the mesh resolution are not provided by the model and have, therefore, not been considered. Of course this effect is of importance and should be analyzed in future.

5 Conclusion

In this study, an unstructured mesh, adjoint, tidal model was generated from the forward code. We inferred from the adjoint sensitivities of the misfit between observed and simulated tidal characteristics with respect to bathymetric depth the regions or coastal features, that were not well represented by the model setup, but which were important for the tidal dynamics. The mesh resolution requires tuning prior to the parameter optimization. In case of the open boundary forcing, we could identify situations when systematic errors from inside the model domain are mapped on the open boundary sensitivities. This would interfere with a subsequent parameter optimization and we advise to avoid such situations.

Because of the large number of model parameters and possible origins of model deficiencies tuning a model close to observations requires a strategy. We propose to use AD for models in development, since the adjoint model can easily be updated without much user input and it showed to cope with complex source codes as well.

Since ocean model equations use approximations and simplifications, the elimination of numerical and discretization errors (like the mesh) is just the first step in model development. It follows the inclusion

of missing physical processes and then as a last step the optimization of parameters for the specific model domain. Here, the adjoint model is also very helpful.

This study concentrated on the dependency of adjoint sensitivities with respect to depth, bottom friction and open boundary forcing on different wave types and meshes. We did not compute the sensitivities with respect to the mesh resolution. In general this information can be used to improve the mesh and a future task would be to derive an appropriate error norm similar to ones used in adaptive mesh methods. We think that this direction of research is in principle possible on unstructured meshes and should be pursued in the future.

The analysis presented above represents first steps towards designing a tidal model fitting to the observational data. Including other constituents and improving the mesh are the obvious direction for future research. It is hope that systematic analysis of sensitivities will be a guiding principle through this work.

Acknowledgements The author is indebted to Sven Harig for the help with mesh design and smoothing. Many thanks also goes to Sergey Danilov and Martin Losch for proof-reading the article and their constructive critiques. I would also like to thank the anonymous reviewers of this paper for their suggestions and comments.

References

Andersen OB (1996) Inverse methods, vol 63. Lecture notes in Earth Sciences, chap application of inversion to global ocean tide mapping. pp 239–246

Andersen OB (2008) Personal communication. DTU, National Space Institute, Denmark

Androsov AA, Klevanny KA, Salusti ES, Voltzinger NE (1995) Open boundary conditions for horizontal 2-d curvilinear-grid long-wave dynamics of a strait. Advances in Water Resources 18:267–276

Ayoub N, Stammer D, Wunsch C (2001) Estimating the North Atlantic circulation with nesting and open boundary conditions using an adjoint model. The ECCO Report Series 10

Bennett AF (1992) Inverse Methods in Physical Oceanography. Cambridge University Press, Cambridge, UK

Blayo E, Debreu L (2005) Revisiting open boundary conditions from the point of view of characteristic variables. Ocean Model 9:231–252

Bourret A, Devenon JL, Chevalier C (2005) Investigation on passive open boundary conditions adapted to the conjunction of strong currents, standing tidal wave and high stratification: application to the French Guiana Continental Shelf. Cont Shelf Res 25:1353–1373

Das SK, Lardner RW (1992) Variational parameter estimation for a two-dimensional numerical tidal model. Int J Numer Methods fluids 15:313–327

Davies AM, Kwong SCM, Flather RA (1997) A three-dimensional model of diurnal and semidiurnal tides on the european shelf. J Geophys Res 102(C4):8625–8656

Dick S, Kleine E, Müller-Navarra S (2001) The operational circulation model of BSH (BSHcmod)—model description and validation. Berichte des Bundesamtes für Seeschifffahrt und Hydrographie 29:49

Dobrindt U, Schröter J (2003) An adjoint ocean model using finite elements: an application to the South Atlantic. J Atmos Ocean Technol 20(3):392–407

Egbert GD, Erofeeva SY (2002) Efficient inverse modeling of barotropic ocean tides. J Atmos Ocean Technol 19(2):183–204

Evensen G, Dee DP, Schröter J (1998) Parameter estimation in dynamical models. In: Chassignet E, Verron J (eds) Ocean modeling and parametrization. Kluwer Academic Publishers, Series C: Mathematical and Physical Sciences, vol 516, Boston, MA. pp. 373–398

Flather RA (1976) A tidal model of the North-West European Continental Shelf. Memoires Societe Royale des Sciences de Liege 6(X):141–164

Giering R (1999) Tangent linear and adjoint model compiler. Users manual 1.4

Giering R, Kaminski T (1998) Recipes for adjoint code construction. ACM Trans Math Softw 24(4):437–474

Goodwillie A (1997) Centenary Edition of the GEBCO Digital Atlas: user guide to the GEBCO one minute grid. Scripps Institution of Oceanography, La Jolla, CA

Gustafsson B, Sundström A (1978) Incompletely parabolic problems in fluid dynamics. SIAM J Appl Math 35(2):343–357

Hanert E, Walters RA, Le Roux DY, Pietrzak JD (2009) A tale of two elements: P_1^{NC} and RT_0. Ocean Model 28:24–33

Harig S, Chaeroni, Pranowo WS, Behrens J (2008) Tsunami simulations on several scales: comparison of approaches with unstructured meshes and nested grids. Ocean Dynamics 58(5-6):429–440

He R, Wilkin JL (2006) Barotropic tides on the southeast New England shelf: A view from a hybrid data assimilative modeling approach. J Geophys Res 111(C08002). doi:10.1029/2005JC003254

He Y, Lu X, Qui Z, Zhao J (2004) Shallow water tidal constituents in the Bohai Sea and the Yellow Sea from a numerical adjoint model with TOPEX/POSEIDON altimeter data. Cont Shelf Res 24:1521–1529

Heemink AW, Mouthaan EEA, Roest MRT, Vollebregt EAH, Robaczewska KB, Verlaan M (2002) Inverse 3D shallow water flow modelling of the continental shelf. Cont Shelf Res 22:465–484

Heimbach P, Hill C, Giering R (2005) An efficient exact adjoint of the parallel MIT general circulation model, generated via automatic differentiation. Future Gener Comput Syst 21(8):1356–1371. doi:10.1016/j.future.2004.11.010

Hervouet JM (2007) Hydrodynamics of free surface flows. Wiley

Hervouet JM, Van Haren L (1994) TELEMAC-2d Version 3.0 Principal Note. Tech. rep., EDF Électricité de France, Departement Laboratoire National d'Hydraulique

Hua B, Thomasset F (1984) A noise free finite element scheme for the two layer shallow equations. Tellus 36A:157–165

Janssen F (2009) Personal communication. Bundesamt für Seeschiffahrt und Hydrographie

Lardner RW, Al-Rabeh AH, Gunay N (1993) Optimal estimation of parameters for a two-dimensional hydrodynamical model of the Arabian Gulf. J Geophys Res 98(C10):18229–18242

Le Provost C, Bennett AF, Cartwright DE (1995a) Ocean tides for and from TOPEX/POSEIDON. Science 267(5198):639–642. doi:10.1126/science.267.5198.639

Le Provost C, Genco ML, Lyard F (1995b) Coastal and estuarine studies: quantitative skill assessment for coastal ocean models. American Geophysical Union, chap Modeling and predicting tides over the World Ocean, pp 175–201

Le Roux DY (2005) Dispersion relation analysis of the $P^{NC} - P^1$ finite-element pair in shallow-water models. SIAM J Sci Statist Comput 27(2):394–411

Logutov OG, Lermusiaux PFJ (2008) Inverse barotropic tidal estimation for regional ocean applications. Ocean Model 25:17–34. doi:10.1016/j.ocemod.2008.06.004

Lynch DR, Hannah CG (2001) Inverse model for limited-area hindcasts on the continental shelf. J Atmos Ocean Technol 18(6):962–981

Marotzke J, Giering R, Zhang KQ, Stammer D, Hill C, Lee T (1999) Construction of the adjoint MIT ocean general circulation model and application to atlantic heat transport sensitivity. J Geophys Res 104(C12):29529–29547

Marshall J, Adcroft A, Hill C, Perelman L, Heisey C (1997) A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. J Geophys Res 102(C3):5753–5766

Maßmann S (2010) Tides on unstructured meshes. Ph.D. thesis, Fachbereich Physik, Universität Bremen. Available at: http://deposit.d-nb.de/cgi-bin/dokserv?idn=1001723007. Accessed 20 April 2010

Maßmann S, Androsov A, Danilov S (2010) Intercomparison between finite-element and finite-volume approaches to model North Sea tides. Cont Shelf Res 30(6):680–691. doi:10.1016/j.csr.2009.07.004

Proudman J (1952) Dynamical oceanography. Dover, New York

Schröter J, Seiler U, Wenzel M (1993) Variational assimilation of Geosat data into an eddy-resolving model of the Gulf Stream extension area. J Phys Oceanogr 23:925–953

Shewchuk JR (1996) Triangle: engineering a 2D quality mesh generator and delaunay triangulator. In: Lin MC, Manocha D (eds) Applied computational geometry: towards geometric engineering. Lecture notes in computer science, vol 1148. Springer, from the First ACM Workshop on Applied Computational Geometry, pp 203–222

Shulman I, Lewis JK, Blumberg AF, Kim BN (1998) Optimized boundary conditions and data assimilation with applictation to the m_2 tide in the Yellow Sea J Atmos Ocean Technol 15(4):1066–1071

Sidorenko D (2004) The North Atlantic circulation derived from inverse models. Ph.D. thesis, Fachbereich Physik, Universität Bremen

Taguchi E (2004) Inverse Modellierung nichtlinearer Flachwassergezeiten und ihre Anwendungen auf ein Randmeer. In: Berichte aus dem Zentrum für Meeres- und Klimaforschung, no. 47 in Reihe B: Ozeanographie, Zentrum für Meeres- und Klimaforschung des Universität Hamburg, Institut für Meereskunde

Taylor GI (1918) Tidal friction in the Irish Sea. Philos Trans R Soc Lond Ser A 220:1

Ten Brummelhuis PGJ, Heemink AW (1993) Identification of shallow sea models. Int J Numer Methods Fluids 17:637–665

Verlaan M, Mouthaan EEA, Kuijper EVL, Philippart ME (1996) Hydroinformatics'96: proceedings of the second international conference on hydroinformatics. Zürich, Switzerland, 9–13 September 1996, A A Balkema, Rotterdam, chap Parameter estimation tools for shallow water flow models. pp. 341–348

Verlaan M, Zijderveld A, de Vries H, Kroos J (2005) Operational storm surge forecasting in the Netherlands: developments in the last decade. Philos Trans R Soc Lond Ser A 363:1441–1453. doi:10.1098/rsta.2005.1578

Wenzel M, Schröter J (2002) Assimilation of TOPEX/Poseidon data in a global ocean model: differences in 1995–1996. Phys Chem Earth 27:1433–1437

Wenzel M, Schröter J, Olbers D (2001) The annual cycle of the global ocean circulation as determined by 4D VAR data assimilation. Prog Oceanogr 48:73–119

Williams PD (2009) A proposed modification to the Robert–Asselin time filter. Mon Weather Rev 137(8):2538–2546. doi:10.1175/2009MWR2724.1

Yang Z, Hamrick JH (2005) Optimal control of salinity boundary condition in a tidal model using a variational inverse method. Estuar Coast Shelf Sci 62:13–24

Zhang A, Wei E, Parker BB (2003) Optimal estimation of tidal open boundary conditions using predicted tides and adjoint data assimilation technique. Cont Shelf Res 23:1055–1070

Ocean Dynamics (2010) 60:1479–1495
DOI 10.1007/s10236-010-0345-8

Nested circulation modelling of inter-tidal zones: details of a nesting approach incorporating moving boundaries

Stephen Nash · Michael Hartnett

Received: 2 July 2010 / Accepted: 23 September 2010 / Published online: 10 October 2010

Abstract Nested circulation models developed to date either exclude the flooding and drying process or prohibit flooding and drying of nested boundaries; they are therefore ill-suited to the accurate modelling of inter-tidal areas. The authors have developed a nested model with moving boundaries which permits flooding and drying of both the interior domain and the nested boundaries. The model uses a novel approach to boundary formulation; ghost cells are incorporated adjacent to the nested boundary cells so that the nested boundaries operate as internal boundaries. When combined with a tailored adaptive interpolation technique, the approach facilitates a dynamic internal boundary. Details of model development are presented with particular emphasis on the treatment of the nested boundary. Results are presented for Cork Harbour, a natural coastal system with an extensive inter-tidal zone and a complex flow regime which provided a rigorous test of model performance. The nested model was found to achieve the accuracy of a high resolution single grid model for a much lower computational cost.

Keywords Nesting · Circulation model · Ghost cells · Internal boundary · Flooding and drying · Moving boundary

1 Introduction

A common problem in coastal modelling is the correct location of open boundaries; they must be located such that their conditions will not adversely affect model predictions within the area of interest (AOI). This often leads to a situation which requires a large computational domain, of which the AOI comprises only a small percentage. For a structured grid, a high spatial resolution requirement in the AOI can lead to excessively high computational costs. This may necessitate the use of a lower resolution resulting in a loss in accuracy. A common solution to this problem is the use of a nesting method which allows the increase of spatial resolution in a sub-region of the model domain without incurring the computational expense of fine resolution over the entire domain.

In coastal models, the problem of insufficient resolution and poor accuracy is typically associated with areas of complex bathymetry and irregular coastlines; these areas are generally in the inter-tidal zone. A nested domain in such an area will require simulation of flooding and drying both within the domain and along its boundaries. Most nested models developed to date do not incorporate flooding and drying as they have been specifically developed for ocean basin or large-scale regional applications where flooding and drying is not important. Those nested models that do incorporate flooding and drying, e.g. Mike21 (DHI Software 2007), prohibit the flooding and drying of open boundaries as it tends to cause instabilities in the model solution. This is primarily because flooding and drying of an open boundary necessitates a moving boundary where the boundary extents shift during the course of a simulation so as to exclude grid cells upon drying and reinstate those same grid cells upon flooding.

The nested model developed by the authors was designed to facilitate moving boundaries for the nested domains; to the authors' knowledge, it is the first nested circulation model to do so. The model is therefore particularly applicable to the high-resolution modelling of

Responsible Editor: Eric Deleersnijder

S. Nash (✉) · M. Hartnett
Civil Engineering, College of Engineering and Informatics,
National University of Ireland, Galway,
Galway, Ireland
e-mail: stephen.nash@nuigalway.ie

the inter-tidal zone. Ghost cells are specified adjacent to nested boundaries so that the governing equations at the nested boundary grid cells are formulated in a similar manner to interior grid cells; the nested boundaries thus become internal boundaries. The internalisation of the boundaries, when combined with an adaptive interpolation technique, enables the stable flooding and drying of boundary grid cells. The internal boundary formulations also give improved simulation of the effect of the external environment on the internal solution resulting in improved conservation of mass and momentum between the coarse and fine grids and thus improved model performance.

The Depth Integrated Velocity and Solute Transport (DIVAST) model was chosen for nested development as it contains a robust flooding and drying routine (for details, see Falconer and Chen (1991)). DIVAST is a well-accepted, two-dimensional, depth-integrated model which allows simulation of hydrodynamics, solute transport and water quality (e.g. Falconer 1984; Lin and Falconer 1997). The hydrodynamic module is based on the solution of the depth integrated Navier–Stokes equations and includes the effects of local and advective accelerations, the rotation of the earth, barotropic and free-surface pressure gradients, wind action, bed resistance and a simple mixing length turbulence model. Solute transport and water quality are computed using the general depth integrated advection–diffusion equations and incorporate local and advective effects, turbulent dispersion and diffusion, wind effects, source and sink inputs and decay and kinetic transformation processes.

The model was developed using idealised rectangular harbour bathymetries. Post-development, it was important to assess model performance in a natural environment. Cork Harbour, located on the southwest coast of Ireland, was chosen as the test case. Its long, irregular coastline and extensive intertidal zone create a complex circulation regime; it therefore provided a rigorous test of model performance.

An overview of the nested model is presented outlining the main features of the nesting technique adopted. Particular emphasis is placed on the treatment of the nested boundary which was designed to reduce error generation within the nested domain. Aspects of the spatial interpolation scheme, the boundary condition and the use of ghost cells in the boundary formulations are presented. Finally, results are presented for Cork Harbour demonstrating both the accuracy of the new model for a natural system and its computational efficiency.

2 Overview of nesting technique

Nested models are categorised as either one-way (passive) models or two-way (interactive) models. One-way models use output from a coarse grid model to provide boundary conditions for the nested fine grid model. The coarse grid therefore affects the fine grid but there is no mechanism by which the evolution in the fine grid can affect the coarse grid (and hence, its own boundary conditions). Two-way models, in addition to providing boundary conditions for the fine grid, allow the evolution within the fine grid to influence the evolution on the coarse grid. In oceanography, one-way nesting was first implemented by Spall and Robinson (1989) and has since become an established method for studying highly resolved sub-domains. In later years, Spall and Holland (1991) developed the first two-way nested oceanographic model and numerous two-way nested models have since been developed (e.g. Fox and Maskell 1995; Ginis et al. 1998; Barth et al. 2005). Although there are advantages to the interactive system, they are necessarily more complicated and computationally expensive (Spall and Holland 1991).

While much recent research has focussed on two-way nesting, one-way nesting of oceanographic models is still very much an active research area. One-way nesting techniques are best-suited to studying localised phenomena such as land–sea interactions in the coastal zone. In particular, one-way nesting techniques are favoured for limited-area operational modelling where downscaling to coastal areas is required (e.g. Korres and Lascaratos 2003; Leitão et al. 2005; Nittis et al. 2006; Staneva et al. 2009; Holt et al. 2009). In such cases, feedback from the limited-area models to the larger-scale ocean models is not required and two-way nesting techniques are therefore unnecessary. The present research concerned down-scaling to an AOI in the coastal zone. The interaction of the high resolution AOI with the low-resolution parent domain was not of interest; a one-way nested approach was therefore most suitable.

The nested model allows finer spatial resolution to be focused over an AOI by introducing an additional grid (or grids) into the simulation. The present model consists of one outer coarse grid (the parent grid, PG) into which one or more inner fine grids are nested (the child grids, CG). Studies have shown that acceptable results are obtained for spatial nesting ratios ($\Delta x_{\text{parent}}/\Delta x_{\text{child}}$) of 3:1 and 5:1 (e.g. Spall and Holland (1991)) but that substantial degradation in model performance occurs for higher nesting ratios (Barth et al. 2005).The model has been developed to enable multiple nesting so that each child may also be a parent to further children. In this way, numerous levels of nesting can be specified and child grids may be telescoped to achieve virtually any required spatial resolution.

The interaction between a parent and child model is crucial to the accuracy of the nested solution. In general, the interaction involves the following processes: the parent model is first advanced to the next time level. Water surface elevations and velocities are then extracted along the nested

boundaries of the child grid. Since the grid resolutions differ, the parent and child grid solutions at the boundary must be blended. The boundary operator is responsible for blending the solutions and is a critical element of a nested model. The boundary data extracted from the parent grid solution is interpolated in space (and usually in time) to obtain a full set of boundary data for the child. These data are then assigned to the child grid according to some boundary condition and the child model is advanced to the time level of the parent. A boundary operator thus comprises an interpolation scheme and a boundary condition.

The nesting procedure used by the model is presented in graphical form in Fig. 1. For clarity, the procedure is only presented for a single level of nesting and boundary data is only shown for water surface elevations, ζ. The procedure may be summarised as follows:

1. advance parent model one full timestep to time level ($t + \Delta t_p$)
2. extract required boundary data from parent grid and interpolate (in time and space) to required time levels for child grid ($t + \frac{1}{2}\Delta t_c$) and ($t + \Delta t_c$)
3. integrate child grid one full timestep to time level ($t + \Delta t_c$)
4. repeat steps 2 and 3 until time level of child grid matches that of parent grid ($t + \Delta t_p$)
5. return to step 1 and continue

The order of time integration within the model is also shown in Fig. 1. The parent grid may only be advanced in time when the child has been advanced to the time level of its parent. The model uses the ADI solution technique to solve the governing continuity and momentum equations; this requires each timestep be split into two. This does not affect the order of time integration as the parent grid is advanced by one full timestep before any computations begin in the child. However, the time splitting does affect the temporal interpolation process as the child grid requires boundary data at each half-timestep as shown.

3 Treatment of the nested boundary

Errors are introduced into a one-way nested model at the interface between the parent and child grids. It is widely accepted that there are two sources of error at the boundary interface: (1) boundary specification errors and (2) boundary formulation errors (Koch and McQueen 1987). The former are errors arising from the use of incorrect boundary data; they are therefore the result of inaccuracies in the PG solution at the CG boundary. The latter are errors arising from problems in blending the solutions of the parent and child grids at the boundary; they are therefore a result of the boundary operator. However, the authors suggest that a third source of errors exists, arising from the mathematical formulations of the governing equations on the nested boundary. Simplifications are sometimes made to the formulations of the governing equations at a nested boundary, as opposed to at interior grid points, due to finite differencing requirements. As the CG model assimilates the boundary data through these mathematical formulations, their simplification can result in a source of error distinct from the boundary operator. The authors therefore propose a more correct description of error sources as follows:

(1) boundary specification (BS) errors: result from incorrect boundary data
(2) boundary operator (BO) errors: result from a poorly designed boundary operator
(3) boundary formulation errors: result from simplified boundary formulations

BS errors may be reduced by locating nested boundaries in areas of high PG accuracy. However, this is not always easy in coastal models as errors tend to be highest in areas

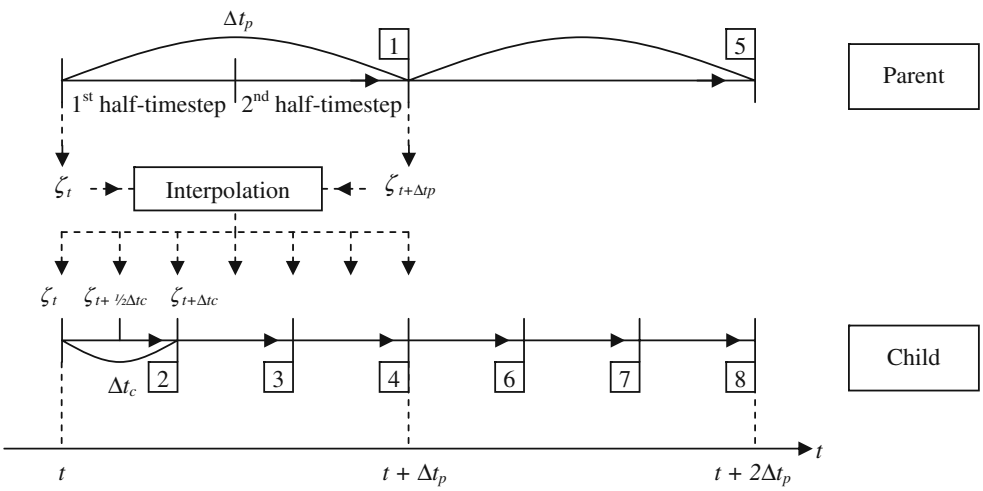

Fig. 1 Schematic of the nesting procedure (p and c signify parent and child grids)

of land–sea interaction. In addition, the location of nested boundaries in inter-tidal zones requires a moving boundary where sections of the boundary can alternately flood and dry. The main focus of model development was the design of a moving nested boundary which aimed to minimise the three error sources defined.

3.1 Boundary operator

Zhang et al. (1986) suggest that an optimal boundary operator has the following properties:

(1) all resolvable waves must propagate across the boundary interface smoothly without generating significant noise
(2) properties such as mass and momentum, exchanged between the two grids, must be conserved

The boundary operator was designed and tested with a view to satisfying these two criteria, thereby reducing BO errors.

3.1.1 The interpolation scheme

The different resolutions of the parent and child grids mean that data from the parent grid must be interpolated in both space and time in order to obtain a full set of boundary data for the child. An unsuitable, or inaccurate, interpolation scheme can lead to errors as a result of loss of mass or momentum, or imperfect reproduction of the parent waveform. Linear interpolation is almost always employed for temporal interpolation (e.g. Miyakoda and Rosati (1977); Pullen and Allen (2001); Korres and Lascaratos (2003)) and was therefore used in the model. Linear and quadratic schemes are most often used for spatial interpolation; however, a number of schemes were evaluated in relation to the conservation of mass and momentum. Four schemes were tested: (1) a zeroth order scheme (ZS), (2) a linear scheme (LS), (3) a mass-conserving quadratic scheme (QS), and (4) an inverse distance weighted scheme (IS). Details of the interpolation formulae are provided in Appendix A. Results from the testing of the interpolation scheme are presented in Section 4.1. The linear interpolation scheme was found to be most accurate; it was therefore implemented in the model.

3.1.2 The boundary condition

The choice of relevant boundary conditions for nested models is a difficult one and has been the focus of many studies. Blayo and Debreu (2005) and Marchesiello et al. (2001) suggest that the best approach is one where incoming and outgoing fluxes are treated separately. The boundary should in effect specify incoming data and evacuate outgoing data. While the application of such a boundary condition works well in idealised experiments, it is more difficult to implement in the complex flows of natural systems where, for example, boundaries may exhibit recirculation.

The most common implementations of boundary conditions in nested models are either relaxation or radiation conditions. Three different types of boundary conditions were tested during model development:

- a Dirichlet (or clamped) condition
- a flow relaxation condition
- a radiation condition

The Dirichlet condition is the most basic form of the relaxation condition. The parent model solution ϕ_p is directly imposed on the child grid solution ϕ_c along the nested boundary Γ such that:

$$\phi_c = \phi_p \tag{1}$$

While this condition usually ensures conservation of model properties, it can also cause reflection of outgoing waveforms if ϕ_p is incompatible with the outgoing solution ϕ_c. Incompatibilities will occur when ϕ_p is inaccurate.

The flow relaxation condition attempts to reduce the wave reflection generated by the Dirichlet condition. Here, the CG domain is extended by defining an additional domain adjacent to the CG boundary, known as the relaxation zone or *sponge layer*. The CG solution is relaxed towards the PG solution in the sponge layer so that incompatibilities between the solutions are minimised. The solution within the sponge layer is calculated according to:

$$\phi_s = (1-\tau)\phi_c + \tau\phi_p \tag{2}$$

where τ is a relaxation function which decreases from 0 on Γ (the inner boundary of the sponge layer), to 1 on the outer boundary of the sponge layer.

Finally, radiation conditions attempt to solve the problem of outgoing wave reflection by allowing them to radiate freely out through the boundary. Radiation conditions are derived from the Sommerfeld radiation condition and assume that the outgoing solution propagates through the boundary in a wave-like form according to:

$$\frac{\partial \phi_c}{\partial t} + C_\phi \frac{\partial \phi_c}{\partial n} = 0 \tag{3}$$

where C_ϕ is the phase speed of the outgoing wave and n is the direction normal to the boundary Γ. The radiation condition is only applied to outgoing variables. Incoming boundary data is therefore interpolated from ϕ_p while outgoing boundary data is calculated using (3). The determination of C_ϕ can prove problematic in complex flows (Blayo and Debreu 2005) where C_ϕ is not constant

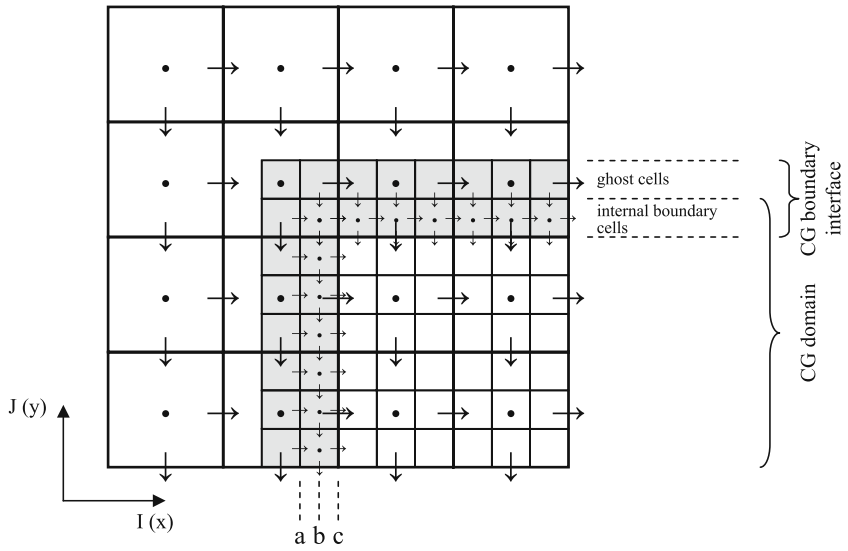

Fig. 2 Grid configuration for a 3:1 nesting ratio showing the CG boundary interface (*shaded*)

along Γ. This was found to be the case during model development. A radiation condition based on advective extrapolation was thus employed where boundary values for outgoing variables were extrapolated from the child grid solution. Due to space limitation, results from the boundary condition testing are only summarised herein. For the complete details of the implementation and testing of the boundary conditions, the reader is referred to Nash (2010).

As expected, the radiation condition was found to minimise outgoing wave reflection when incompatibilities between the child and parent solutions were large. However, neither mass nor momentum was accurately conserved for incoming waveforms, particularly in locations where the flow was strongly influenced by momentum. In addition, instabilities were also observed adjacent to the boundary at times of complex flows, i.e. when waveforms were simultaneously flowing inward and outward across the boundary. Blayo and Debreu (2005) have questioned the relevance of radiation boundaries in such complex flows and such boundary conditions have proved to give rather poor results in several comparative studies (e.g. Roed and Cooper (1987); Palma and Matano (1998); Nycander and Doos (2003)). Where incompatibilities between the child and parent solution were small, the Dirichlet condition was found to give improved accuracy over the radiation condition.

Relaxation methods have been found to be one of the best methods in comparative numerical studies (e.g. Roed and Cooper (1987); Palma and Matano (1998); Nycander and Doos (2003)). However, the meaning of the relaxed solution in the sponge layer is questionable as it is not computed by any of the system's governing equations. In addition, parent model inaccuracies, if present, are introduced into the child grid solution not only on the nested boundary but throughout the extent of the sponge layer. During testing, it was found that the sponge layer solution resulted in inaccuracies in the conservation and propagation of mass and momentum. The relaxation condition therefore produced a less accurate solution than the Dirichlet condition.

The boundary condition experimentations proved that the most stable and accurate model solution was achieved by the Dirichlet boundary condition; for this reason, it was implemented in the present model. Data was obtained from the parent solution, interpolated to the child grid, and

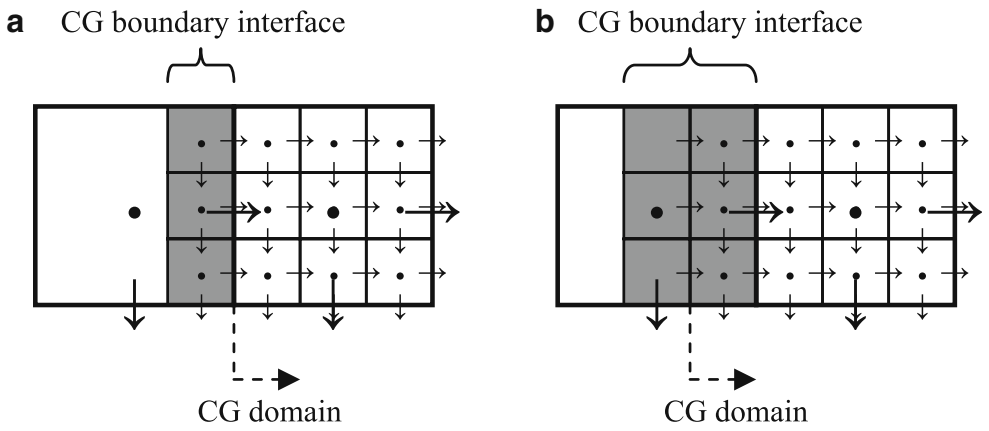

Fig. 3 CG boundary configurations used in **a** previous nesting approaches incorporating ghost cells and **b** the authors' ghost cell approach (*shaded cells* indicate PG data specification)

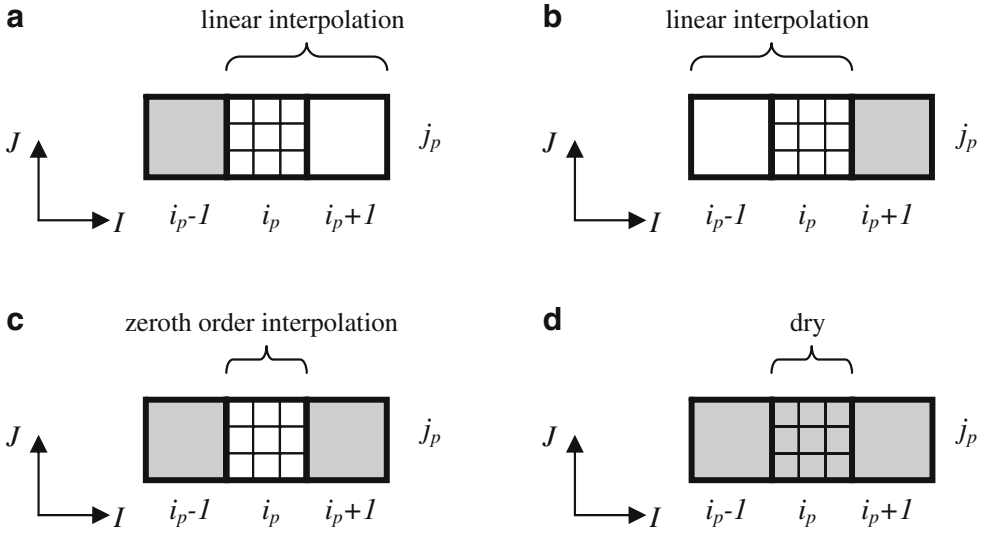

Fig. 4 Graphical illustration of the adaptive linear interpolation scheme (*shaded cells* are dry)

specified at the boundary for water elevations ζ and depth integrated velocity components in the x- and y-directions, U and V. Using these data, boundary values were also calculated for the water depths H_x and H_y and the volumetric flux components q_x and q_y. The specification of elevations alone was sufficient to ensure conservation of mass and momentum when nested boundaries were located in flows primarily driven by mass. However, for more complex boundary flows governed more by momentum, the specification of the additional variables was required to ensure conservation of both mass and momentum (see Nash (2010)).

3.2 Boundary formulation

Upon initial implementation of the Dirichlet condition, it was found that while mass conservation between the parent and child grids was relatively good, momentum conservation was poor. Investigation of this problem led to the more detailed analysis of the boundary formulation of the momentum equation. The formulation of the governing equations of flow is necessarily different for open boundary cells of a model than for interior grid cells. For example, the advective acceleration term in the x-direction momentum equation may be written in partial differential form as:

$$\beta\left[\frac{\partial Uq_x}{\partial x}+\frac{\partial Uq_y}{\partial y}\right] \tag{4}$$

where β is the momentum correction factor for a non-uniform vertical velocity profile (i.e. the ratio of the true total momentum flux through a cross-sectional area and the momentum flux estimated from the depth-averaged flow velocity).

The calculation of the partial derivative ($\partial Uq_x/\partial x$) at grid cell (x,y) using the central difference method on the model's space-staggered grid system requires values of U and q_x at the cell (x,y) and at the two adjacent grid cells ($x-\Delta x,y$) and ($x+\Delta x,y$). For an interior grid cell this is not a problem; however, at a grid cell on the boundary of the model domain, the adjacent external grid cell does not exist. At open-boundary grid cells, such normal derivatives are therefore typically set to zero; hence, the boundary cannot completely model the effect of the exterior environment on the internal solution.

Regarding a nested domain, the situation is somewhat different; the exterior environment is known. The boundary formulations were thus modified so that the normal derivatives were computed as for an internal grid cell. The missing external

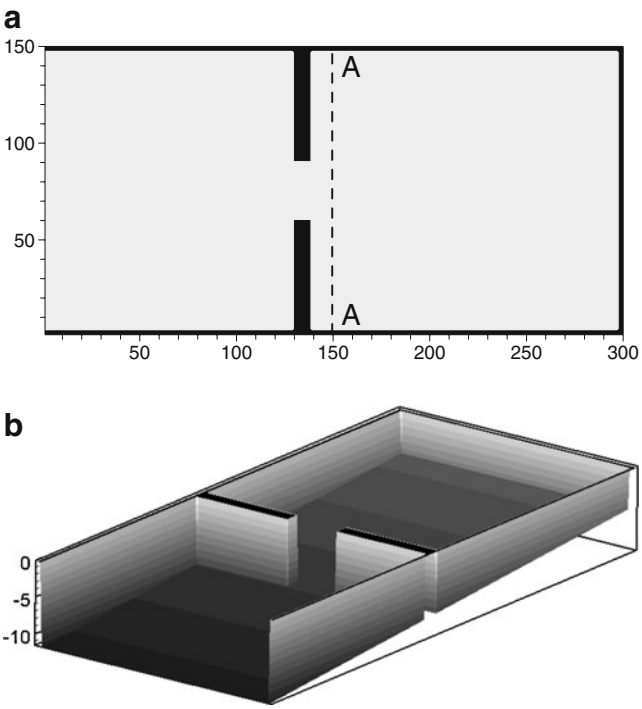

Fig. 5 **a** Plan view and **b** 3d view of model domain (depths in metres)

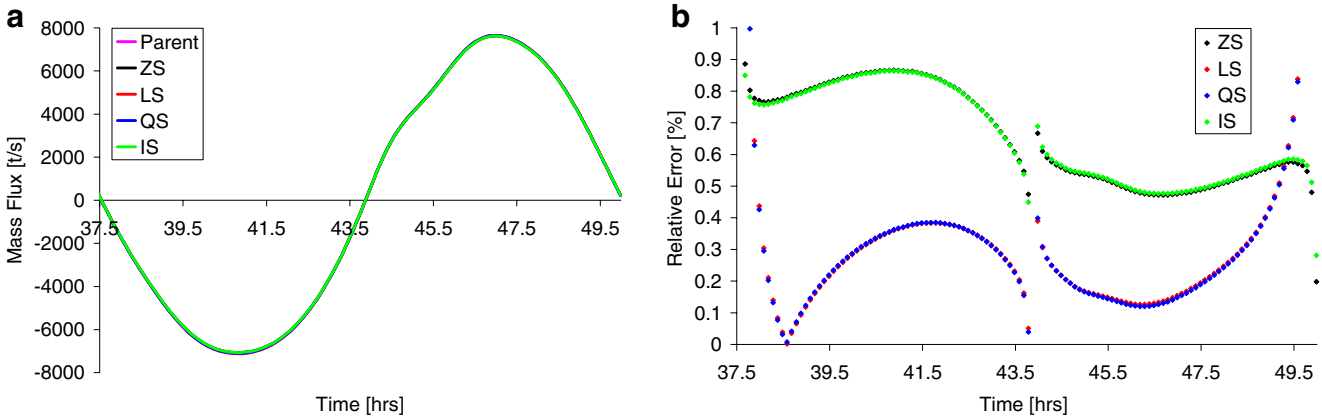

Fig. 6 **a** Boundary mass fluxes and **b** errors relative to parent model

grid cells were included in the domain as ghost cells and were incorporated in the modified boundary formulations. Using this technique, nested boundaries, in effect, become part of the interior domain or in other words, internal boundaries. To demonstrate, the finite difference formulation of the derivative ($\partial Uq_x/\partial x$) previously set to zero at an open boundary cell was instead calculated, using the ghost cells, as:

$$\frac{\partial Uq_x}{\partial x} = \left[\frac{[U(x+\Delta x,y)+U(x,y)]}{2}\cdot\frac{[q_x(x+\Delta x,y)+q_x(x,y)]}{2} - \frac{[U(x,y)+U(x-\Delta x,y)]}{2}\cdot\frac{[q_x(x,y)+q_x(x-\Delta x,y)]}{2}\right] \quad (5)$$

The CG boundary configuration developed is shown in Fig. 2. In the diagram, water elevations are indicated by ● while velocities and volumetric fluxes are indicated by →. Large symbols relate to the parent grid and small symbols to the child. It can be seen that the boundary configuration consists of two rows/columns of CG cells; the internal boundary cells and the adjacent exterior ghost cells. For clarity, child grid variables are only shown for these cells. Boundary data, interpolated from the PG, are specified at both the ghost cells and the internal boundary cells. Ghost cell data are only required for the computation of the derivatives of the components of velocity and volumetric flux normal to the boundary, thus only normal velocity and flux components are specified. For the internal boundary cells, all variable values are specified. Taking the example of the boundary parallel to the *j*-axis in Fig. 2, normal velocities and fluxes, U and q_x respectively, are specified along interface 'a' for the ghost cells and interface 'c' for the internal boundary cells. Water surface elevations, ζ, and tangential velocity and flux components, V and q_y, respectively, are also specified along interface 'b' for the internal boundary cells.

Fictitious grid cells (see Fox and Maskell (1995) and Blayo and Debreu (2008)) similar to ghost cells have previously been used in nesting algorithms. The main difference between these approaches and that presented herein is the configuration of the CG boundary interface; this difference is demonstrated graphically in Fig. 3. Based on interpretation of the literature, it is the authors' understanding that all of the previous ghost cell approaches employ the CG boundary configuration shown in Fig. 3a

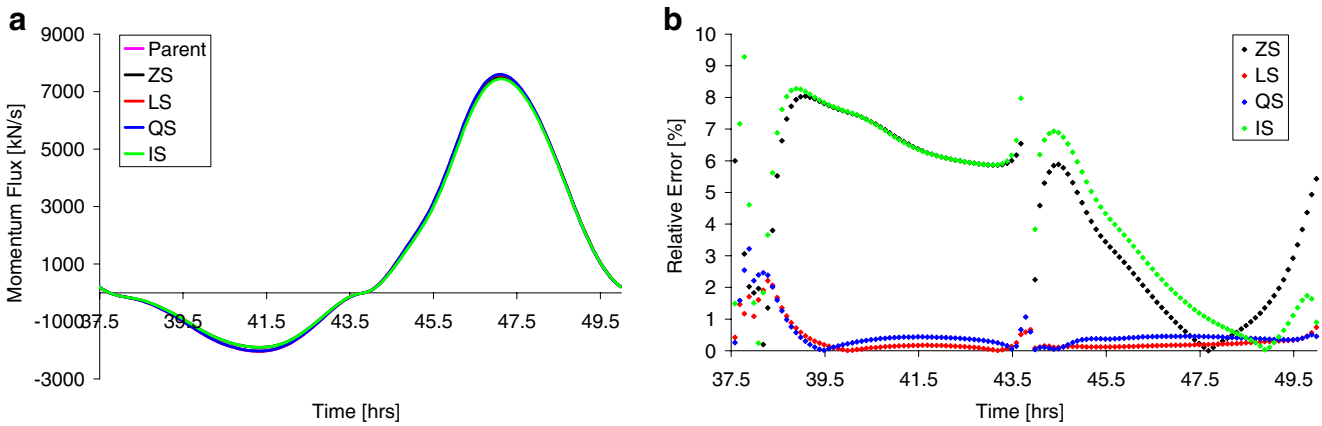

Fig. 7 **a** Boundary momentum fluxes and **b** errors relative to parent model

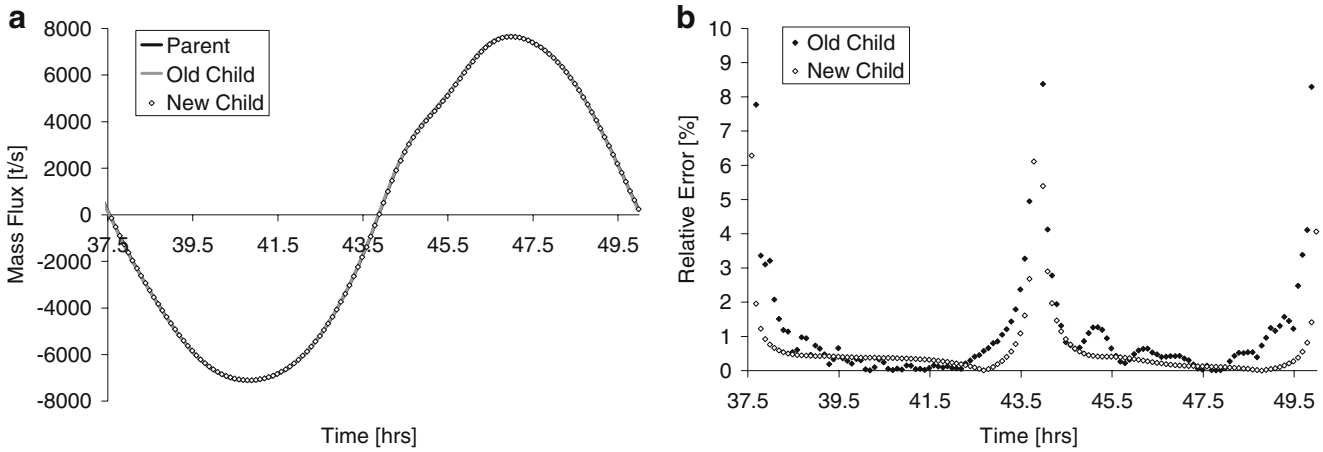

Fig. 8 **a** Boundary mass fluxes and **b** errors relative to parent model

where PG boundary data is only specified to a single column/row of cells outside of the CG domain—the ghost cells. Data values at all grid cells within the CG domain are computed by the CG model. The calculation of normal derivatives at the CG boundary is therefore equally influenced by the PG boundary data and the computed CG solution. In the authors' internal boundary approach, PG boundary data is specified to both the ghost cells outside the CG domain and to the internal boundary cells. Although the internal boundary cells are included in the CG domain, the computation of their data values by the CG model is prevented; the PG forcing is therefore strictly imposed at these cells. The approach also ensures that the calculation of normal derivatives at the CG boundary is more strongly influenced by the parent grid data; this is consistent with the strict imposition of the PG forcing.

The implementation of the internal boundary in the nested model developed by the authors was found to significantly improve the conservation of mass and, more importantly, momentum between the parent and child grids. Model results presented in Section 4.2 demonstrate the high conservation levels achieved by the authors' consistent approach.

3.3 Flooding and drying on the boundary

The flooding and drying routine used in the model was originally developed and incorporated in DIVAST by Falconer and Chen (1991). The present research was not concerned with the performance of the flooding and drying routine itself—the routine has previously been extensively tested in both idealised channels and natural estuaries and has been shown to be both stable and accurate (Falconer and Chen 1991). Rather, the research was concerned with the stability of a nested boundary located in an area of flooding and drying. The stability of a nested boundary in such an area will depend on the quality of the boundary data obtained from the parent grid and how those data are assimilated into the child grid.

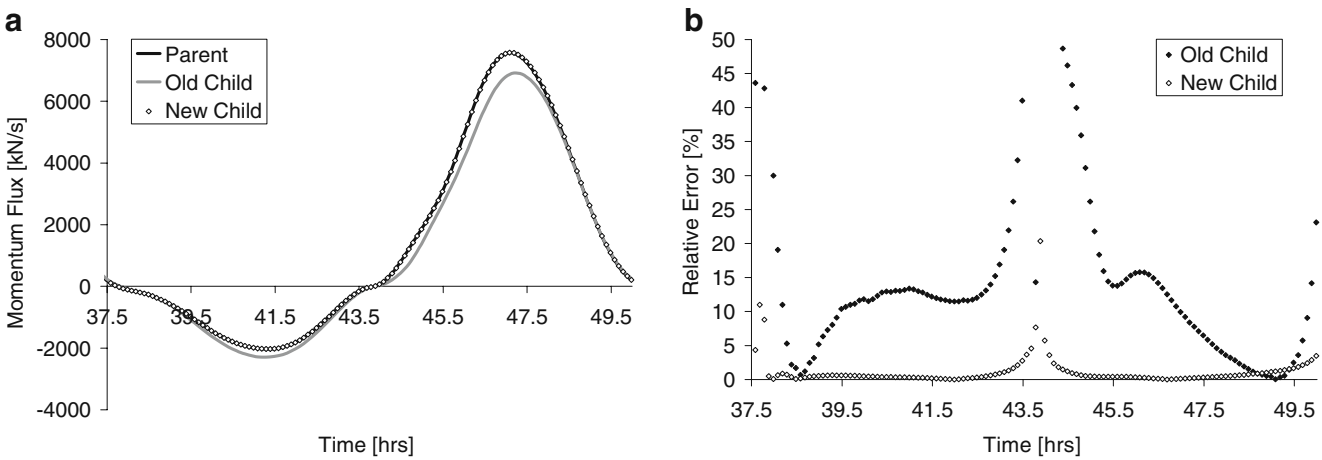

Fig. 9 **a** Boundary momentum fluxes and **b** errors relative to parent model

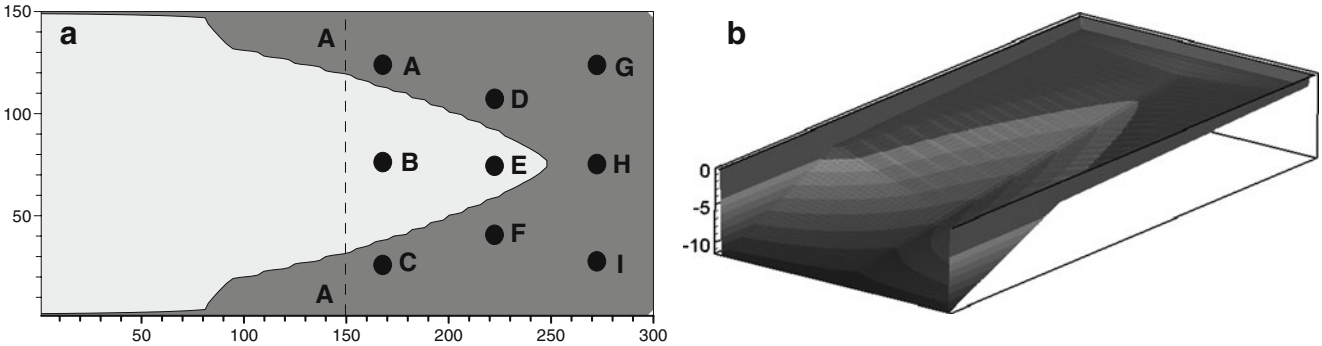

Fig. 10 **a** Plan view of model domain showing CG boundary A–A and timeseries locations (A–I), and **b** 3d view of model domain (depths in metres)

The improvements in mass and momentum conservation, resulting from the internalisation of the nested boundary, were found to improve the performance of the moving boundary. However, significant errors were still found to occur near the boundary interface in those areas subject to flooding and drying. The reason for these errors was found to be a combination of the flooding and drying routine and the boundary interpolation scheme. In the model, when a grid cell dries out, velocity components are set to zero and water level is maintained at the existing level immediately prior to drying. During interpolation of the child grid boundary data, parent grid cells which had dried out were being used in the interpolation procedure resulting in spurious boundary data values. An adaptive linear interpolation scheme was therefore designed and implemented in the model.

Figure 4 graphically illustrates how the adaptive linear interpolation scheme operates. Take the example of a parent grid cell (i_p,j_p) located on a child grid boundary interface for which the enclosed child grid cell values are required. If one of the parent grid cells to either side of (i_p,j_p) is dry, then the other adjacent parent grid cell is used in the linear interpolation formula (Fig. 4a and b); if both of the adjacent parent grid cells are dry then a zeroth-order interpolation scheme is applied (Fig. 4c) and finally, if the enclosing parent grid cell (i_p,j_p) is dry then all enclosed child grid cells are deemed dry and their variable values treated accordingly (Fig. 4d).

4 Boundary results

For each test case, two models were developed: a nested model and a single grid (SG) model. The nested model yielded two solutions for the AOI, the parent solution and the child solution. The SG model had the same extents as the PG model but was resolved at the same high resolution as the CG model; this allowed like-for-like comparisons with the CG in the AOI. The SG solution was termed the *reference solution*. Model performance was assessed by comparing PG and CG solutions with the reference solution. The determination of nested model accuracy in this way is a standard approach in cases where an analytic solution of the test case does not exist. For example, this method has been used by Spall and Holland (1991), Ginis et al. (1998), Blayo and Debreu (1999) and Rowley and Ginis (1999).

The performance of the nested boundary was primarily determined by analysis of net fluxes of mass and momentum across the boundary interface whilst the accuracy of the model solution at particular grid points was assessed by comparing timeseries of model variables. On the other hand, the accuracy of the domain-wide solution was determined using an original mathematical approach to error quantification—the tidally averaged relative error field, RE_T (Nash 2010). This approach allows the simultaneous analysis of spatial and temporal errors in a model solution in a single spatial distribution plot. The errors in both the PG and CG solutions are quantified by calculating RE_T for both solutions relative to the reference solution. Spatial distributions of model variables are output at regular time intervals (snapshots) for each domain and,

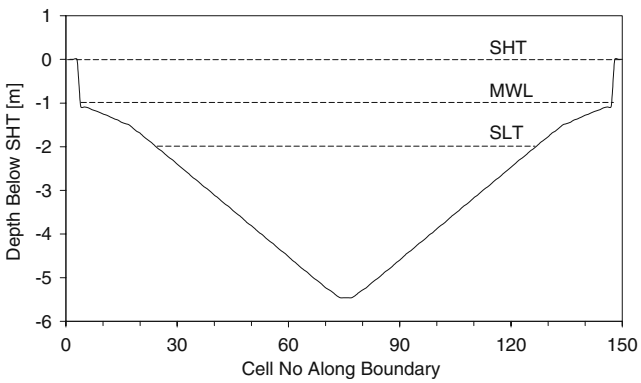

Fig. 11 Bathymetric cross-section along CG boundary (*SHT* spring high tide, *MWL* mean water level, *SLT* spring low tide)

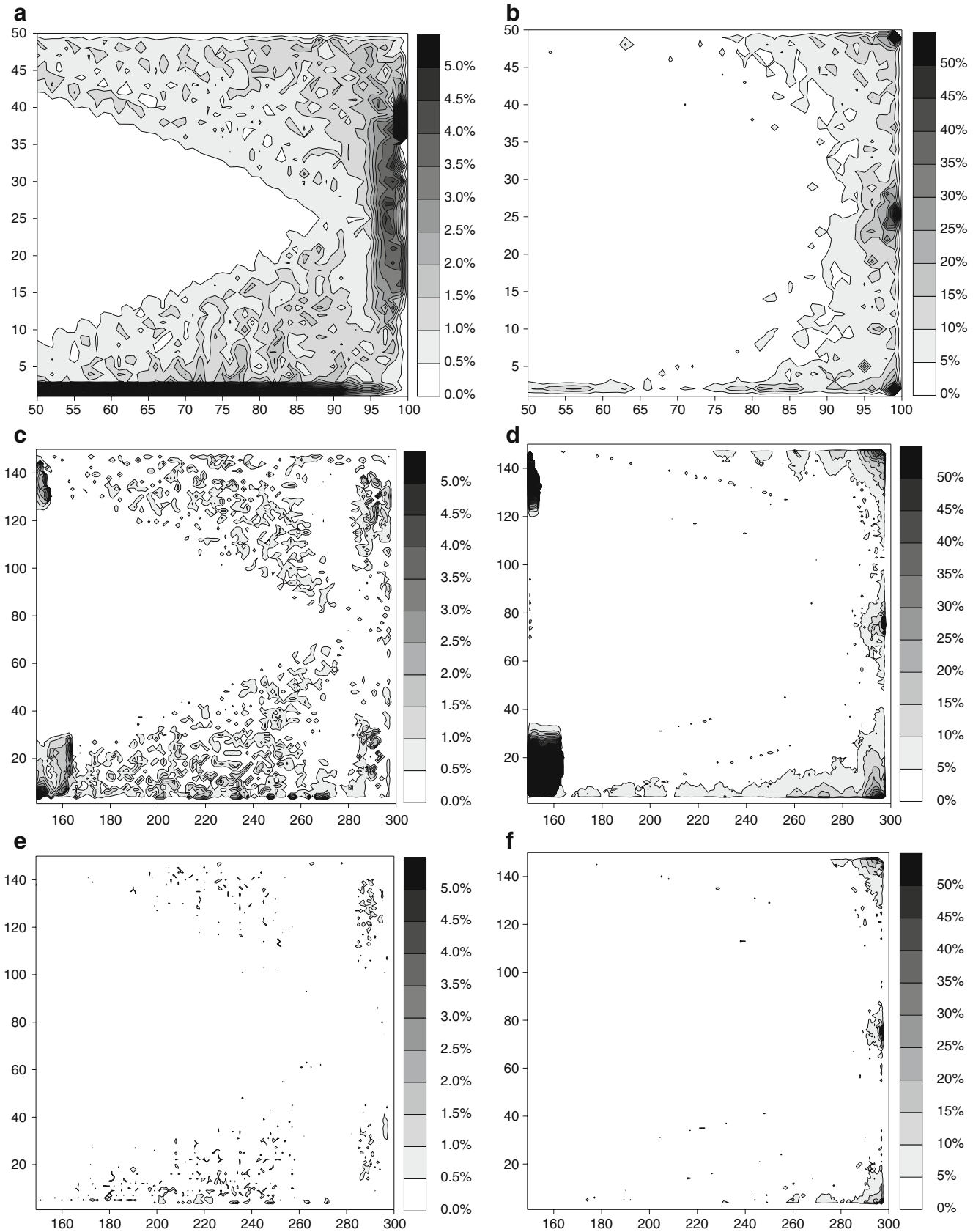

Fig. 12 **a**, **b** RE_T in PG elevations and velocities respectively, **c**, **d** RE_T in CG elevations and velocities, respectively, for standard CG boundary configuration and **e**, **f** RE_T in CG elevations and velocities, respectively, for dynamic internal CG boundary configuration

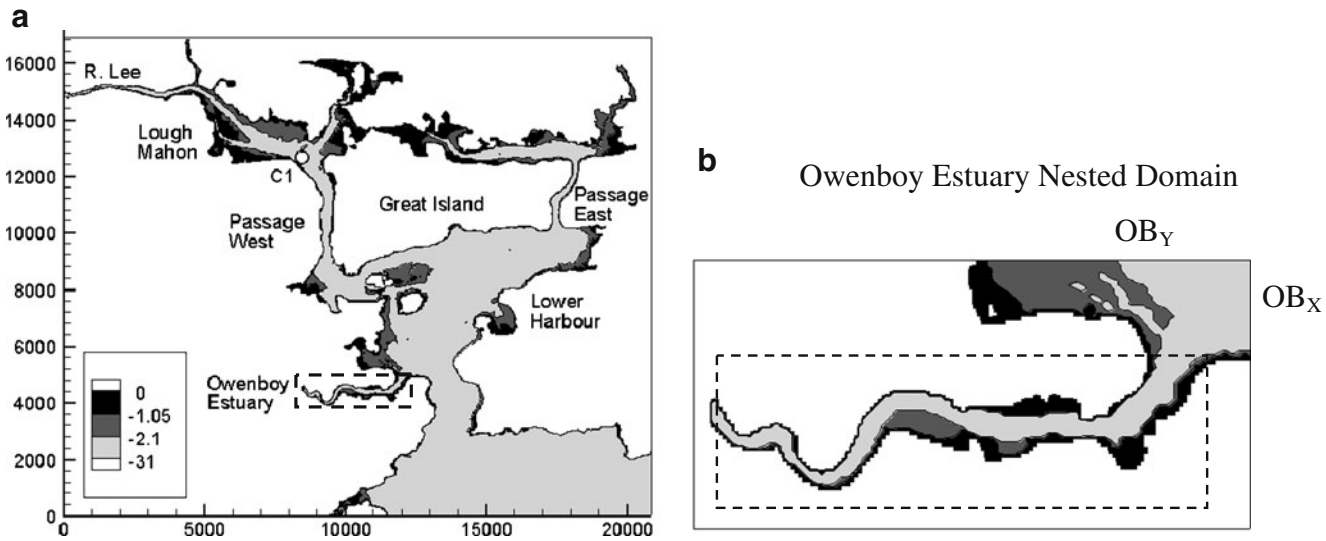

Fig. 13 **a** Cork Harbour domain and AOI (*dashed*) and **b** extents of nested domain with AOI (*dotted*). Mudflats on a spring tide are indicated by the −2.1 m contour

taking the example of the PG solution the tidally averaged relative error at a grid cell (*i,j*), RE_T (*i,j*), expressed as a percentage is then calculated as:

$$\mathrm{RE}_T(i,j) = \frac{\sum_{n=1}^{N} \left|\phi_p^n(i,j) - \phi_s^n(i,j)\right|}{\sum_{n=1}^{N} \phi_p^n(i,j)} .(100) \qquad (6)$$

where $\phi_p^n(i,j)$ and $\phi_s^n(i,j)$ are the hydrodynamic variables calculated by the PG and SG models, respectively, at grid cell (*i,j*) and at the output time of snapshot *n*. *N* is the total number of output times. To compute the tidally averaged relative error field for an AOI, Eq. 6 must be applied to each grid cell within the AOI. For a 12.5-h tidal cycle, tests established that *N= 25* was the lowest number of datasets required to give an accurate approximation of RE_T at a particular grid point (see Nash (2010) for more details). Datasets were therefore output at half-hourly intervals over the course of a single tidal cycle.

4.1 Boundary operator

The boundary operator was developed and tested using an idealised rectangular harbour with constant bed slope (see Fig. 5). The harbour dimensions were 12×6 km and the bed depth decreased from 10 m at the open boundary to 4 m at the back of the harbour. For simplicity a single nested grid was used at a 3:1 nesting ratio. The parent model used a grid spacing of 120 m and a timestep of 120 s while the child model used a grid spacing of 40 m and a timestep of 40 s. The CG boundary (A–A in Fig. 5) spanned the full width of the harbour and was located exactly halfway along

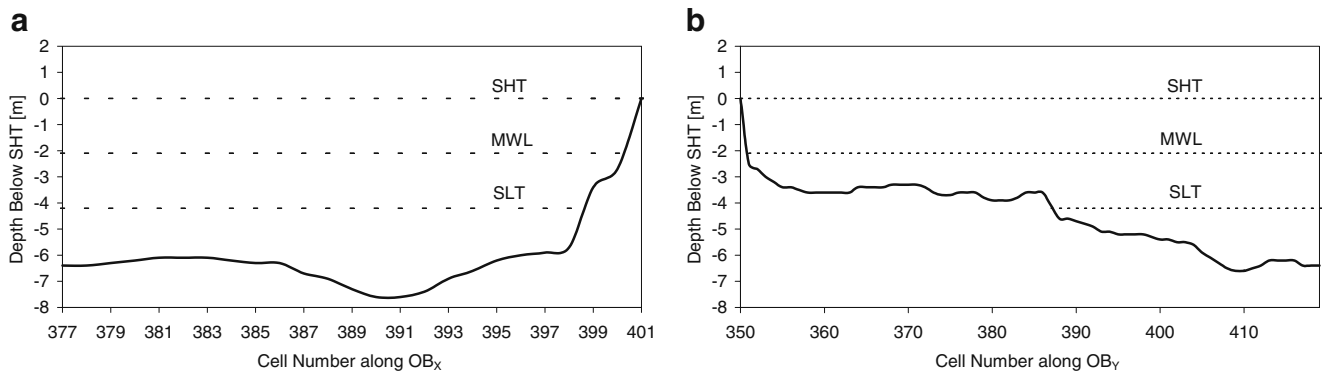

Fig. 14 Bathymetric cross-sections along **a** boundary $\mathrm{OB_X}$ and **b** boundary $\mathrm{OB_Y}$ (*SHT* spring high tide, *MWL* mean water level, *SLT* spring low tide)

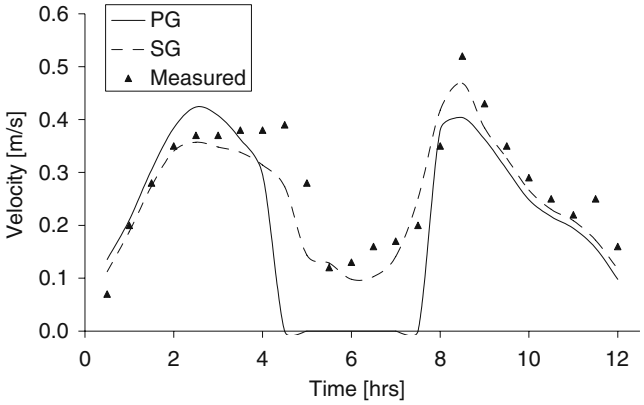

Fig. 15 Comparison of computed and measured velocities at point C1 in Fig. 13a

the longitudinal axis; the CG domain therefore comprised the inner half of the harbour. A tide of constant period (12.5 h) and range (3 m) was used as the model forcing.

A harbour wall was included in the model to induce momentum-driven flows. The CG boundary was intentionally located in an area where flows were significantly influenced by momentum. Unfortunately, this also corresponded to a location of high parent model inaccuracy meaning that BS errors were introduced to the child model; however, the choice of location was judicious in developing a rigorous test of the boundary operator.

4.2 Spatial interpolation

Conservation of mass and momentum between parent and child grids is directly affected by the accuracy of the spatial interpolation scheme. Performance of four different schemes was assessed by analysing net fluxes of mass and momentum across the nested boundary interface (i.e. in the direction normal to the boundary interface). Child model fluxes were computed and compared with the corresponding fluxes from the parent model. Comparisons for a full tidal cycle are shown in Figs. 6 and 7; the errors in child model flux relative to parent flux are also shown as these errors are not easily apparent from the flux graphs due to the scales involved. Fluxes were output immediately after the spatial interpolation of the boundary data so that any error could only be attributed to the interpolation scheme.

The LS and QS were found to achieve similarly high levels of conservation. The ZS and IS also achieved similar, but lower, levels of conservation. The superior performance of the LS and QS was attributed to their mass conservation properties. In the case of the LS, conservation was naturally inherent to the scheme but in the case of the QS conservation was ensured by the introduction of a mass conservation parameter following Clark and Farley (1984).

Overall, the LS achieved the highest levels of conservation of both mass and momentum; the tidally averaged errors were acceptably low at 0.27% for mass and 0.32% for momentum. While the QS produced almost the same level of mass conservation, the level of momentum conservation was slightly lower than that of the LS. The results from the analysis agree with literature; Alapaty et al. (1998) state that higher-order schemes often lead to spurious oscillations of the interpolated variable in regions of sharp gradients and are thus less accurate than linear schemes. It should be noted that the highest errors were observed at the times of near-zero fluxes and were therefore not significant in terms of conservation.

4.2.1 Boundary formulations

The effect of the internal boundary formulation on model performance was also best demonstrated by the comparison of fluxes across the boundary interface. Figures 8 and 9 compare the boundary fluxes of mass and momentum from the child model, before and after the incorporation of the ghost cells, with those from the parent model. The time of

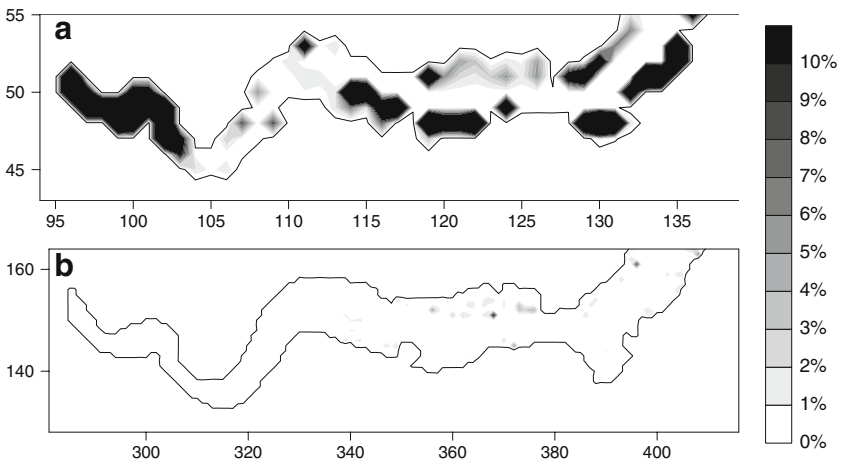

Fig. 16 RE_T in water elevation in the AOI for **a** parent model and **b** child model

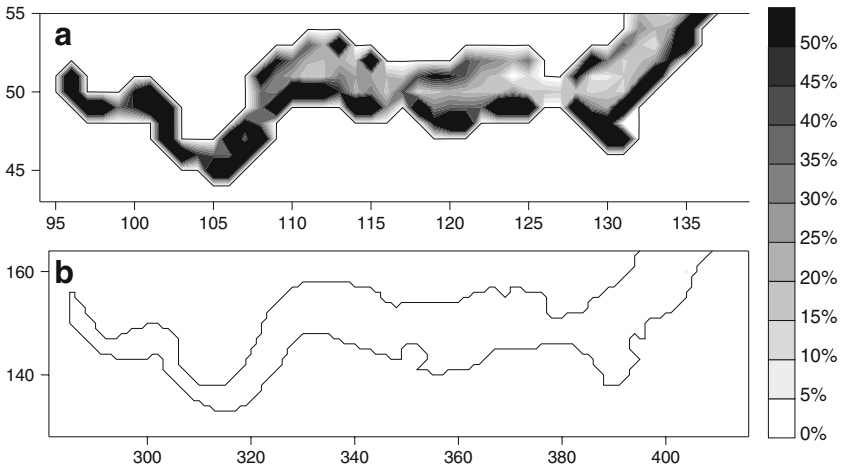

Fig. 17 RE_T in current velocity in the AOI for **a** parent model and **b** child model

output of the fluxes differed to those shown in Figs. 6 and 7; in this case, fluxes were output at each timestep following the solution of the governing equations of flow.

The ghost cell approach to boundary formulation significantly improved momentum conservation between the parent and child grids. The tidally averaged error in boundary momentum flux decreased from 10.5% prior to the implementation of the internal boundary to 0.4% afterwards. The error in mass conservation, although already quite low prior to implementation, also decreased from 0.6% to 0.3%. The improvement in mass and momentum conservation led to significant improvements in the accuracy of the nested model.

4.3 Moving nested boundary

The new approach to the moving boundary problem was tested on an idealised harbour which was subject to flooding and drying (see Fig. 10). The parent and child grid extents were the same as those used in the rectangular harbour experiment. The models were forced with a tide of constant period (12.5 h) and range (4 m). In order to properly test the moving boundary, the harbour bed was shaped to ensure that flooding and drying occurred both inside the CG domain and on the CG boundary (A–A in Fig. 10a). Figure 11 shows a bathymetric cross-section along the CG boundary A–A. From the diagram, it can be seen that approximately one third of the boundary's length is dry at low tide.

The improvement in the simulation of the moving nested boundary, resulting from the implementation of the internal dynamic boundary, was best demonstrated by examination of the RE_T in CG water elevations and current velocities. Figure 12 compares the RE_T in CG water elevations and current velocities computed by the nested model using the standard open boundary configuration with those computed by the nested model using the dynamic internal boundary. To demonstrate the improvement in accuracy achieved by the CG model, the RE_T in the PG solution is also shown. It can be seen that whilst using the standard open boundary configuration resulted in an improvement in accuracy in the nested domain, large errors still occurred on and near the boundary in those areas subject to flooding and drying. This was due to the poor conservation of mass and momentum and the use of data values from dry PG cells in the interpolation of the CG boundary data. Errors can also be seen just inside the central section of the boundary; these were also a result of the poor conservation of momentum resulting from the simplified standard open boundary formulations. Neither type of error occurred when the dynamic internal boundary was used; as a result, the accuracy of the CG solution was similar to that of the SG reference solution with RE_T in both elevations and velocities very close to zero across the full CG domain. It should be noted that the elevated relative errors at the back of the harbour in Fig. 12f represented absolute errors of less than 0.001 m/s; these relative errors were therefore insignificant.

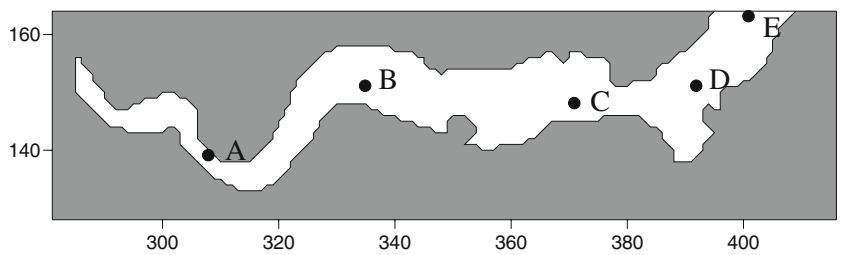

Fig. 18 Time series locations in the AOI

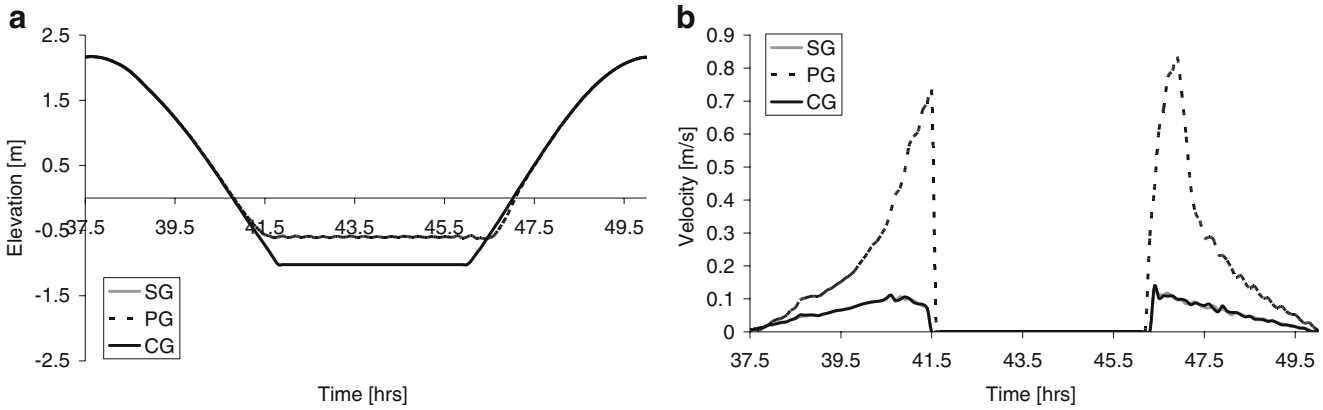

Fig. 19 **a** Water elevations and **b** current velocities at point A (SG and CG coincident)

5 Application to a natural harbour

The most important elements of nested model development were the application of the model to a natural coastal system and the assessment of its performance. Cork Harbour, located on the southwest coast of Ireland, was chosen as the test case for model application. The Harbour, shown in Fig. 13a, is one of the largest sea inlets in Ireland, with just under 120 miles of coastline. It is a macro-tidal estuary with typical spring tide ranges of 4.2 m. The deepest point is 29 m below mean water level (MWL) and occurs in the entrance channel to the Lower Harbour while the average water depth is 8.4 m below MWL. At low water, extensive areas of mud and sand flats are exposed throughout the Harbour; it was primarily for this reason that it was chosen as a test case. In addition, the complex coastline and bathymetry of the region mean that the accuracy of a numerical model is highly dependent on the spatial resolution.

The domain chosen for nesting was the Owenboy Estuary; the extents of the domain and the AOI are shown in Fig. 13b. The region was chosen as it has an extensive inter-tidal zone with large areas of the wetted domain being exposed at low spring tide. Most importantly, flooding and drying was prevalent on both open boundaries OB_X and OB_Y. Indeed, almost half of the boundary OB_Y dries out at low tide; this can be seen in Fig. 14 which shows bathymetric cross-sections of both open boundaries.

The parent model comprising the full domain of Fig. 13a was resolved at a grid spacing of 90 m and a timestep of 18 s. A 3:1 nesting ratio was used, thus the child model was resolved at a grid spacing of 30 m and a timestep of 6 s. The accuracy of the PG and CG solutions was determined relative to a reference solution—the solution from a SG model with the same domain extents as the parent grid and the same high resolution as the child grid. The models were run for a constant spring tide so that the areas of flooding and drying were maximised. The SG model had been extensively calibrated and validated during previous research (see Costello et al. (2001)). To demonstrate that the high resolution did indeed give improved model accuracy, Fig. 15 compares current velocities computed by the high resolution SG model and the low resolution PG model with measured velocities at point C1 (see Fig. 13a) over a single spring tidal cycle. It can be seen that the PG model erroneously computed a period of drying at low water (indicated by the zero velocities); this was not reflected in

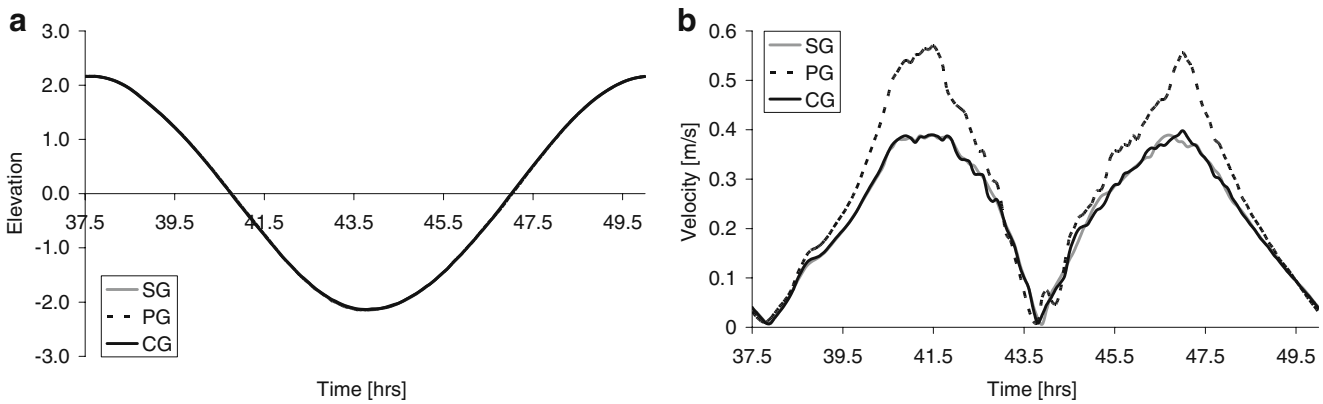

Fig. 20 **a** Water elevations and **b** current velocities at point C (SG and CG coincident)

the measured data, nor was it computed by the SG model. The improved accuracy of the SG solution is clear to be seen.

5.1 Model accuracy

In order to assess model performance, analysis of model results was only carried out for the AOI. Figure 16 compares the RE_T in PG and CG water elevations in the AOI. Typically, water elevations will tend to be quite accurate irrespective of model resolution. Overall, errors were quite low with the areas of largest errors occurring in the inter-tidal areas (see Fig. 13b). The nested model achieved a significant reduction in model error. While the RE_T in the parent model exceeded 1% in 47% of the AOI, this level was only exceeded in the child model in 4% of the AOI. The average RE_T value for the AOI was found to decrease from 8.3% in the parent model to 0.6% in the child.

The nested model also significantly improved the accuracy of the current velocity computations. Figure 17 compares the RE_T in PG and CG velocities in the AOI. Firstly, it can be seen that errors in the parent model were much greater for current velocities than for water surface elevations; this was as expected. Secondly, the RE_T in the parent model velocities exceeded 5% in 86% of the AOI; by contrast, this level was never exceeded in the child model. In general, the average RE_T in velocities for the AOI decreased from 45% in the parent model to 0.1% in the child.

In order to verify the accuracy of the nested model, time series of water elevations and current velocity magnitudes were analysed at a number of different locations (see Fig. 18). Due to space limitations, time series comparisons are only presented herein for points A (Fig. 19) and C (Fig. 20). Water elevations computed by the parent model were quite accurate for the most part; however, large inaccuracies were present in the current velocities; this agreed with the spatial error distributions. An improvement in the accuracy of water elevations in those areas subject to flooding and drying can be seen at location A. The lower resolution of the parent model was unable to resolve the drying of grid cells to the same level of accuracy as the high resolution model; a 0.5-m error was recorded in the level at which the grid cell was deemed to have dried out. In contrast, the child model was capable of resolving the inter-tidal region to the same accuracy as the single grid model.

Significant improvements in the accuracy of velocity computations were achieved by the child model. At point A, the tidally averaged error in velocities was reduced from 329% in the parent model to 3% in the child model; at point C, the error was reduced from 25% to 2%. The time series show that the nested model achieves a similar level of accuracy to the high-resolution model. This was verified by using linear regression of the time series data (see Fig. 21) to calculate the degree of correlation, r^2, between the CG and SG velocities. As can be seen, r^2 values of 0.99 and 1 were computed for the CG velocities at points A and C, respectively. In addition to the high r^2 values, it can be seen from the equations of the CG trend lines that the slopes were equal to 1 and the intercepts were zero; this further demonstrates the high level of accuracy of the nested model.

6 Conclusions

The authors have developed a nested circulation model incorporating flooding and drying which has significant and practical applications in the modelling of coastal areas and,

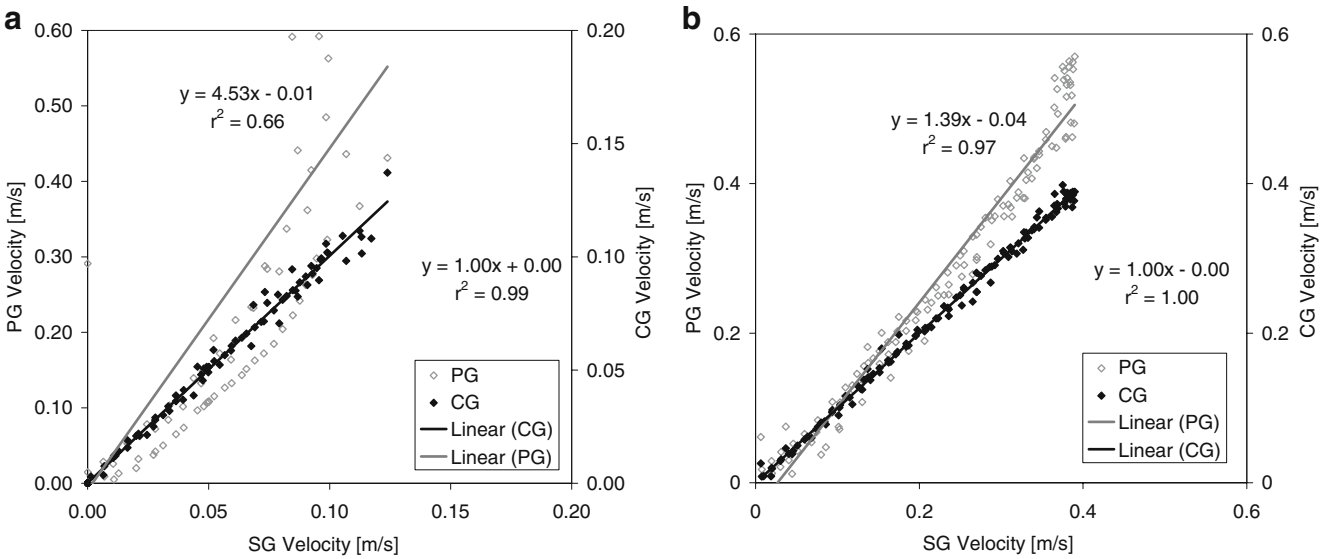

Fig. 21 Linear regression analyses of velocity time series data showing trend lines and r^2 values for **a** Point A and **b** Point C

in particular, inter-tidal zones. The accurate modelling of inter-tidal zones requires a high resolution which can be cost prohibitive. The nested model allows high-resolution modelling in an AOI but at a significantly lower computational cost than a high-resolution single-grid model. For example, the application of the model to the Owenboy estuary presented herein resulted in a 92% saving in computation time over an equivalent large domain high-resolution model. The model results demonstrated that the model not only offered improved accuracy over the lower resolution parent model but that the level of accuracy achieved in the AOI was similar to that of the high-resolution model.

The nested modelling of inter-tidal zones requires the incorporation of moving nested boundaries to accommodate flooding and dying of boundary grid cells. Faced with this problem, nested models have traditionally either prohibited the flooding and drying of nested boundaries or omitted the process completely. The model developed by the authors solves the moving boundary problem through the incorporation of ghost cells in the mathematical formulations of the nested boundaries. Nested boundary cells are formulated in a similar manner to interior grid cells by specifying ghost cells adjacent to the boundary and obtaining data for both the ghost cells and the internal boundary cells from the parent model. In combination with an adaptive linear interpolation scheme designed specifically for the model, nested boundaries are, in effect, converted into dynamic internal boundaries; this enables the stable flooding and drying of boundary grid cells.

Finally, errors are introduced to a nested model at the boundary interface; particular attention was, therefore, focused on the development of the boundary operator. A linear interpolation technique and a Dirichlet boundary condition were found to give the highest levels of conservation of mass and momentum between parent and child grids. The internalisation of the boundary was also found to significantly improve conservation at the boundary by allowing the boundary to better model the effects of the external environment on the internal hydrodynamics.

Acknowledgment The authors would like to acknowledge financial support from INTERREG IV -Atlantic Area Programme under the Marine Renewable Energy (MAREN) project.

Appendix A: spatial interpolation formulae

Zeroth-order interpolation scheme

If ϕ_i represents the value of a parent grid variable at cell i, then ϕ_k for all the child grid cells within its confines can be written as:

$$\phi_k = \phi_i \quad k = 1, ..., m \tag{9}$$

where m is the number of child grid cells within the parent grid cell, i.e. $m = r_s^2, r_s$ being the spatial nesting ratio.

Linear interpolation scheme

The variable ϕ in a child grid cell k which lies within the parent grid cell i may be calculated as:

$$\phi_k = \phi_{i-1} + \omega(\phi_i - \phi_{i-1}) \tag{10}$$

with ω, the proportional interpolation coefficient, further expressed as:

$$\omega = \frac{2k + r_s - 1}{2r_s}, \quad k = 1, ..., r_s \tag{11}$$

Quadratic interpolation scheme

For a given nesting ratio r_s, the interpolation of the parent grid cell variable ϕ_i to any child grid cell k of any horizontal row of enclosed child grid cells may be specified as (adapted from Alapaty et al. (1998)):

$$\phi_k = E_{i-1}^k \phi_{i-1} + E_i^k \phi_i + E_{i+1}^k \phi_{i+1} \tag{12}$$

with the function E further expressed as:

$$E_{i-1}^k = \frac{\lambda_k(\lambda_k - 1)}{2} + \alpha \tag{13}$$

$$E_i^k = \left(1 - \lambda_k^2\right) - 2\alpha \tag{14}$$

$$E_{i+1}^k = \frac{\lambda_k(\lambda_k + 1)}{2} + \alpha \tag{15}$$

where λ represents a normalised local coordinate pointing in the same direction as the global coordinate and whose origin coincides with the centre of the parent grid cell i. The value of λ for the child grid cell k is defined as:

$$\lambda_k = \frac{(2k-1)\Delta x_c - \Delta x_p}{2\Delta x_p}, \quad k = 1, ..., r_s \tag{16}$$

and

$$\alpha = \frac{1}{24}\left[\left(\frac{\Delta x_c}{\Delta x_p}\right)^2 - 1\right] \tag{17}$$

Δx_p and Δx_c are the parent and child grid spacings, respectively. The parameter α is introduced to ensure mass conservation following Clark and Farley (1984).

Inverse distance weighted interpolation scheme

The variable ϕ in a child grid cell k which lies within an enclosing parent grid cell is calculated as:

$$\phi_k = \sum_{i=1}^{n} w_i \phi_i / \sum_{i=1}^{n} w_i \quad k = 1, ..., m \tag{18}$$

where n is the number of parent grid cells used in the interpolation and w_i is the weighting function. For the nested model n=9, i.e. the enclosing grid cell and the eight adjacent grid cells were used. The weighting function is written as:

$$w_i = \frac{1}{d_{(k,i)}{}^2} \tag{19}$$

where $d_{(k,i)}$ is the distance from the child grid point k to the parent grid point i. The weighting function varies from a value of unity at the child grid point of interest to a value approaching zero as the distance from the grid point increases.

References

Alapaty K, Mathur R, Odman T (1998) Intercomparison of spatial interpolation schemes for use in nested grid models. Mon Weather Rev 126:243–249

Barth A, Alvera-Azcárate A, Rixen M, Beckers J-M (2005) Two-way nested model of mesoscale circulation features in the Ligurian Sea. Prog Oceanogr 66:171–189

Blayo E, Debreu L (1999) Adaptive mesh refinement for finite-difference ocean models: first experiments. J Phys Oceanogr 29:1239–1250

Blayo E, Debreu L (2005) Revisiting open boundary conditions from the point of view of characteristic variables. Ocean Model 9:231–252

Blayo E, Debreu L (2008) Two-way embedding algorithms: a review. Ocean Dyn 58:415–428

Clark TL, Farley RD (1984) Severe downslope windstorm calculations in two and three dimensions using anelastic interface grid nesting: a possible mechanism for gustiness. J Atmos Sci 41:329–350

Costello MJ, Hartnett M, Mills P, O'Mongain E, Collier L, Johnson M, Nash S, Leslie R, Berry A, Emblow C, Collins A, McCrea M (2001) Nutrient dynamics of two estuaries in Ireland: Wexford and Cork Harbours, Irish Environmental Protection Agency

DHI Software (2007) Mike 21 flow model: hydrodynamic module user guide, DHI Water and Environment

Falconer RA (1984) A mathematical model study of the flushing characteristics of a shallow tidal bay. Proc Inst Civil Eng 2 Res and Theory 77(3):311–332

Falconer RA, Chen YP (1991) An improved representation of flooding and drying and wind stress effects in a 2-D tidal numerical model. Proc Inst Civil Eng 2 Res and Theory 91:659–678

Fox AD, Maskell SJ (1995) Two-way interactive nesting of primitive equation ocean models with topography. J Phys Oceanogr 25:2977–2996

Ginis I, Richardson RA, Rothstein LM (1998) Design of a multiply nested primitive equation ocean model. Mon Weather Rev 126:1054–1079

Holt J, Harle K, Proctor R, Michel S, Ashworth M, Batstone C, Allen I, Holmes R, Smyth T, Haines K, Bretherton D, Smith G (2009) Modelling the global coastal ocean. Philos Trans R Soc A 367:939–951

Koch SE, McQueen JT (1987) A survey of nested grid techniques and their potential for use within the MASS weather prediction model, NASA Technical Memorandum 87808, National Aeronautics and Space Administration, USA

Korres G, Lascaratos A (2003) A one-way nested eddy resolving model of the Aegean and Levantine Basins: implementation and climatological runs. Ann Geophys 21:205–220

Leitão P, Coelho H, Santosa A, Nevesa R (2005) Modelling the main features of the Algarve coastal circulation during July 2004: a downscaling approach. J Atmos Ocean Sci 10(4):421–462

Lin B, Falconer RA (1997) Tidal flow and transport modelling using the ULTIMATE QUICKEST scheme. J Hydraul Eng 123(4):303–314

Marchesiello P, McWilliams J, Shchepetkin A (2001) Open boundary conditions for long-term integration of regional oceanic models. Ocean Model 3:1–20

Miyakoda K, Rosati A (1977) One-way nested grid models: the interface conditions and the numerical accuracy. Mon Weather Rev 105:1092–1107

Nash S (2010) Development of an adaptive mesh inter-tidal circulation model, PhD Thesis, College of Engineering & Informatics, National University of Ireland Galway

Nittis K, Perivoliotis L, Korres G, Tziavos C, Thanos I (2006) Operational monitoring and forecasting for marine environmental applications in the Aegean Sea. Environ Modell Softw 21:243–257

Nycander J, Doos K (2003) Open boundary conditions for barotropic waves. J Geophys Res 108(C5):3168–3187

Palma ED, Matano RP (1998) On the implementation of passive open boundary conditions for a general circulation model: the barotropic mode. J Geophys Res 103(C1):1319–1341

Pullen J, Allen JS (2001) Modeling studies of the coastal circulation off the northern coast of California: statistics and patterns of wintertime flow. J Geophys Res 106(C11):26959–26984

Roed LP, Cooper C (1987) A study of various open boundary conditions for wind-forced barotropic numerical ocean models. In: Nihoul JCJ, Jamart BN (eds) Three-dimensional models of marine and estuarine dynamics. Elsevier, pp 305–335

Rowley C, Ginis I (1999) Implementation of a mesh movement scheme in a multiply nested ocean model and its application to air -sea interaction studies. Mon Weather Rev 127:1879–1896

Spall MA, Holland WR (1991) A nested primitive equation model for oceanic applications. J Phys Oceanogr 21:205–220

Spall MA, Robinson AR (1989) A new hybrid coordinate open ocean primitive equation model. Math Comput Simul 31:241–269

Staneva J, Stanev EV, Wolff JO, Badewien TH, Reuter R, Flemming B, Bartholoma A, Boldinge K (2009) Hydroynamics and sediment dynamics in the German Bight: a focus on observations and numerical modelling in the East Frisian Wadden Sea. Cont Shelf Res 29:302–319

Zhang D-L, Chang H-R, Seaman NL, Wamer TT, Fritsch JM (1986) A two-way interactive nesting procedure with variable terrain resolution. In: Koch SE, McQueen JT (1987)

Ocean Dynamics (2010) 60:1497–1537
DOI 10.1007/s10236-010-0349-4

Multiscale two-way embedding schemes for free-surface primitive equations in the "Multidisciplinary Simulation, Estimation and Assimilation System"

Patrick J. Haley Jr. · Pierre F. J. Lermusiaux

Received: 22 June 2010 / Accepted: 10 October 2010 / Published online: 11 November 2010

Abstract We derive conservative time-dependent structured discretizations and two-way embedded (nested) schemes for multiscale ocean dynamics governed by primitive equations (PEs) with a nonlinear free surface. Our multiscale goal is to resolve tidal-to-mesoscale processes and interactions over large multiresolution telescoping domains with complex geometries including shallow seas with strong tides, steep shelfbreaks, and deep ocean interactions. We first provide an implicit time-stepping algorithm for the nonlinear free-surface PEs and then derive a consistent time-dependent spatial discretization with a generalized vertical grid. This leads to a novel time-dependent finite volume formulation for structured grids on spherical or Cartesian coordinates, second order in time and space, which preserves mass and tracers in the presence of a time-varying free surface. We then introduce the concept of two-way nesting, implicit in space and time, which exchanges all of the updated fields values across grids, as soon as they become available. A class of such powerful nesting schemes applicable to telescoping grids of PE models with a nonlinear free surface is derived. The schemes mainly differ in the fine-to-coarse scale transfers and in the interpolations and numerical filtering, specifically for the barotropic velocity and surface pressure components of the two-way exchanges. Our scheme comparisons show that for nesting with free surfaces, the most accurate scheme has the strongest implicit couplings among grids. We complete a theoretical truncation error analysis to confirm and mathematically explain findings. Results of our discretizations and two-way nesting are presented in realistic multiscale simulations with data assimilation for the middle Atlantic Bight shelfbreak region off the east coast of the USA, the Philippine archipelago, and the Taiwan–Kuroshio region. Multiscale modeling with two-way nesting enables an easy use of different sub-gridscale parameterizations in each nested domain. The new developments drastically enhance the predictive capability and robustness of our predictions, both qualitatively and quantitatively. Without them, our multiscale multiprocess simulations either were not possible or did not match ocean data.

Keywords Embedding schemes · Multiscale ocean modeling · Shelfbreak regions · Coastal dynamics · Tidal forcing · Multiresolution · Multigrid · CFD

Responsible Editor: Eric Deleersnijder

We thank the Office of Naval Research for research support under grants N00014-07-1-1061 and N00014-08-1-1097 (6.1), N00014-07-1-0501 (AWACS-SW06), N00014-07-1-0473 (PHILEX), and N00014-07-1-0241 and N00014-08-1-0586 (QPE) and to the Massachusetts Institute of Technology.

P. J. Haley Jr. · P. F. J. Lermusiaux (✉)
Massachusetts Institute of Technology, 77 Mass. Avenue, Cambridge, MA 02139, USA
e-mail: pierrel@mit.edu

P. J. Haley Jr.
e-mail: phaley@mit.edu

1 Introduction

Ocean dynamics is now known to involve multiple scales and dynamical interactions with inherent transient effects and intense localized gradients. Sources

of interacting scales and intermittent behavior include internal nonlinear dynamics, steep bathymetries, complex geometries, and remote and boundary forcing. To predict such dynamics, ocean modeling systems must be capable of multiresolution, multiscale, and multidynamic numerical simulations. A major objective of our present research is to derive and study robust and accurate two-way embedding (nesting) schemes for telescoping ocean domains governed by primitive equation (PE) dynamics with a nonlinear free surface. The intent is to resolve tidal-to-mesoscale dynamics over large multiresolution domains with complex coastal geometries from embayments and shallow seas with strong tidal flows to the steep shelfbreaks and the deep ocean with frontal features, jets, eddies, and other larger-scale current systems.

Most structured-grid models have been developed to be general, with applications in varied ocean regions (e.g., Lynch and Davies 1995; Mooers 1999). These modeling systems include Modular Ocean Model (Griffies et al. 2007, 2010), Navy Layer Ocean Model/Deep-ocean Assessment and Reporting of Tsunamis (Carnes et al. 1996; Wallcraft et al. 2003), Regional Ocean Model System (Haidvogel et al. 2000; Shchepetkin and McWilliams 2005), Princeton Ocean Model (POM; Blumberg and Mellor 1987; Mellor 2004), Parallel Ocean Program (Smith et al. 1992), MIT General Circulation Model (Marshall et al. 1997), Terrain-following Ocean Modeling System (Ezer et al. 2002; Arango et al. 2010), Hybrid Coordinate Ocean Model (HYCOM; Bleck 2002; Chassignet et al. 2009), and Harvard Ocean Prediction System (HOPS; Robinson 1999; Haley et al. 1999). Examples of applications of these models as well as others include for the US eastern coastal oceans (Signell et al. 2000; Lynch et al. 2001; Robinson et al. 1999, 2001), northwestern Atlantic (Chassignet et al. 2000; Chassignet and Malanotte-Rizzoli 2000), Atlantic Ocean (Chassignet et al. 2003; Stammer and Chassignet 2000), Pacific Ocean and US western coastal oceans (de Szoeke et al. 2000; Chao et al. 2009), Mediterranean Sea (Pinardi and Woods 2002; Onken et al. 2003, 2008), European North Seas (Berntsen and Svendsen 1999), and basins and the global ocean (Semtner 2000; Dutay et al. 2002; Marshall et al. 1997; Gent et al. 1998). More recently, unstructured algorithms have been applied to simulate multiscale ocean dynamics and processes (Deleersnijder and Lermusiaux 2008a, b). Here we focus only on the use of conservative structured and embedded grid approaches to multiscale dynamics that are ubiquitous around the world: tidal-to-mesoscale dynamics at shelfbreaks, including interactions with shallow seas, complex coastal geometries, and deep oceans.

To our knowledge, none of the above structured models includes fully implicit two-way embedding schemes for nonlinear free-surface PEs. With fully implicit and two-way embedding, all of the updated field values are exchanged across scales among nested domains, as soon as they become available, within the same time step. This is challenging but found most valuable with nonlinear free-surface PEs. Major contributions of this manuscript are to derive a class of such embedding schemes, implicit in space and time, to compare them to alternatives using simulations and theoretical truncation error analysis, and to illustrate them in a set of realistic applications. Another contribution is a time-dependent spatial discretization of the nonlinear free-surface PEs, including generalized vertical coordinates. These computational algorithms are derived and developed next. Specific new developments include a nonlinear formulation of the free surface and its boundary conditions, a modification of an implicit time-stepping algorithm (Dukowicz and Smith 1994) to handle the nonlinear formulation, a consistent spatial discretization for a time-dependent finite volume method, a generalized vertical grid, and a fully implicit two-way nesting scheme for the nonlinear free-surface PE. Implicit two-way nesting schemes are shown to have truncation errors of higher order than other nesting schemes across the multiresolution domains. Two-way nesting also enables us to easily use different parametrizations for the sub-gridscale physics in each nested domain. The additions of these improvements are shown to drastically enhance the predictive capability and robustness of our ocean prediction system. Without them, our multiscale multiprocess simulations were either not possible or their predictions did not match ocean data.

All of the above new computational schemes have been derived and implemented as part of our MIT Multidisciplinary Simulation, Estimation and Assimilation System (MSEAS; MSEAS Group 2010). This allowed us to evaluate robustness in several ocean regions, including the middle Atlantic Bight, Californian coast around Monterey Bay, Philippine archipelago, and Taiwan–Kuroshio region of the eastern Pacific (e.g., see Section 5). These applications utilized various components of MSEAS including our free-surface generalization of the original rigid-lid PE model of the HOPS (see "Appendix 3" and Haley et al. 1999); a coastal objective analysis scheme based on fast-marching methods (Agarwal and Lermusiaux 2010); uncertainty estimation, data assimilation, and adaptive sampling schemes (Lermusiaux 1999, 2002, 2007; Lermusiaux et al. 2000, 2002); a stochastic representation for sub-gridscale processes (Lermusiaux

2006); nested tidal inversion schemes (Logutov and Lermusiaux 2008); multiple biological models (Tian et al. 2004); and several acoustic models (Lam et al. 2009; Lermusiaux and Xu 2010).

A recent and comprehensive review of nesting algorithms can be found in Debreu and Blayo (2008), including discussions on time-stepping and time-splitting issues. They review methods for the conservation of quantities across the nesting interface and compare a variety of schemes for the transfer of information between grids. They conclude with a review of methods to control noise, including relaxation methods, sponge layers and open boundary conditions suitable for nesting. One-way nesting and two-way nesting with PE models are relatively common (e.g., Spall and Holland 1991; Fox and Maskell 1995; Sloan 1996; Penven et al. 2006; Haley et al. 2009; Mason et al. 2010), and we refer to Debreu and Blayo (2008) for a review. Focusing on scheme comparisons, Cailleau et al. (2008) contrasted methods to control the open boundaries of a modeling domain (Bay of Biscay embedded in a North Atlantic domain), specifically one-way nesting, two-way nesting, and "full coupling based on domain decomposition" (Schwarz method: Martin 2003, 2004). They found that this "full coupling" gave the most regular solutions at interfaces but was computationally much more expensive (a factor of 5) than two-way nesting, without demonstrating significant improvements. Other recent examples include Barth et al. (2005) who use nesting and the free-surface GHER model (Beckers 1991; Beckers et al. 1997) to obtain high-resolution simulations in the Ligurian Sea nested in Mediterranean domains. A new feature of their nesting algorithm is their interpolation of normal velocities from the coarse-to-fine domains. They employ a constrained minimization of the second derivatives to obtain smoothly continuous boundary fields while maintaining the conservation of volume. In Barth et al. (2007), this same setup is coupled with an ensemble-based data assimilation algorithm to assimilate sea surface temperature (SST) and sea surface height (SSH). Estournel et al. (2009) applied "scale-oriented" one-way multi-model nesting to the northwestern Mediterranean Sea (MFSTEP: Pinardi et al. 2002), using a variational scheme to ensure mass balance. Guo et al. (2003) used one-way nesting and the POM for their studies of the Kuroshio, using three telescoping domains. They found that higher resolution not only improved bathymetry reproduction but also JEBAR (joint effect of baroclinicity and relief: Sarkisya and Ivanov 1971) and the Kuroshio dynamics. Other developments include attempts at using improved physics in the refined nested domain. Shen and Evans (2004) developed such a modeling system based on a semi-Lagrangian scheme: A fully nonhydrostatic simulation can be embedded in a larger weakly nonhydrostatic simulation which, in turn, can be embedded in a still larger compatible hydrostatic simulation. Maderich et al. (2008) developed a system to model the transport and mixing of industrial cooling water in freshwater and marine environments, combining free-surface hydrostatic physics with a buoyant jet model or a nonhydrostatic model, using buffer zones to reduce noise due to physics mismatches.

The nesting schemes in all above works fall under the categories we define as "explicit" or "coarse-to-fine implicit" nesting. As shown in Fig. 1, in explicit two-way nesting, the coarse and fine domain fields are only exchanged at the start of a discrete time integration or time step: The two-way exchanges are explicit. In "coarse-to-fine implicit" two-way nesting, the coarse domain feeds the fine domain during its time step: Usually, fine domain boundary values are computed from the coarse domain integration, but the fine domain interior values are only fed back at the end of the coarse time step. In "fine-to-coarse implicit" two-way nesting, it is the opposite; fine domain updates are fed to the coarse domain during its integration but the coarse domain feedback only occur at the end of the fine domain discrete integration. In this paper, we derive two-way nested schemes, fully implicit in space and time: The fine and large domains exchange all updated information during their time integration, as soon as updated fields become available. A type of such scheme consists of computing fine domain boundary values from the coarse domain but with feedback from the fine domain. Some of the algorithmic details of our multiscale fully implicit two-way nesting schemes are specific to MSEAS, but the approach and schemes are general and applicable to other modeling systems.

In what follows, in Section 2, we give the equations of motion, provide an implicit time discretization for the nonlinear free-surface PEs, and develop a time-dependent, spatial discretization of the PEs. In Section 3, we derive and describe our fully implicit two-way nesting scheme and contrast it from traditional explicit and coarse-to-fine implicit schemes. In Section 4, we compare nesting schemes and show that for nesting with free surfaces, the most accurate schemes are those with stronger implicit couplings among grids, especially for the velocity components. We also complete a theoretical truncation error analysis to mathematically confirm and explain our findings. In Section 5, we illustrate the use of our novel discretization and nesting schemes in the middle Atlantic Bight, Philippine archipelago, and Taiwan–

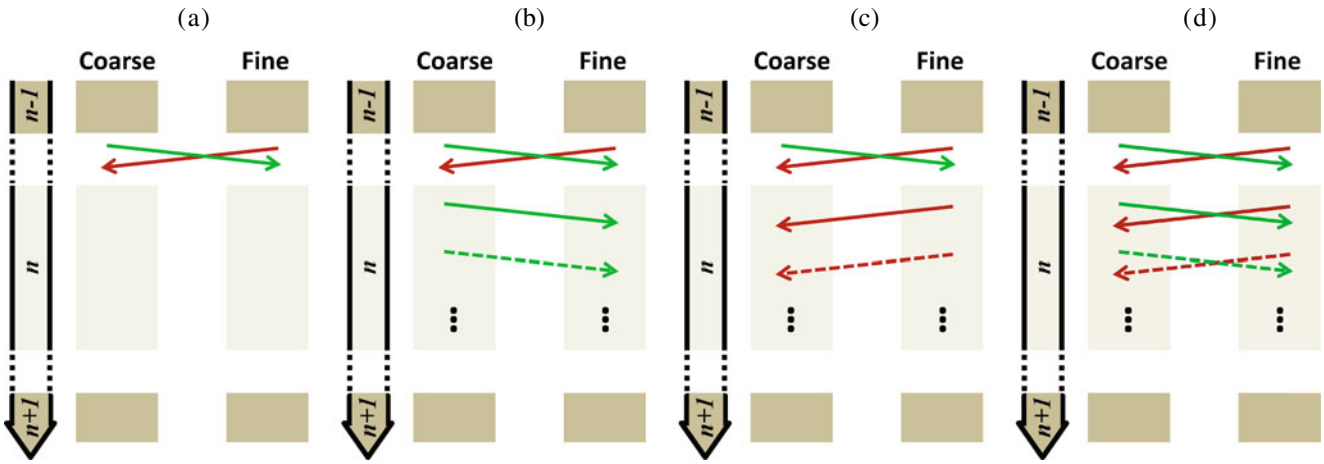

Fig. 1 Schematic of **a** explicit, **b** coarse-to-fine implicit, **c** fine-to-coarse implicit, and **d** fully implicit two-way nesting. *Green arrows* sketch coarse-to-fine transfers; *red arrows* sketch fine to coarse. The *left arrow* indicates discrete time integrations or time steps ($n-1$, n, and $n+1$). Nesting transfers occur before (explicit) or during (implicit) discrete time step n. If the time steps of two nested models are not equal, the duration of step n would in general be the longest of the two

Kuroshio region of the eastern Pacific. Conclusions are in Section 6. Details on vertical and horizontal discretizations and fluxes, open boundary conditions, and conservation properties are in "Appendix 1". Multiscale nesting procedures for setting up multigrid domains and bathymetries, for multiresolution initialization, for tidal forcing, and for solving the free-surface equation are given in "Appendix 2". Our original two-way nesting scheme for rigid-lid PEs is outlined in "Appendix 3".

2 Formulation of a new scheme for free-surface primitive equation modeling

In this section, we derive the discretized equations of motion for our new nested nonlinear free-surface ocean system. We have encoded both the spherical and Cartesian formulations (see "Appendix 1") and most often use the spherical one, but for ease of notation, we present the equations in only one form, the Cartesian one. In Section 2.1, we give the differential form of the free-surface PEs. In Section 2.2, we recast these equations in their integral control volume form in order to easily derive a mass preserving scheme. In Section 2.3, we introduce our novel implicit time discretization of these PEs. Finally, in Section 2.4, we derive the corresponding time-dependent, spatial discretization which preserves mass and tracers in the presence of a time-varying free surface.

2.1 Continuous free-surface primitive equations

The equations of motion are the PEs, derived from the Navier–Stokes equations under the hydrostatic and Boussinesq approximations (e.g., Cushman-Roisin and Beckers 2010). Under these assumptions, the state variables are the horizontal and vertical components of velocity ($\mathbf{u}, w$), the temperature (T), and the salinity (S). Denoting the spatial positions as (x, y, z) and the temporal coordinate with t, the PEs are:

$$\text{Cons. Mass} \quad \nabla\cdot\mathbf{u}+\frac{\partial w}{\partial z}=0, \tag{1}$$

$$\text{Cons. Horiz. Mom.} \quad \frac{D\mathbf{u}}{Dt}+f\hat{k}\times\mathbf{u}=-\frac{1}{\rho_0}\nabla p+\mathbf{F}, \tag{2}$$

$$\text{Cons. Vert. Mom.} \quad \frac{\partial p}{\partial z}=-\rho g, \tag{3}$$

$$\text{Cons. Heat} \quad \frac{DT}{Dt}=F^T, \tag{4}$$

$$\text{Cons. Salt} \quad \frac{DS}{Dt}=F^S, \tag{5}$$

$$\text{Eq. of State} \quad \rho=\rho(z,T,S) \tag{6}$$

where $\frac{D}{Dt}$ is the 3D material derivative, p is the pressure, f is the Coriolis parameter, ρ is the density, ρ_0 is the (constant) density from a reference state, g is the acceleration due to gravity, and $\hat{k}$ is the unit direction vector in the vertical direction. The gradient operators, ∇, in Eqs. 1 and 2 are 2D (horizontal) operators. The turbulent sub-gridscale processes are represented by $\mathbf{F}$, F^T, and F^S.

Since we are considering free-surface applications in regions with strong tides, we need a prognostic equation for the evolution of the surface elevation, η. We integrate Eq. 1 over the vertical column and apply the kinematic conditions at the surface and bottom to arrive at the nonlinear free-surface transport constraint

$$\frac{\partial \eta}{\partial t} + \nabla \cdot \left(\int_{-H}^{\eta} \mathbf{u}\, dz \right) = 0 \tag{7}$$

where $H = H(x, y)$ is the local water depth in the undisturbed ocean.

We decompose the horizontal velocity into a depth averaged ("barotropic") component, $\mathbf{U}$, and a remainder ("baroclinic"), $\mathbf{u}'$

$$\mathbf{u} = \mathbf{u}' + \mathbf{U} \; ; \quad \mathbf{U} = \frac{1}{H+\eta} \int_{-H}^{\eta} \mathbf{u}\, dz. \tag{8}$$

To further isolate the effects of the free surface, we decompose the pressure into a hydrostatic component (employing the terminology of Dukowicz and Smith 1994), p_{h}, and a surface component, p_{s}:

$$p = p_{\mathrm{s}} + p_{\mathrm{h}} \; ; \; p_{\mathrm{h}}(x, y, z, t) = \int_{z}^{\eta} g\rho \, d\zeta \; ;$$
$$p_{\mathrm{s}}(x, y, t) = \rho_0 g \eta. \tag{9}$$

Note that the definition of the hydrostatic pressure automatically enforces Eq. 3. Using Eqs. 8 and 9, we split Eq. 2 into two equations, one for $\mathbf{U}$ obtained by taking the vertical average of Eq. 2 and one for $\mathbf{u}'$ by removing the vertical average from Eq. 2:

$$\frac{\partial \mathbf{U}}{\partial t} - \frac{\mathbf{u}'|_{\eta}}{H+\eta} \frac{\partial \eta}{\partial t} + f\hat{k} \times \mathbf{U} = \overline{\widehat{\mathcal{F}}} - g \nabla \eta \tag{10}$$

$$\frac{\partial \mathbf{u}'}{\partial t} + \frac{\mathbf{u}'|_{\eta}}{H+\eta} \frac{\partial \eta}{\partial t} + f\hat{k} \times \mathbf{u}' = \widehat{\mathcal{F}} - \overline{\widehat{\mathcal{F}}}. \tag{11}$$

In Eqs. 10 and 11, we now have additional terms of the form $\frac{\mathbf{u}'|_{\eta}}{H+\eta} \frac{\partial \eta}{\partial t}$. These small terms are often neglected but are kept here since our dynamical focus ranges from the deep ocean to the very shallow ocean with strong tides. In Eqs. 10 and 11, we have introduced the following notation for the terms we group on the RHS:

$$\widehat{\mathcal{F}} = -\frac{1}{\rho_0} \nabla p_h - \mathbf{\Gamma}(\mathbf{u}) + \mathbf{F} \; ; \quad \overline{\widehat{\mathcal{F}}} = \frac{1}{H+\eta} \int_{-H}^{\eta} \widehat{\mathcal{F}}\, dz$$

and for the advection operator

$$\mathbf{\Gamma}(\mathbf{u}) = \begin{pmatrix} \Gamma(u) \\ \Gamma(v) \end{pmatrix} \; ; \quad \Gamma(\phi) = u \frac{\partial \phi}{\partial x} + v \frac{\partial \phi}{\partial y} + w \frac{\partial \phi}{\partial z}.$$

Note that instead of directly solving for $\mathbf{u}'$ using Eq. 11, we instead solve for $\mathbf{u}$ using Eq. 2 recast in the following form

$$\frac{\partial \mathbf{u}}{\partial t} + f\hat{k} \times \mathbf{u} = \widehat{\mathcal{F}} - g \nabla \eta, \tag{12}$$

then obtain $\mathbf{u}'$ from definition 8. By using Eqs. 12 and 8 instead of Eq. 11, we reduce the truncation error for our time-splitting procedure in Section 2.3.1.

2.2 Control volume formulation of the free-surface primitive equations

We now rewrite the governing Eqs. 1, 4, 5, and 12 in a conservative integral formulation. With this transformation at the continuous level, it is easier to derive a new discrete system that correctly accounts for the temporal changes in the ocean volume due to a moving free surface.

We integrate Eq. 1 and the conservative forms of Eqs. 4, 5, and 12 over a control volume $\mathcal{V}$ and use the divergence theorem to arrive at the following system of equations:

$$\int_{\mathcal{S}} (\mathbf{u}, w) \cdot d\boldsymbol{\mathcal{A}} = 0, \tag{13}$$

$$\frac{\partial}{\partial t} \left(\int_{\mathcal{V}} \mathbf{u}\, d\mathcal{V} \right) + \int_{\mathcal{V}} f\hat{k} \times \mathbf{u}\, d\mathcal{V} = \widetilde{\mathcal{F}} - g \nabla \eta, \tag{14}$$

$$\mathbf{u}' = \mathbf{u} - \frac{1}{H+\eta} \int_{-H}^{\eta} \mathbf{u}\, dz, \tag{15}$$

$$\frac{\partial \mathbf{U}}{\partial t} - \frac{\mathbf{u}'|_{\eta}}{H+\eta} \frac{\partial \eta}{\partial t} + f\hat{k} \times \mathbf{U} = \overline{\widetilde{\mathcal{F}}} - g \nabla \eta, \tag{16}$$

$$\frac{\partial}{\partial t} \left(\int_{\mathcal{V}} T\, d\mathcal{V} \right) + \tilde{\Gamma}(T) = \int_{\mathcal{V}} F^T\, d\mathcal{V}, \tag{17}$$

$$\frac{\partial}{\partial t} \left(\int_{\mathcal{V}} S\, d\mathcal{V} \right) + \tilde{\Gamma}(S) = \int_{\mathcal{V}} F^S\, d\mathcal{V}, \tag{18}$$

$$\rho = \rho(z, T, S), \tag{19}$$

$$\frac{\partial \eta}{\partial t} + \nabla \cdot [(H+\eta)\, \mathbf{U}] = 0 \tag{20}$$

where

$$\widetilde{\mathcal{F}} = -\frac{1}{\rho_0}\int_{\mathcal{S}} p_h\,\hat{n}_h\cdot d\boldsymbol{\mathcal{A}} - \tilde{\boldsymbol{\Gamma}}(\mathbf{u}) + \int_{\mathcal{V}} \mathbf{F}\,d\mathcal{V},$$

$$\overline{\widetilde{\mathcal{F}}} = \frac{1}{H+\eta}\int_{-H}^{\eta} \widetilde{\mathcal{F}}\,dz,$$

$\mathcal{S}$ is the surface of the control volume, and $d\boldsymbol{\mathcal{A}}$ is an infinitesimal area element vector pointing in the outward normal direction to $\mathcal{S}$. In Eqs. 14–18, we have introduced the following notation for the surface advective fluxes:

$$\tilde{\boldsymbol{\Gamma}}(\mathbf{u}) = \begin{pmatrix} \tilde{\Gamma}(u) \\ \tilde{\Gamma}(v) \end{pmatrix} \;;\quad \tilde{\Gamma}(\phi) = \int_{\mathcal{S}} \phi\ (\mathbf{u}, w)\times d\boldsymbol{\mathcal{A}}$$

where $\phi\ (\mathbf{u}, w)$ denotes the local advective flux of ϕ.

2.3 Temporal discretization

We now derive our novel implicit time discretization for the nonlinear free-surface PEs (Eqs. 13–20). Using the following discrete time notation:

$$t_n = n\,\Delta t \;;\quad \phi\,(t_n) = \phi^n$$

where Δt is the discrete time step, and using the second-order leap-frog time differencing operator:

$$\delta(\phi) = \phi^{n+1} - \phi^{n-1},$$

we obtain the following temporal discretization of Eqs. 13–20

$$\int_{\mathcal{S}^n} (\mathbf{u}^n, w^n)\cdot d\boldsymbol{\mathcal{A}} = 0, \tag{21}$$

$$\frac{1}{\tau}\delta\left(\int_{\mathcal{V}} \mathbf{u}\,d\mathcal{V}\right) + \left(\int_{\mathcal{V}} f\hat{k}\times\mathbf{u}\,d\mathcal{V}\right)^{\alpha} = \widetilde{\mathcal{F}}^{n,n-1} - \left(\int_{\mathcal{V}} g\nabla\eta\,d\mathcal{V}\right)^{\alpha}, \tag{22}$$

$$\mathbf{u}'^{n+1} = \mathbf{u}^{n+1} - \frac{1}{H+\eta^{n+1}}\int_{-H}^{\eta^{n+1}} \mathbf{u}^{n+1}\,dz, \tag{23}$$

$$\frac{\delta(\mathbf{U})}{\tau} - \frac{\mathbf{u}'^n\big|_\eta}{H+\eta^n}\frac{\delta\eta}{\tau} + f\hat{k}\times\mathbf{U}^{\alpha} = \overline{\widetilde{\mathcal{F}}^{n,n-1}} - g\nabla\eta^{\alpha}, \tag{24}$$

$$\frac{1}{\tau}\delta\left(\int_{\mathcal{V}} T\,d\mathcal{V}\right) = \int_{\mathcal{V}^n} F^{T^n}\,d\mathcal{V} - \tilde{\Gamma}(T^n), \tag{25}$$

$$\frac{1}{\tau}\delta\left(\int_{\mathcal{V}} S\,d\mathcal{V}\right) = \int_{\mathcal{V}^n} F^{S^n}\,d\mathcal{V} - \tilde{\Gamma}(S^n), \tag{26}$$

$$\frac{\eta^{n+1}-\eta^n}{\Delta t} + \nabla\cdot\left[(H+\eta^n)\,\mathbf{U}^{\theta}\right] = 0 \tag{27}$$

where

$$\widetilde{\mathcal{F}}^{n,n-1} = -\frac{1}{\rho_0}\int_{\mathcal{S}^n} p_h^n\,\hat{n}_h\cdot d\boldsymbol{\mathcal{A}} - \tilde{\boldsymbol{\Gamma}}(\mathbf{u}^n) + \int_{\mathcal{V}^n}\mathbf{F}^n\,d\mathcal{V} + \int_{\mathcal{V}^{n-1}}\mathbf{F}^{n-1}\,d\mathcal{V},$$

$$\overline{\widetilde{\mathcal{F}}^{n,n-1}} = \frac{1}{H+\eta^n}\int_{-H}^{\eta^n}\left\{-\frac{1}{\rho_0}\int_{\mathcal{S}^n} p_h^n\,\hat{n}_h\cdot d\boldsymbol{\mathcal{A}} - \tilde{\boldsymbol{\Gamma}}(\mathbf{u}^n) + \int_{\mathcal{V}^n}\mathbf{F}^n\,d\mathcal{V}\right\}dz + \frac{1}{H+\eta^{n-1}}\int_{-H}^{\eta^{n-1}}\left\{\int_{\mathcal{V}^{n-1}}\mathbf{F}^{n-1}\,d\mathcal{V}\right\}dz,$$

and $\tau = 2\Delta t$ is twice the time step. Following the results of the stability analyses in Dukowicz and Smith (1994), we have introduced semi-implicit time discretizations for the Coriolis force

$$\phi^{\alpha} = \alpha\phi^{n+1} + (1-2\alpha)\phi^n + \alpha\phi^{n-1}$$

and for the barotropic continuity:

$$\phi^{\theta} = \theta\phi^{n+1} + (1-\theta)\phi^n.$$

In practice, we run using the stabilizing choices $\alpha = \frac{1}{3}$ (C. Lozano and L. Lanerolle, private communication) and $\theta = 1$ (Dukowicz and Smith 1994). A stability analysis of the explicit leap-frog algorithm can be found in Shchepetkin and McWilliams (2005), while Dukowicz and Smith (1994) analyze the linearized implicit algorithm. Note that even though our discretization parallels Dukowicz and Smith (1994), we do not make the linearizing assumption $\eta \ll H$ in Eqs. 8, 9, and 27. This generalization allows our system to be deployed in littoral regions of high topographic variations and strong tides.

A couple of observations are worth making. First, we are considering the case in which the control volume is time dependent. Therefore, in the new time discretizations (Eqs. 21–26), all terms involving control volume integrals must be evaluated at the proper discrete times as a whole, not just the integrands. The second is that Eqs. 22–24 and 27 form a coupled system of equations to solve for $\mathbf{u}^{n+1}$, $\mathbf{u}'^{n+1}$, $\mathbf{U}^{n+1}$, and η^{n+1}. We decouple these equations using a time-splitting algorithm. Another approach would have been to use an iterative method (e.g., Newton solver). However, time-splitting is usually more efficient and for their similar

time-splitting approach, Dukowicz and Smith (1994) showed that no significant physics was lost, provided $f\Delta t \leq 2$. Our time steps are always much smaller than that limit.

2.3.1 Time-splitting procedure

Similar to Dukowicz and Smith (1994), we employ a time-splitting approach by first introducing the splitting variables, $\left(\widehat{\int_{\mathcal{V}} \mathbf{u}\, d\mathcal{V}}\right)^{n+1}$ and $\hat{\mathbf{U}}^{n+1}$:

$$\left(\widehat{\int_{\mathcal{V}} \mathbf{u}\, d\mathcal{V}}\right)^{n+1} \equiv \left(\int_{\mathcal{V}} \mathbf{u}\, d\mathcal{V}\right)^{n+1} + \alpha\tau\delta\left(\int_{\mathcal{V}} g\nabla\eta\, d\mathcal{V}\right), \tag{28}$$

$$\hat{\mathbf{U}}^{n+1} \equiv \mathbf{U}^{n+1} + \alpha\tau g\nabla\delta\eta - \frac{\mathbf{u}'^{n}\big|_{\eta}}{H+\eta^{n}}\delta\eta. \tag{29}$$

The novel portions of this, needed to deal with the full nonlinear free-surface dynamics, are the introduction of Eq. 28 and the last term in Eq. 29. Substituting Eqs. 28 and 29 into Eqs. 22 and 24, we obtain

$$\begin{aligned}
&\hat{\delta}\left(\int_{\mathcal{V}} \mathbf{u}\, d\mathcal{V}\right) + \alpha\tau\hat{\delta}\left(\int_{\mathcal{V}} f\hat{k}\times\mathbf{u}\, d\mathcal{V}\right)\\
&\quad = \tau\widetilde{\mathcal{F}}^{n,n-1} - \tau\left(\int_{\mathcal{V}} g\nabla\eta\, d\mathcal{V}\right)^{\tilde{\alpha}} - \tau\left(\int_{\mathcal{V}} f\hat{k}\times\mathbf{u}\, d\mathcal{V}\right)^{\tilde{\alpha}}\\
&\qquad + \alpha^2\tau^2\delta\left(\int_{\mathcal{V}} f\hat{k}\times\nabla\eta\, d\mathcal{V}\right),
\end{aligned} \tag{30}$$

$$\begin{aligned}
\hat{\delta}\mathbf{U} + \alpha f\tau\hat{k}\times\hat{\delta}\mathbf{U} = \tau\left\{\mathcal{F}^{n,n-1} - g\nabla\eta^{\tilde{\alpha}}\right\}\\
+\alpha^2 g f\tau^2\hat{k}\times\nabla\delta\eta + \alpha f\tau\delta\eta\hat{k}\\
\times\frac{\mathbf{u}'^{n}\big|_{\eta}}{H+\eta^{n}},
\end{aligned} \tag{31}$$

where we have introduced the following notation

$$\phi^{\tilde{\alpha}} = (1-2\alpha)\phi^{n} + 2\alpha\phi^{n-1},$$

$$\hat{\delta}\mathbf{U} = \hat{\mathbf{U}}^{n+1} - \mathbf{U}^{n-1} \quad ;$$

$$\hat{\delta}\left(\int_{\mathcal{V}} \mathbf{u}\, d\mathcal{V}\right) = \left(\widehat{\int_{\mathcal{V}} \mathbf{u}\, d\mathcal{V}}\right)^{n+1} - \left(\int_{\mathcal{V}} \mathbf{u}\, d\mathcal{V}\right)^{n-1},$$

$$\begin{aligned}
\mathcal{F}^{n,n-1} = \frac{1}{H+\eta^{n}}\int_{-H}^{\eta^{n}}\left\{-\frac{1}{\rho_0}\int_{\mathcal{S}^{n}} p_h^{n}\hat{n}_h\cdot d\mathcal{A} - \tilde{\mathbf{\Gamma}}(\mathbf{u}^{n})\right.\\
\left. + \int_{\mathcal{V}^{n}} \mathbf{F}^{n}\, d\mathcal{V}\right\} dz\\
+ \frac{1}{H+\eta^{n-1}}\int_{-H}^{\eta^{n-1}}\left\{\int_{\mathcal{V}^{n-1}} \mathbf{F}^{n-1}\, d\mathcal{V}\right\} dz\\
- f\hat{k}\times\mathbf{U}^{\tilde{\alpha}}.
\end{aligned}$$

To decouple Eqs. 30–31, we first notice that the last term in Eq. 30 and the second to last term in Eq. 31 are both $O\left(\tau^2\delta\eta\right)$. These terms are the same order as the second-order truncation errors already made and hence can be discarded. The last term in Eq. 31 is $O\left(\tau\delta\eta\right)$. Although this represents a first-order error term in the free-surface elevation, it is still comparable to the error in the free-surface integration scheme (Eq. 27). Furthermore, the term is divided by $H+\eta$, meaning that it is $O\left(\frac{\tau\delta\eta}{H+\eta}\right)$ which is never larger than $O\left(\frac{\tau\delta\eta}{\eta}\right)$ in a single time step and often much smaller. Hence, we discard this term too. Discarding these terms results in the following decoupled momentum equations

$$\begin{aligned}
&\hat{\delta}\left(\int_{\mathcal{V}} \mathbf{u}\, d\mathcal{V}\right) + \alpha\tau\hat{\delta}\left(\int_{\mathcal{V}} f\hat{k}\times\mathbf{u}\, d\mathcal{V}\right)\\
&\quad = \tau\widetilde{\mathcal{F}}^{n,n-1} - \tau\left(\int_{\mathcal{V}} g\nabla\eta\, d\mathcal{V}\right)^{\tilde{\alpha}} - \tau\left(\int_{\mathcal{V}} f\hat{k}\times\mathbf{u}\, d\mathcal{V}\right)^{\tilde{\alpha}},
\end{aligned} \tag{32}$$

$$\hat{\delta}\mathbf{U} + \alpha f\tau\hat{k}\times\hat{\delta}\mathbf{U} = \tau\left\{\mathcal{F}^{n,n-1} - g\nabla\eta^{\tilde{\alpha}}\right\}, \tag{33}$$

To finish the decoupling, we take Eq. 27, average it with itself evaluated a time step earlier, and substitute Eq. 29 for $\mathbf{U}^{n+1}$. The result is the following decoupled equation for η^{n+1}

$$\begin{aligned}
&\alpha\theta g\tau\nabla\cdot\left[\left(H+\eta^{n}\right)\nabla\delta\eta\right] - \theta\nabla\cdot\left(\mathbf{u}'^{n}\big|_{\eta}\,\delta\eta\right) - \frac{2\delta\eta}{\tau}\\
&\quad = \nabla\cdot\left[\left(H+\eta^{n}\right)\left(\theta\hat{\mathbf{U}}^{n+1} + \mathbf{U}^{n} + (1-\theta)\mathbf{U}^{n-1}\right)\right]
\end{aligned} \tag{34}$$

In conclusion, the new elements of temporal discretization are in Eqs. 28, 29, 32, and 34. In particular, the nonlinear free-surface parametrization is maintained by the $H+\eta^{n}$ factors in the divergences in Eq. 34 and by the second term on the left-hand side of Eq. 34.

Note that it is this decoupling procedure that inspired us to keep the full momentum equation (Eq. 12) instead of the baroclinic equation (Eq. 11; see Section 2.1). Had we worked with the baroclinic momentum equation directly, the barotropic equations (Eqs. 29, 31, and 33) would have been unchanged; however, the

truncation term in going from Eqs. 30 to 32 would have been $\alpha\tau f\delta\left(\int_{\mathcal{V}} \frac{\hat{k}\times \mathbf{u}'^{n}|_{\eta}}{H+\eta^{n}}\eta\, d\mathcal{V}\right)$ instead of the higher-order term we obtained in Eq. 30. Further, the error term in Eq. 30 is more uniform, while the error term that would have been obtained from the baroclinic equations would have grown as the topography shoaled.

2.4 Time-dependent, nonlinear "distributed-σ" spatial discretization of the free-surface primitive equations

Using temporal discretization (Eqs. 21, 23, 25, 26, and 32–34), we can derive our new, time-dependent, spatial discretization. This discretization distributes with depth the temporal volume changes in the water column due to the time-variable free surface. We found that these variations of cell volumes must all be accounted for to avoid potentially large momentum and tracer errors in regions of strong tides and shallow topography.

Following Bryan (1969), we discretize Eqs. 21, 23, 25, 26, and 32–34 on the staggered Arakawa B-grid (Arakawa and Lamb 1977). We retain the B-grid of the PE model of HOPS based on its ability to simulate geostrophy and any potentially marginally resolved fronts and filaments in our multiscale simulations (Webb et al. 1998; Griffies et al. 2000; Wubs et al. 2006). We employ a finite volume discretization in which the average of a variable over the volume is approximated by the value of the variable at the center of the finite volume (see Section 4.7.1). As shown in Fig. 2, the tracers and free surface (T, S, η) are horizontally located at the centers of "tracer cells" while velocities ($\mathbf{u}'$, $\mathbf{U}$, $\hat{\mathbf{U}}$) are located at the centers of "velocity cells" which are offset $\frac{1}{2}$ grid-point to the northeast from the "tracer cells". In the vertical, the 3D tracers and velocities (T, S, $\mathbf{u}'$) are, again, located at the centers of their respective cells, while the vertical velocities are calculated at the tops of the tracer and velocity cells. By choosing this type of discretization, the control volumes of Eqs. 21, 23, 25, 26, and 32–34 become structured-grid finite volumes.

In the vertical, our new time-dependent, terrain-following coordinates are defined as follows: First, the terrain-following depths for the (undisturbed) mean sea level, $z^{\mathrm{MSL}}_{i,j,k}$, are set (see "Appendix 1.1"). We then define the time variable model depths such that the change in cell thickness is proportional to the relative thickness of the original (undisturbed) cell. Hence, along model level k, the depths can be found from

$$z_k(x,y,t) = \eta(x,y,t) + \left(1+\frac{\eta(x,y,t)}{H(x,y)}\right) z^{\mathrm{MSL}}_k(x,y). \tag{35}$$

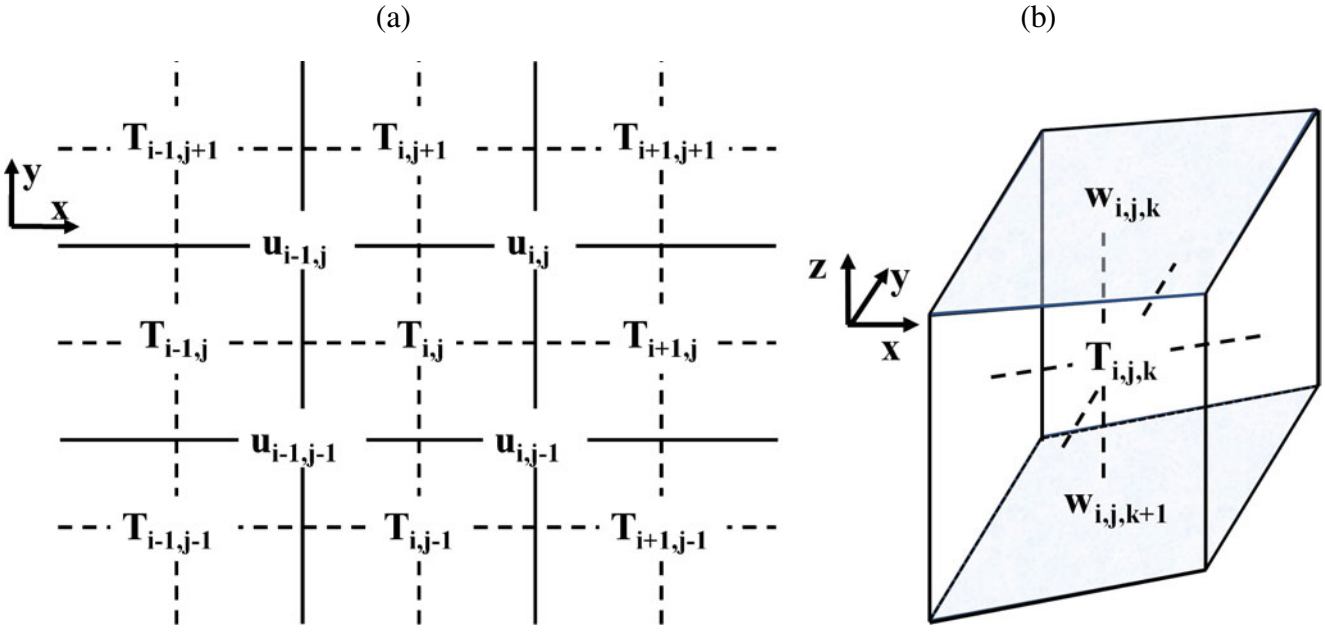

Fig. 2 B-grid indexing scheme. **a** Horizontal lay-out. Here T stands for variables centered in tracer cells (T, S, η) and $\mathbf{u}$ represents variables centered in velocity cells ($\mathbf{u}$, $\mathbf{u}'$, $\mathbf{U}$). **b** Vertical lay-out. Tracer cells are shown, velocity cells have the same lay-out, merely shifted $\frac{1}{2}$ grid-point, and w represents the vertical velocity

By distributing the temporal change in the free surface across all the model levels, we simplify the discretization in shallow regions with large tides (e.g., we avoid requiring that the top level be thick enough to encompass the entire tidal swing, which in the case of very shallow depth can mean most of the total depth). An additional computational benefit is that the time dependence of the computational cell thickness decouples from the vertical index. This provides us the following properties:

$$\frac{1}{H+\eta^n}\sum_{k=1}^{K}\phi^n_{i,j,k}dz^n_{i,j,k}=\frac{1}{H}\sum_{k=1}^{K}\phi^n_{i,j,k}dz^{\mathrm{MSL}}_{i,j,k};$$

$$\frac{\Delta\mathcal{V}^n_{i,j,k}}{\Delta\mathcal{V}^{\mathrm{MSL}}_{i,j,k}}=1+\frac{\eta^n_{i,j}}{H_{i,j}}$$

both of which are used to derive Eq. 39 below.

Since our vertical grid is both terrain following and time variable, we also define a new vertical flux velocity, ω, normal to the top, ζ, of finite volumes as

$$\omega=w-\mathbf{u}\cdot\nabla\zeta-\frac{\partial\zeta}{\partial t}. \tag{36}$$

An important consequence of this definition is that the kinematic conditions at the surface and bottom reduce to

$$\omega|_\eta=0\ ;\quad \omega|_{-H}=0$$

Using these definitions, along with the second-order mid-point approximation

$$\int_{\mathcal{V}}\phi\,d\mathcal{V}=\phi\Delta\mathcal{V}+O\left(\Delta\mathcal{V}^2\right),$$

we discretize Eqs. 21, 23, 25, 26, and 32–34 as

$$\int_{\mathcal{S}^n_{\mathrm{lat}}}\mathbf{u}\cdot d\boldsymbol{\mathcal{A}}+\int_{\mathcal{S}^n_{\mathrm{TB}}}\omega\cdot d\boldsymbol{\mathcal{A}}=0, \tag{37}$$

$$\frac{\hat{\delta}\left(\mathbf{u}\Delta\mathcal{V}\right)}{\tau}+\alpha f\hat{k}\times\hat{\delta}\left(\mathbf{u}\Delta\mathcal{V}\right)=\hat{\mathcal{F}}^{n,n-1}-g\left(\Delta\mathcal{V}\nabla\eta\right)^{\tilde{\alpha}}-f\hat{k}\times\left(\mathbf{u}\Delta\mathcal{V}\right)^{\tilde{\alpha}}, \tag{38}$$

$$\left(\mathbf{u}'\Delta\mathcal{V}\right)^{n+1}=\widehat{\left(\mathbf{u}\Delta\mathcal{V}\right)}^{n+1}-\frac{\Delta\mathcal{V}^{\mathrm{MSL}}}{H}\sum_{k=1}^{K}\frac{\widehat{\left(\mathbf{u}\Delta\mathcal{V}\right)}^{n+1}}{\Delta\mathcal{V}^{\mathrm{MSL}}}dz^{\mathrm{MSL}}, \tag{39}$$

$$\frac{\delta\left(T\Delta\mathcal{V}\right)}{\tau}=F^{T^n}\Delta\mathcal{V}^n-\breve{\Gamma}(T^n), \tag{40}$$

$$\frac{\delta\left(S\Delta\mathcal{V}\right)}{\tau}=F^{S^n}\Delta\mathcal{V}^n-\breve{\Gamma}(S^n), \tag{41}$$

$$\hat{\delta}\mathbf{U}+\alpha f\tau\hat{k}\times\hat{\delta}\mathbf{U}=\tau\left\{\boldsymbol{\mathcal{F}}^{n,n-1}-g\nabla\eta^{\tilde{\alpha}}\right\}, \tag{42}$$

$$\alpha\theta g\tau\nabla\cdot\left[\left(H+\eta^n\right)\nabla\delta\eta\right]-\theta\nabla\cdot\left(\mathbf{u}'^n\big|_\eta\,\delta\eta\right)-\frac{2\delta\eta}{\tau}=\nabla\cdot\left[\left(H+\eta^n\right)\left(\theta\hat{\mathbf{U}}^{n+1}+\mathbf{U}^n+(1-\theta)\mathbf{U}^{n-1}\right)\right], \tag{43}$$

$$\mathbf{U}^{n+1}=\hat{\mathbf{U}}^{n+1}-\alpha\tau g\nabla\delta\eta+\frac{\mathbf{u}'^n\big|_\eta}{H+\eta^n}\delta\eta \tag{44}$$

where

$$\breve{\boldsymbol{\Gamma}}(\mathbf{u})=\begin{pmatrix}\breve{\Gamma}(u)\\ \breve{\Gamma}(v)\end{pmatrix};\quad \breve{\Gamma}(\phi)=\int_{\mathcal{S}^n_{\mathrm{lat}}}\phi\,\mathbf{u}\cdot d\boldsymbol{\mathcal{A}}+\int_{\mathcal{S}^n_{\mathrm{TB}}}\phi\,\omega\cdot d\boldsymbol{\mathcal{A}},$$

$$\hat{\mathcal{F}}^{n,n-1}=-\frac{1}{\rho_0}\int_{\mathcal{S}^n}p^n_h\hat{n}_h\cdot d\boldsymbol{\mathcal{A}}-\breve{\boldsymbol{\Gamma}}(\mathbf{u})^n+\mathbf{F}^n\Delta\mathcal{V}^n+\mathbf{F}^{n-1}\Delta\mathcal{V}^{n-1},$$

$$\overline{\hat{\mathcal{F}}^{n,n-1}}=\frac{1}{H_{i,j}+\eta^n_{i,j}}\int_{-H_{i,j}}^{\eta^n_{i,j}}\left\{-\frac{1}{\rho_0}\int_{\mathcal{S}^n}p^n_h\hat{n}_h\cdot d\boldsymbol{\mathcal{A}}-\breve{\boldsymbol{\Gamma}}(\mathbf{u})^n+\mathbf{F}^n\Delta\mathcal{V}^n\right\}dz+\frac{1}{H_{i,j}+\eta^{n-1}_{i,j}}\int_{-H_{i,j}}^{\eta^{n-1}_{i,j}}\left\{\mathbf{F}^{n-1}\Delta\mathcal{V}^{n-1}\right\}dz,$$

$$\boldsymbol{\mathcal{F}}^{n,n-1}=\overline{\hat{\mathcal{F}}^{n,n-1}}-f\hat{k}\times\mathbf{U}^{\tilde{\alpha}},$$

$\mathcal{S}^n_{\mathrm{lat}}$ are the lateral surfaces of a computational cell and $\mathcal{S}^n_{\mathrm{TB}}$ represents the top and bottom surfaces of the computational cell.

With our new choice of vertical discretization, all cell volumes are functions of time. In regions with relatively high tides (compared to the total water depth), not correctly accounting for the time dependence of the volume change can lead to large errors in the tracer and momentum fields. Focusing on the computational aspects, this time dependency of the cell volume means that we solve the tracer and baroclinic velocity fields in two steps. Using temperature as an example, we first solve for $(T\Delta\mathcal{V})^{n+1}$. Then, after we have solved for η^{n+1}, we update the cell volume and compute T^{n+1}. A second computational property is that we do not maintain separate storage for $\widehat{\left(\mathbf{u}\Delta\mathcal{V}\right)}^{n+1}$ and $\left(\mathbf{u}'\Delta\mathcal{V}\right)^{n+1}$. Instead, immediately after solving Eq. 38, we remove the vertical mean according to Eq. 39. All details of the discretization of the fluxes through the

boundaries of the computational volumes are given in "Appendix 1.2". The resultant system of discrete equations is given by Eqs. 38–44 and 64–65.

3 Fully implicit nesting scheme

In this section, we derive and discuss our new fully implicit (in space and time) two-way nesting scheme. Deriving this scheme required a detailed exploration of the choices of variables to exchange and the specific algorithms, as discussed in Section 4.

Considering first traditional "explicit" and "coarse-to-fine implicit" two-way nesting (Debreu and Blayo 2008), fields are often interpolated from a coarser resolution domain to provide boundary values for a finer resolution domain. Then fields from the finer domain are averaged to replace corresponding values in the coarser domain. This is a natural order of operations in the sense that often a refined (smaller) time step is used for the finer domain, and hence, not all refined time steps have corresponding coarse field values. However, once updated, the coarse domain fields are no longer the same fields that were interpolated for the finer domain boundaries. This results in a weakened coupling (Section 4.6) between the domains which can be rectified either with an iteration scheme or with fully implicit nesting.

In our new implicit nesting, the goal is to exchange all of the updated fields values as soon as they become available. This is analogous to an implicit time-stepping algorithm, which simultaneously solves for all unknowns. It is only analogous because here updated values are exchanged across multiple scales and nested grids within the same time step, for several fields. Hence, we refer to such schemes as being implicit in space and time; the nested solutions are intertwined. Such tightly coupled implicit nesting can, in some sense, be seen as refining grids in a single domain (e.g., Ginis et al. 1998). However, there are some advantages to the nesting paradigm. First, the time stepping can be easily refined for the finer domains. Second, the model dynamics can be tuned for the different scales in the different domains. Most notably, different sub-gridscale physics can easily be employed in the different domains, and we have used this in several regions. Finally, fundamentally different dynamics can be employed in the different domains (e.g., Shen and Evans 2004; Maderich et al. 2008). To implement our implicit nesting, we observe that most of our prognos-

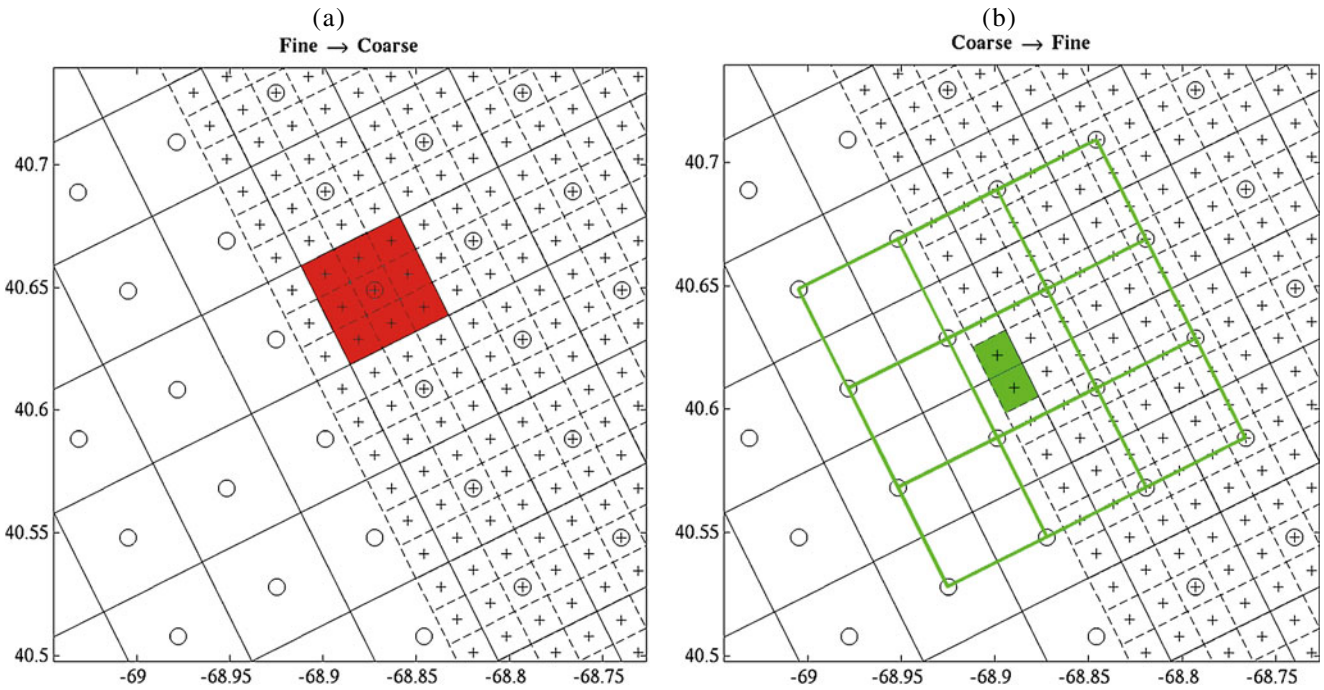

Fig. 3 The basic collocated nesting finite volume domains are shown (for a 3:1 example) with the coarse domain nodal points indicated by *open circles* and the boundaries of the corresponding coarse domain computational cells in *solid lines*. The fine domain nodal points are marked with *plus signs* and the boundaries of the corresponding fine domain computational cells in *dashed lines*. **a** The $r \times r$ array of fine grid cells averaged to update a single coarse grid cell are *highlighted*. **b** The 4×4 stencil of coarse grid nodes bi-cubically interpolated to update boundary nodes of the fine domain are *highlighted* as are the updated fine grid cells

tic variables in our free-surface PE model (Eqs. 37–44) are coded with explicit time stepping. Therefore, reversing the order of operations (updating the coarse domain fields with averages from the interior of the fine domain before interpolating to the boundaries of the fine domain) ensures that, for these fields, the updated field values are in place as soon as they are needed. For the remaining variables, such implicit nesting is more complex. The free-surface η has implicit time stepping (Eq. 43), while $\mathbf{U}$ is coupled to η through (Eq. 44) and boundary conditions ("Appendix 1.3"). Furthermore, additional constraints are imposed on η and $\mathbf{U}$ to maintain the vertically integrated conservation of mass ("Appendix 1.4"). Much of the research summarized in Section 4 was centered around these two variables. The final results are presented next, assuming a two-domain configuration (coarse and fine).

We start by defining collocated grids for the coarse and fine domains as shown in Fig. 3. Our nesting algorithm is suitable for arbitrary odd refinement ratios (r:1), subject to the known issues of scale matching (e.g., Spall and Holland 1991). In this paper, we illustrate the nesting with 3:1 examples. We denote fields evaluated at coarse grid nodes with the indices (i_c, j_c) and fields evaluated at fine grid nodes with (i_f, j_f). We distinguish two special subsets of fine grid nodes: (a) fine grid nodes collocated with coarse grid nodes (i_{fc}, j_{fc}) and (b) fine grid nodes at the outer boundary of the fine domain (i_{fb}, j_{fb}). In this presentation, we assume that we have the same number of model levels and distribution of vertical levels in both domains (i.e., no vertical refinement). However, the topography can be refined in the finer domains (it is refined in all of our examples), subject to the constraints described in "Appendix 2.1.1". The algorithms apply to (and are coded for) both Cartesian and spherical coordinates.

At each time step, our nesting algorithm proceeds as follows (also shown graphically in Fig. 4):

1. Solve Eqs. 37–42 simultaneously in each domain for $(\mathbf{u}'^{n+1}\Delta z^{n+1}, \hat{\mathbf{U}}^{n+1}, T^{n+1}\Delta z^{n+1}, S^{n+1}\Delta z^{n+1})$
2. Replace $(\mathbf{u}'^{n+1}\Delta z^{n+1}, (H+\eta^n)\hat{\mathbf{U}}^{n+1}, T^{n+1}\Delta z^{n+1}, S^{n+1}\Delta z^{n+1}, \eta^n)$ in the coarse domain at overlap nodes with the following averages from the fine domain $(\mathbf{u}'^{n+1}\Delta z^{n+1}, (H+\eta^n)\hat{\mathbf{U}}^{n+1}, T^{n+1}\Delta z^{n+1}, S^{n+1}\Delta z^{n+1}, \eta^n)$

$$\phi^{n+1}_{i_c,j_c,k}\Delta z^{n+1}_{i_c,j_c,k} = \frac{1}{\Delta\mathcal{A}_{i_c,j_c}} \sum_{j=j_{fc}-r_h}^{j_{fc}+r_h} \sum_{i=i_{fc}-r_h}^{i_{fc}+r_h} \phi^{n+1}_{i,j,k}\Delta\mathcal{V}^{n+1}_{i,j,k}, \tag{45}$$

$$\eta^n_{i_c,j_c} = \frac{1}{\Delta\mathcal{A}_{i_c,j_c}} \sum_{j=j_{fc}-r_h}^{j_{fc}+r_h} \sum_{i=i_{fc}-r_h}^{i_{fc}+r_h} \eta^n_{i,j}\Delta\mathcal{A}_{i,j}, \tag{46}$$

$$\left(H_{i_c,j_c}+\eta^n_{i_c,j_c}\right)\hat{\mathbf{U}}^{n+1}_{i_c,j_c} = \frac{1}{\Delta\mathcal{A}_{i_c,j_c}} \sum_{j=j_{fc}-r_h}^{j_{fc}+r_h} \sum_{i=i_{fc}-r_h}^{i_{fc}+r_h} \left(H_{i,j}+\eta^n_{i,j}\right)\hat{\mathbf{U}}^{n+1}_{i,j}\Delta\mathcal{A}_{i,j} \tag{47}$$

where $r_h = \lfloor r/2 \rfloor$ is the greatest integer less than or equal to $r/2$,

$$\phi = \mathbf{u}',\ T,\ S; \quad \Delta\mathcal{V}^n_{i,j,k} = \Delta x_{i,j}\Delta y_{i,j}\Delta z^n_{i,j,k};$$

$$\Delta\mathcal{A}_{i,j} = \Delta x_{i,j}\Delta y_{i,j}.$$

3. In the coarse domain, recompute $\mathbf{U}^n$ from Eq. 44 and updated η^n. When the coarse domain estimate of $\mathbf{U}^n$ was computed from Eq. 44 in the $n-1$ time step, the coarse domain estimate η^n had not yet been updated from the fine domain (Eq. 46 in step 2).
4. In the coarse domain, solve Eqs. 43 and 44 for η^{n+1}, $\mathbf{U}^{n+1}$, Δz^{n+1}, $\mathbf{u}'^{n+1}$, T^{n+1}, S^{n+1}.
5. Using piece-wise bi-cubic Bessel interpolation, $\mathcal{B}$, replace values in the fine grid boundary with values interpolated from the coarse grid

$$\phi^{n+1}_{i_{fb},j_{fb},k} = \mathcal{B}\left(\phi^{n+1}_{i_c,j_c,k}\right), \tag{48}$$

$$\mathbf{u}'^{n+1}_{i_{fb},j_{fb},k}\Delta z^{n+1}_{i_{fb},j_{fb},k} = \mathcal{B}\left(\mathbf{u}'^{n+1}_{i_c,j_c,k}\Delta z^{n+1}_{i_c,j_c,k}\right), \tag{49}$$

$$\mathbf{U}^{n+1}_{i_{fb},j_{fb},k} = \mathcal{B}\left[\left(H_{i_c,j_c}+\eta^{n+1}_{i_c,j_c}\right)\mathbf{U}^{n+1}_{i_c,j_c}\right] \frac{1}{H_{i_{fb},j_{fb}}+\eta^{n+1}_{i_{fb},j_{fb}}} \tag{50}$$

where

$$\phi = T,\ S,\ \eta^n,\ \eta^{n+1}.$$

Note that Eqs. 49 and 50 are written in terms of transports rather than velocities. This is done to generate a consistent mass flux as seen by both domains. We have implemented this scheme to either use the interpolated values in Eqs. 48–50 directly or to correct them to allow the outward radiation of scales unrepresented in the coarse domain. The radiation scheme is an extension of Perkins et al. (1997) and updates our previous radiation schemes (Lermusiaux 2007; Haley et al. 2009). Some more promising, recent boundary conditions that we have derived and that improve the continuity of horizontal fluxes and reduce jumps in vertical fluxes across the fine domain boundaries are presented in "Appendix 1.3".

6. In the fine domain, solve Eqs. 43–44 for η^{n+1}, $\mathbf{U}^{n+1}$, Δz^{n+1}, $\mathbf{u}'^{n+1}$, T^{n+1}, and S^{n+1}.

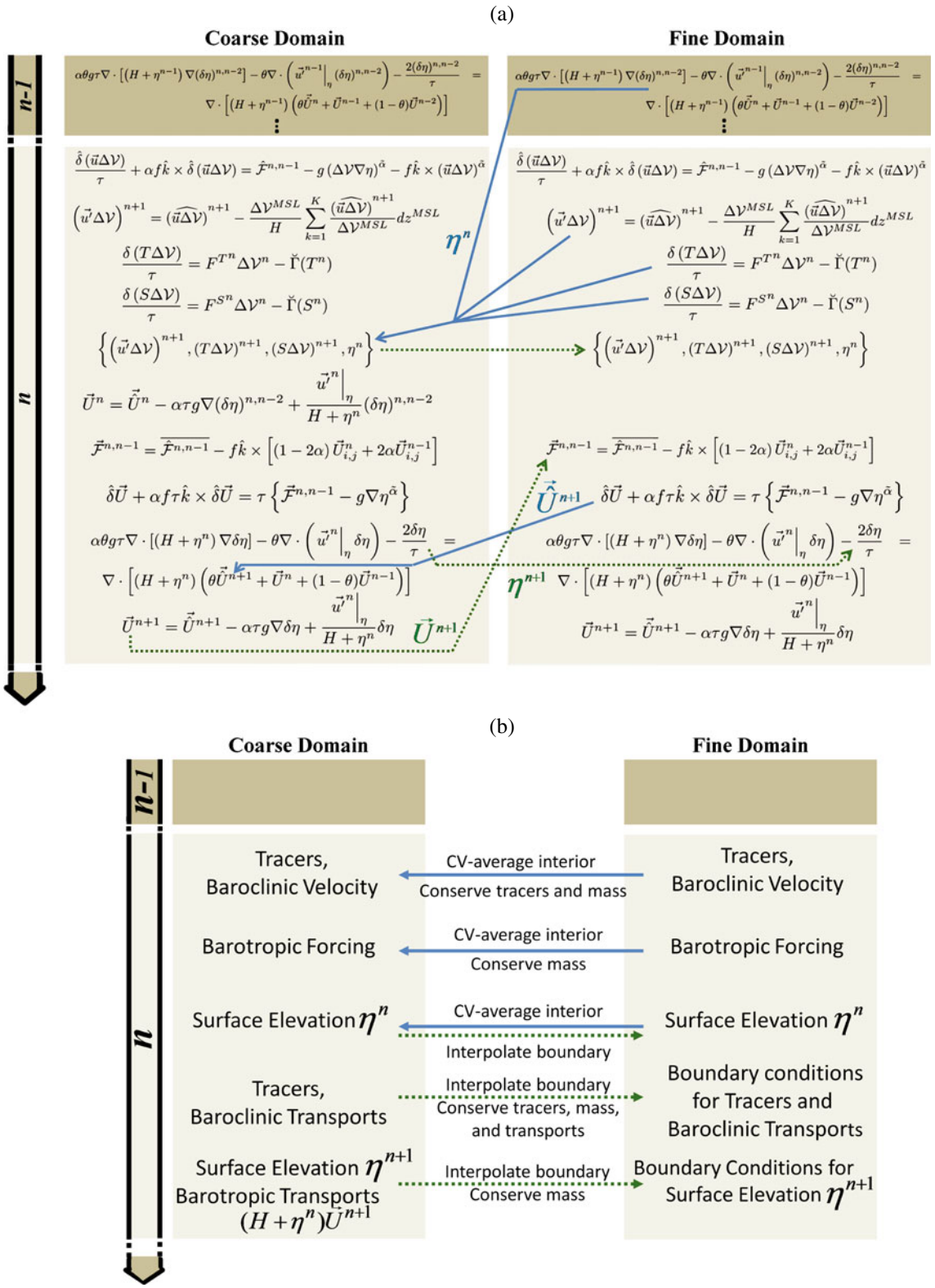

Fig. 4 Present MSEAS-nesting algorithm, two-way implicit in space and time. The nesting algorithm is shown schematically **a** on the discrete structured finite-volume equations (Eqs. 37–44) and **b** in words. *Solid lines* indicate averaging operators from fine domain to coarse. *Dashed lines* indicate interpolation operators from the coarse domain to the boundary of the fine domain

As written in steps 1–6, the new fully implicit nesting scheme requires that both domains be run with the same time step. This is an outgrowth of the applications we have been running, which have strong thermoclines, haloclines, and pycnoclines over shallow areas, steep shelfbreak, and/or open ocean. These applications require a relatively large number of vertical levels (e.g., from 50 to 100 or more). Satisfying the Courant–Friedrichs–Lewy (CFL; Courant et al. 1928) restrictions from the resulting vertical discretizations requires a small enough time step such that the maximum horizontal velocities only reach about 10% of their own CFL limits. Hence, decreasing the horizontal grid spacing by a factor of 3 or 5 does not affect the total CFL limitation much or require a smaller time step.

It is a straightforward problem to restructure this algorithm to handle refined time stepping. First, split the data transfer from the horizontal interpolation in step 5. Before step 2, the values from the coarse grid in the two bands outside of the overlap region (i.e., all the coarse grid points in the interpolation stencil but outside of the overlap region) would be passed to some auxiliary storage in the fine grid model. In the fine grid, these external values would be time interpolated to the current refined time step then spatially interpolated with the averaged fine grid values to the outer boundary. An advantage of our scheme over one with refined time stepping is that the fine grid fields are available to make the update in Eq. 47, which increases the coupling of the barotropic modes between the domains (see Section 4)

Our scheme is directly applicable to an arbitrary number of nonoverlapping, telescoping domains. First, iterate step 2 over all domains from finest to coarsest. Then, apply the series of steps 3–5 for all domains from coarsest to finest.

Finally, since we allow refinement in the topography, our undisturbed vertical terrain-following grid, $z^{\mathrm{MSL}}_{i,j,k}$, requires constraints to maintain consistent interpolation and averaging operations in the above nesting rules. Specifically, in the portion of the coarse domain supported by averages from the fine domain, $z^{\mathrm{MSL}}_{i_c,j_c,k}$ are computed from averages $z^{\mathrm{MSL}}_{i_f,j_f,k}$ following Eq. 46. Along the boundary of the fine domain, $z^{\mathrm{MSL}}_{i_{fb},j_{fb},k}$ are interpolated from $z^{\mathrm{MSL}}_{i_c,j_c,k}$ following Eq. 48. These restrictions, along with the nesting couplings on η, keep the computational cells consistent between domains which, in turn, keeps the averaging operations in Eq. 45 consistent (i.e., as long as the coarse cell is equivalent to the sum of the fine cells then the integral of a field over the coarse cell is conceptually the same as the sum of the integrals of the same field over the corresponding fine cells).

4 Exploring different variations of the fully implicit nesting scheme

We now present and compare a series of two-way nesting schemes that we implemented and tested. Most are simpler versions of the fully implicit nesting scheme (Section 3). All schemes were tested on many common idealized (e.g., a jet meander) and realistic test simulations, for a total of about 1,000 simulations. However, only the results of one of these tests are illustrated next, the same for each scheme. In addition, even though we tried a large number of permutations among all of these schemes, with both small and large variations among them, we only present an organized subset of all schemes tried. Our goal is to illustrate the main canonical schemes. We also limit our comparisons to two-way nesting. Cailleau et al. (2008) compared one-way nesting and two-way nesting. They found significant improvements with two-way nesting, which also reflects our experience. At the end, Section 4.7, we provide an important theoretical analysis of the order of magnitude of the dominant truncation errors of the different schemes. This analysis mathematically explains and contrasts the performance of the various schemes (Section 4.7.2).

The main realistic simulation we selected (see Fig. 5) is based on the real-time AWACS and SW06 exercises (Aug.–Sep. 2006) in the New Jersey Shelf/Hudson Canyon region (WHOI 2006; Lermusiaux et al. 2006; Chapman and Lynch 2010; Lin et al. 2010). It uses Cartesian coordinates. The coarse domain is a 522×447-km domain, with 3-km resolution, to simulate the region of influence. The fine domain is a 172×155-km domain, with 1-km resolution, to refine the simulated dynamics in the main acoustic region just south of the Hudson Canyon. For these nesting tests, both domains employed 30 vertical levels in a double-σ configuration (see "Appendix 1.1"). The bathymetry used was a combination of the NOAA (2006) Coastal Relief Model combined with V8.2 (2000) of the Smith and Sandwell (1997) topography in the deep regions. This combined bathymetry was interpolated and conditioned to coarse 3-km and fine 1-km resolution domains. In the domains overlap, the 1-km bathymetry has sharper scales and is not an interpolation of the 3-km bathymetry (but the 3-km bathymetry is a 3-km control-volume average of the 1-km bathymetry). The estimation of the initial conditions was based on two objective analyses, one inshore and one offshore of the

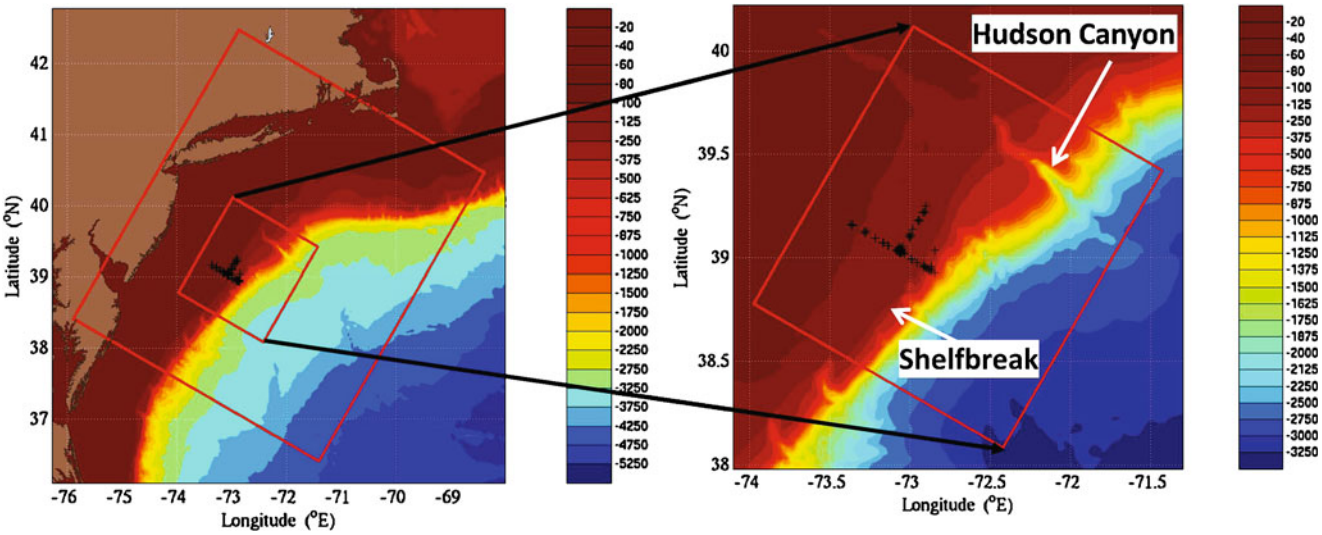

Fig. 5 Nesting domains used for the series of numerical tests we completed in the Shallow Water-06 region. The New Jersey Shelf/Hudson Canyon region of the middle Atlantic Bight is shown along with a pair of domains (3 km, 1 km resolutions) used for two-way nesting

expected shelfbreak front, using both in situ synoptic (gliders, ship deployed conductivity–temperature–depth (CTD), autonomous underwater vehicles) and historical data (National Marine Fisheries Service, World Ocean Database, Gulf Stream Feature analyses, Buoy data, etc.). These two analyses were combined using a shelfbreak front feature model (Sloan 1996; Lermusiaux 1999; Gangopadhyay et al. 2003). The Gulf Stream was initialized based on historical CTD profiles and estimates of its position based on SST and NAVOCEANO feature analyses. The simulations were forced with atmospheric fluxes derived from weather research and forecasting (J. Evans, personal communication) and Fleet Numerical Meteorological and Oceanography Center and laterally forced with linear barotropic tides (Egbert and Erofeeva 2002; Logutov and Lermusiaux 2008). Twice-daily assimilation of the synoptic data is applied to control uncertainties. The nominal duration for this simulation was 43.5 days (two cases with incomplete implicit two-way nesting terminated early due to local CFL violations, Sections 4.2 and 4.3). This duration was chosen by considering the time scales of the dominant processes. For this representative shelfbreak region, they are on the order of 2–7 days. Thus, the simulations are of significant (six to 20 events) duration. Results next are also confirmed by our extensive set of other (not shown) idealized and realistic test simulations.

4.1 Scheme 1: baseline nesting (mimic rigid-lid nesting)

One of the first schemes we tested was a straightforward update of the nesting scheme used for the rigid-lid dynamics (see "Appendix 3"). Comparing this scheme 1 to the consistent implicit scheme of Section 3, Scheme 1 is a five-step scheme; steps 1, 4, and 6 from Section 3 remain unchanged; step 3 is eliminated; and steps 2 and 5 are modified. As a whole, the changes are as follows

- In step 2 (replacing coarse grid values with averages of fine grid values):
  - Eliminate Eq. 46 (averaging surface elevation).
  - In Eq. 47 (averaging barotropic forcing), replace $\left(H_{i,j}+\eta^{n}_{i,j}\right)\hat{\mathbf{U}}^{n+1}_{i,j}$ by $\mathcal{F}^{n,n-1}_{i,j}$ (i.e., move averaging of barotropic forcing one "step" back in PE algorithm).
- Eliminate step 3 (making time-lagged coarse grid barotropic velocity consistent with time-lagged fine grid averaged surface elevation).
- In step 5 (interpolating coarse grid values to boundary of the fine grid; Eq. 48), do not interpolate η^{n}.

The net result of these differences is that there is a much weaker feedback from the fine domain barotropic fields to the coarse domain in this nesting scheme. This "baseline" scheme was first considered because the analogous rigid-lid scheme worked well.

In Fig. 6, we show the results of applying this incompletely implicit nested scheme in the middle Atlantic Bight. In the top row, we present the vector differences between the barotropic velocity computed in the fine domain with the barotropic velocity computed in the corresponding coarse domain simulation, interpolated to the fine domain. These vector differences are overlaid on a map of the magnitude of these

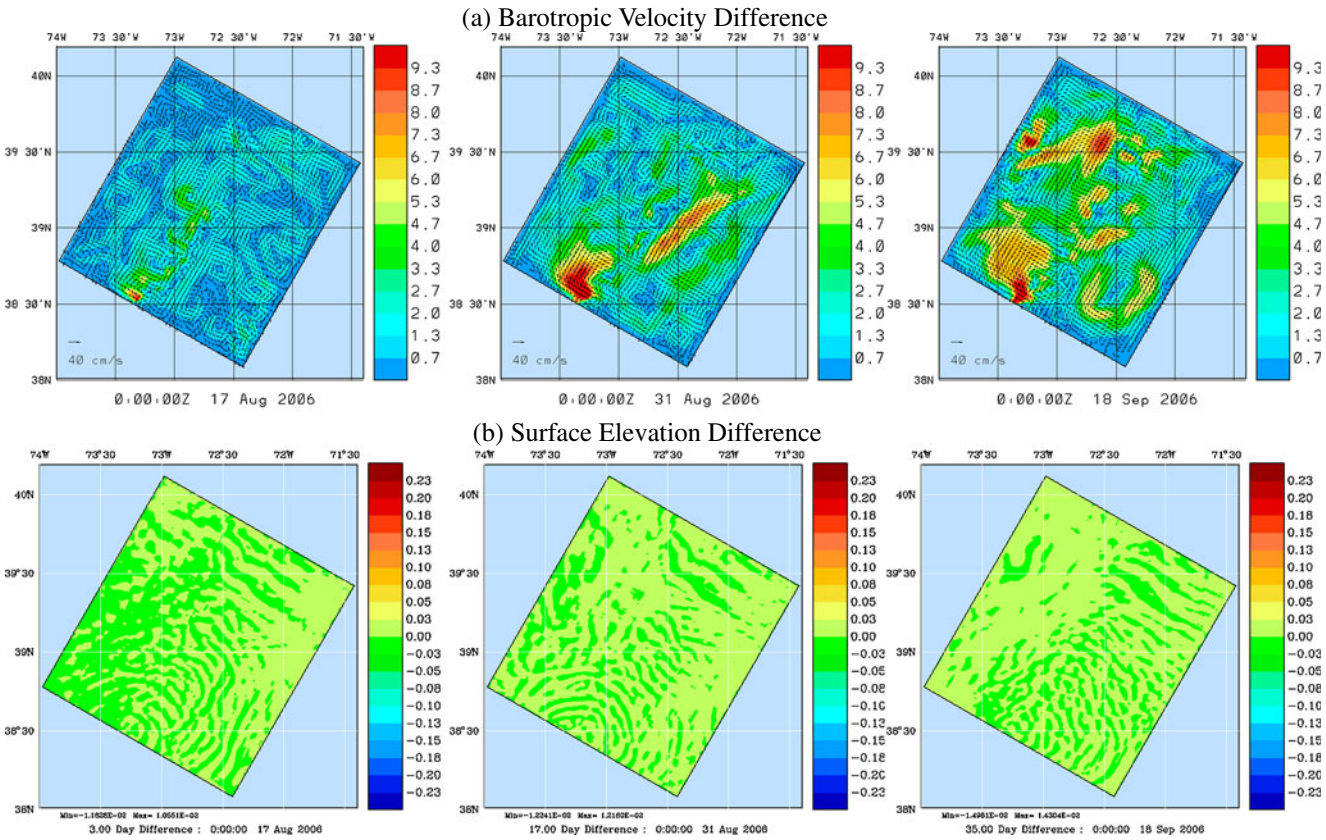

Fig. 6 Scheme 1: baseline nesting. **a** Vector difference between barotropic velocity in coarse and fine domains plotted in the fine domain for 00Z on 17 Aug., 31 Aug., and 18 Sep. (overlain on magnitude of vector difference). **b** Difference between surface elevation in coarse and fine domains plotted in the fine domain for 00Z on 17 Aug., 31 Aug., and 18 Sep. Notice the large (sub)-mesoscale differences in the barotropic velocity

vectors. In the bottom row, we plot the same scalar differences for the surface elevation. Going from left to right, we show these differences at 3 days (after initial adjustment), 17 days (during tropical storm Ernesto), and 35 days (post-Ernesto relaxation) into the simulation. While the coupling of the surface elevation is good, within ± 3 cm everywhere, there is large and growing discrepancy in the barotropic velocity. Not only is the magnitude of the velocity difference large, $O(10$ cm/s), but the velocity differences become similar to (sub)-mesoscale features of the region. These differences are clearly not interpolation error features but represent growing biases between the barotropic velocities estimated on the coarse and fine domains (see Section 4.7.2).

4.2 Scheme 2: average $\hat{\mathbf{U}}$ not $\mathcal{F}$

This scheme improves the barotropic feedback from the fine domain to the coarse of scheme 1 by averaging $\hat{\mathbf{U}}$ instead of $\mathcal{F}$ in Eq. 47. This more strongly couples the barotropic mode by pushing the exchange one step later in the nonlinear free-surface PE algorithm (Eq. 42) and making the feedback closer to the actual barotropic velocity $\mathbf{U}$ (Eq. 44). Comparing this scheme 2 to the consistent implicit scheme of Section 3: Scheme 2 is a five-step scheme; steps 1, 4, and 6 from Section 3 remain unchanged; step 3 is eliminated; and steps 2 and 5 are modified. As a whole, the changes are as follows

- In step 2 (replacing coarse grid values with averages of fine grid values):
  - Eliminate Eq. 46 (averaging surface elevation).
  - In Eq. 47 (averaging barotropic forcing), replace $\left(H_{i,j}+\eta^n_{i,j}\right)\hat{\mathbf{U}}^{n+1}_{i,j}$ by $\hat{\mathbf{U}}^{n+1}_{i,j}$ (i.e., transfer velocity instead of transport).
- Eliminate step 3 (making time-lagged coarse grid barotropic velocity consistent with time-lagged fine grid averaged surface elevation).
- In step 5 (interpolating coarse grid values to boundary of the fine grid; Eq. 48), do not interpolate η^n.

In Fig. 7, we again plot vector differences between the fine and coarse estimates for **U**, η, but using this second scheme. For this particular nesting, a local CFL violation (see below) occurs at 4.5 days into the simulations. We therefore focus on the initial error growth and examined the differences at 0, 0.25, and 0.5 days into a half-day simulation, which is sufficient to illustrate the results. Overall, since we now feedback the barotropic velocity implicitly, the differences between them are much smaller, with no mesoscale organization and amplitudes mainly less than 0.7 cm/s with regions of 0.7–1.3 cm/s along the shelfbreak and in the Hudson canyon and an isolated spot of 10 cm/s at the intersection of the shelfbreak with the southern boundary (where large tidal signal are sensitive to bathymetry resolution). By day 4.5 (not shown), this isolated spot doubles in size and leads to the CFL violation. However, the surface elevation differences are now both large, $O(0.25$ m), and organized on the mesoscale. By day 4.5 (not shown), these differences grow to ± 1 m. By only strengthening the coupling between the **U** estimates, we have simply pushed the interdomain growing bias to η (see Section 4.7.2).

4.3 Scheme 3: exchange η^n

This schemes learns from the advantages of each of the schemes 1 and 2. It further increases the coupling of scheme 2 by also exchanging the surface elevation at a lagged time step. The exchange is both in the averaging from the fine domain to the coarse domain as well as in the interpolation from the coarse domain to the fine domain. Comparing this scheme 3 to the consistent implicit scheme of Section 3: Scheme 3 is a five-step scheme; steps 1, 4, 5, and 6 from Section 3 remain unchanged; step 3 is eliminated; and step 2 is modified. As a whole, the changes are as follows:

– In step 2 (replacing coarse grid values with averages of fine grid values), Eq. 47 (averaging barotropic

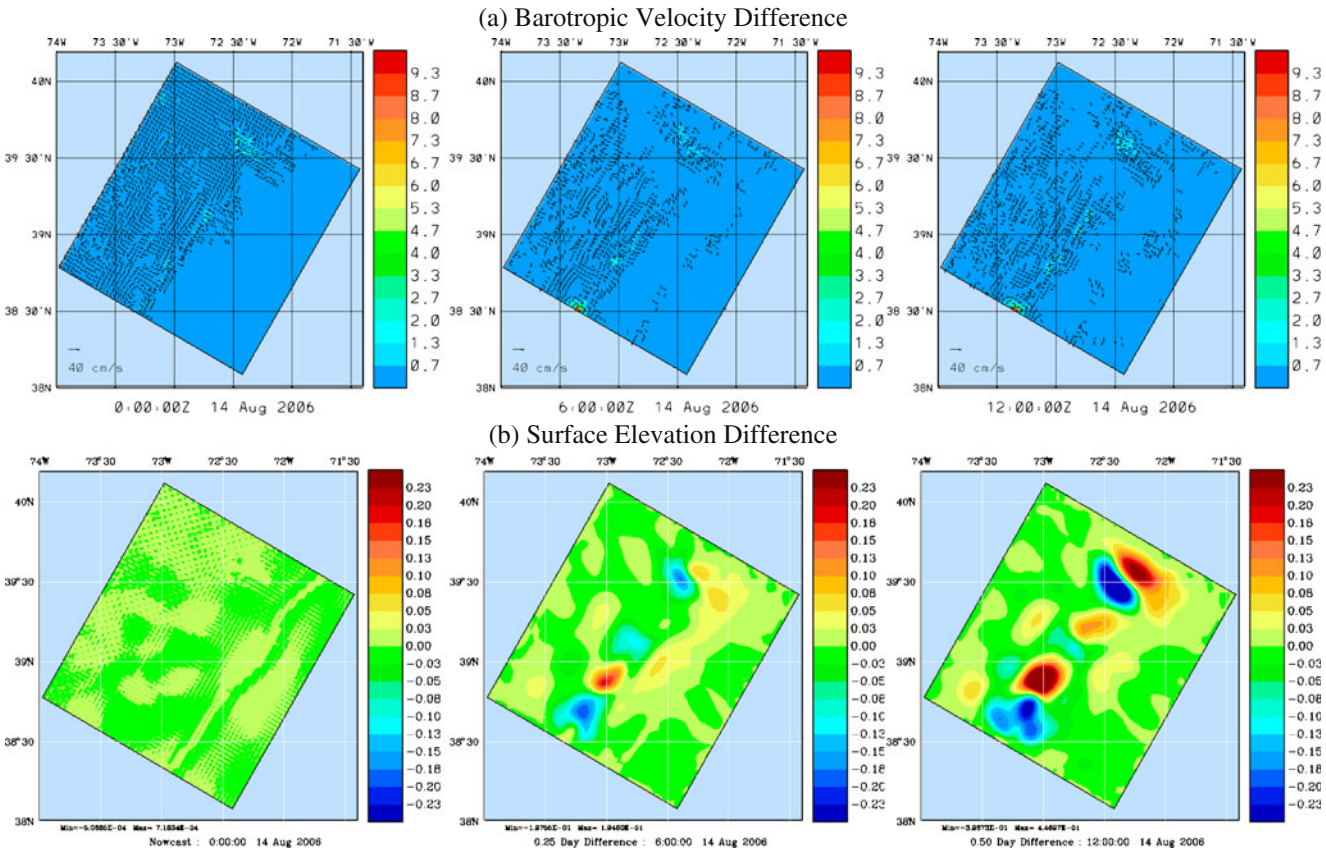

Fig. 7 Scheme 2: average $\hat{\mathbf{U}}$ not $\mathcal{F}$. **a** Vector difference in barotropic velocity between coarse and fine domains plotted in the fine domain for 00Z, 06Z, and 12Z on 14 Aug. (overlain on magnitude of vector difference). **b** Difference in surface elevation difference between coarse and fine domains plotted in the fine domain for 00Z, 06Z, and 12Z on 14 Aug. Notice the large (sub)-mesoscale differences in the surface elevation that develop within a half day

forcing), replace $\left(H_{i,j}+\eta^{n}_{i,j}\right)\hat{\mathbf{U}}^{n+1}_{i,j}$ by $\hat{\mathbf{U}}^{n+1}_{i,j}$ (i.e., transfer velocity instead of transport).

– Eliminate step 3 (making time-lagged coarse grid barotropic velocity consistent with time-lagged fine grid averaged surface elevation).

In Fig. 8, we again plot vector differences between the fine and coarse estimates for $\mathbf{U}$, η, using this third scheme. Here too, the simulation is cut short by a local CFL violation at 14.9 days into the run. The problem takes time to develop and we thus examine the differences at 3, 8, and 14 days. In the domain as a whole, the differences in both $\mathbf{U}$ and η are small in magnitude and scale. The magnitude of the velocity difference is generally <0.7 cm/s with regions of 0.7–1.3 cm/s mainly near the shelfbreak and Hudson canyon. However, where the shelfbreak intersects the southern boundary, there is a growing region where velocity differences reach O(10 cm/s). This eventually leads to a local CFL violation. The difference in the η estimates remains small, in the range ±3 cm over most of the domain and bounded by ±7 cm in the region of large $\mathbf{U}$ differences. Taken as a whole, this indicates that this scheme produces the overall desired level of coupling between the coarse and fine domains but is overly sensitive (see Section 4.7.2).

4.4 Scheme 4: update $\mathbf{U}^n$ as function of η^n

The improvement in this scheme is not in exchanging additional fields between the coarse and fine domains but in making sure that the values that are exchanged are used as consistently as possible in the free-surface PE algorithm. Specifically, we use Eq. 44 to correct the time lagged barotropic velocity in the coarse domain after receiving the averaged time lagged surface elevation from the fine domain. Comparing this scheme 4 to the consistent implicit scheme of Section 3: Scheme 4 is a

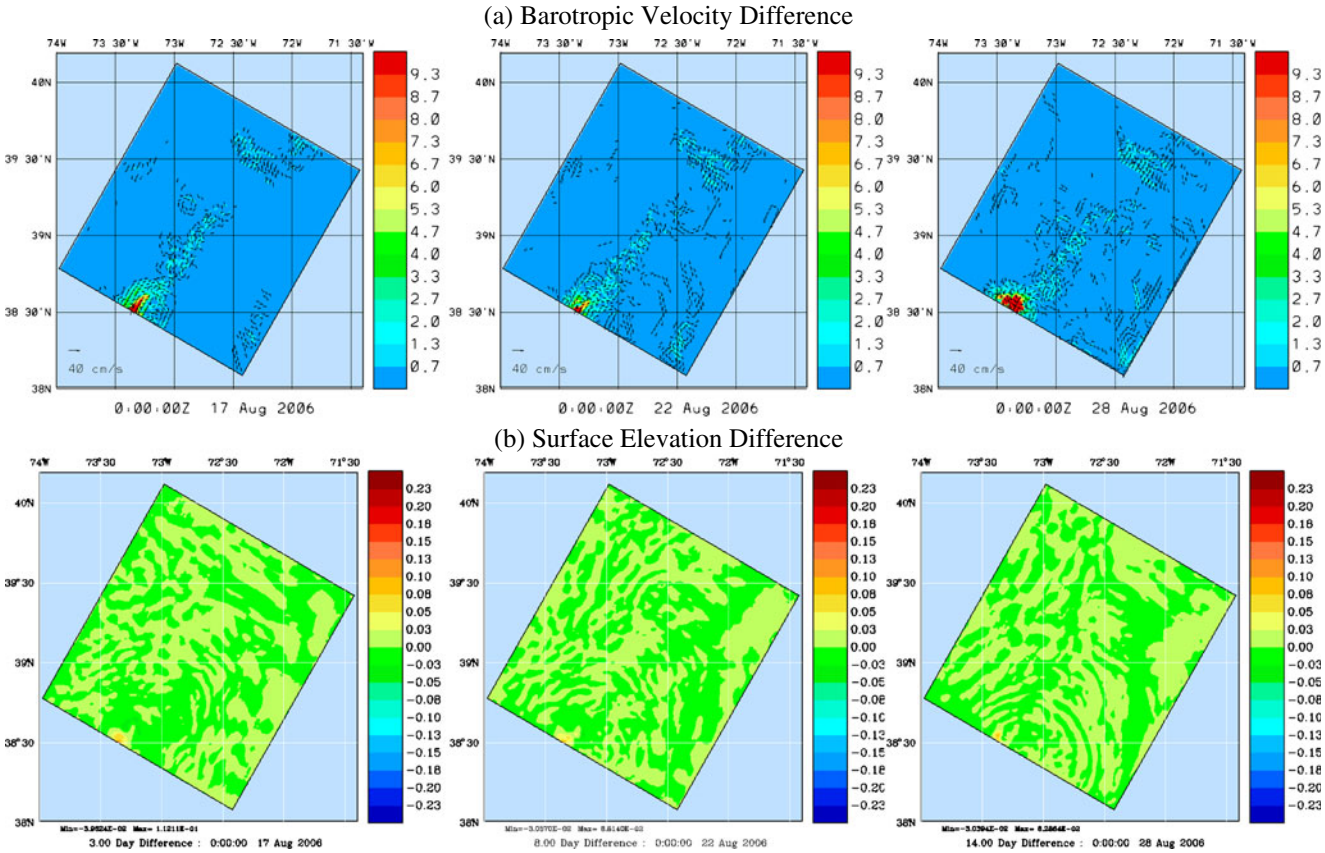

Fig. 8 Scheme 3: exchange η^n. **a** Vector difference between barotropic velocity in coarse and fine domains plotted in the fine domain for 00Z on 17 Aug., 22 Aug., and 28 Aug. (overlain on magnitude of vector difference). **b** Difference between surface elevation in coarse and fine domains plotted in the fine domain for 00Z on 17 Aug., 22 Aug., and 28 Aug. Notice the growing velocity misfit caused by an instability at the shelfbreak/southern boundary intersection

six-step scheme; steps 1, 3, 4, 5, and 6 from Section 3 remain unchanged; and step 2 is modified. As a whole, the changes are as follows:

– In step 2 (replacing coarse grid values with averages of fine grid values), Eq. 47 (averaging barotropic forcing), replace $\left(H_{i,j}+\eta^{n}_{i,j}\right)\hat{\mathbf{U}}^{n+1}_{i,j}$ by $\hat{\mathbf{U}}^{n+1}_{i,j}$ (i.e., transfer velocity instead of transport).

Figure 9 shows vector differences between the fine and coarse grid estimates for $\mathbf{U}$, η, using this implicit nesting scheme, at 3, 17, and 35 days into the coupled simulations. Here differences between $\mathbf{U}$ and η are still small in magnitude and scale. Difference magnitudes for $\mathbf{U}$ are <0.7 cm/s over the majority of the domain with regions of 0.7–2 cm/s generally near the shelfbreak and Hudson canyon. The intersection of the shelfbreak with the southern boundary excites an isolated spot of larger differences, O(1–10) cm/s between the coarse and fine $\mathbf{U}$. However, with this scheme, these boundary differences remain confined to a small region near the boundary and bounded. Moreover, these differences are not monotonic but intermittent, growing, and fading between 4 and 10 cm/s repeatedly during the simulation: They are partly due to tidal and inertial responses that differ slightly in the fine and coarse domain. This can lead to localized small intermittent misfits. Differences in η remain small, in the range ±3 cm over most of the domain and bounded by ±7 cm in the region of larger $\mathbf{U}$ differences. As the velocity differences, the elevation differences in this region are intermittent, growing, and fading repeatedly during the simulation. When compared to the previous schemes, this is the first scheme that possesses sufficient coupling for consistent estimates between the coarse and fine domains and sufficient robustness for use in realistic simulations. However, it does not conserve transport (see Section 4.7.2).

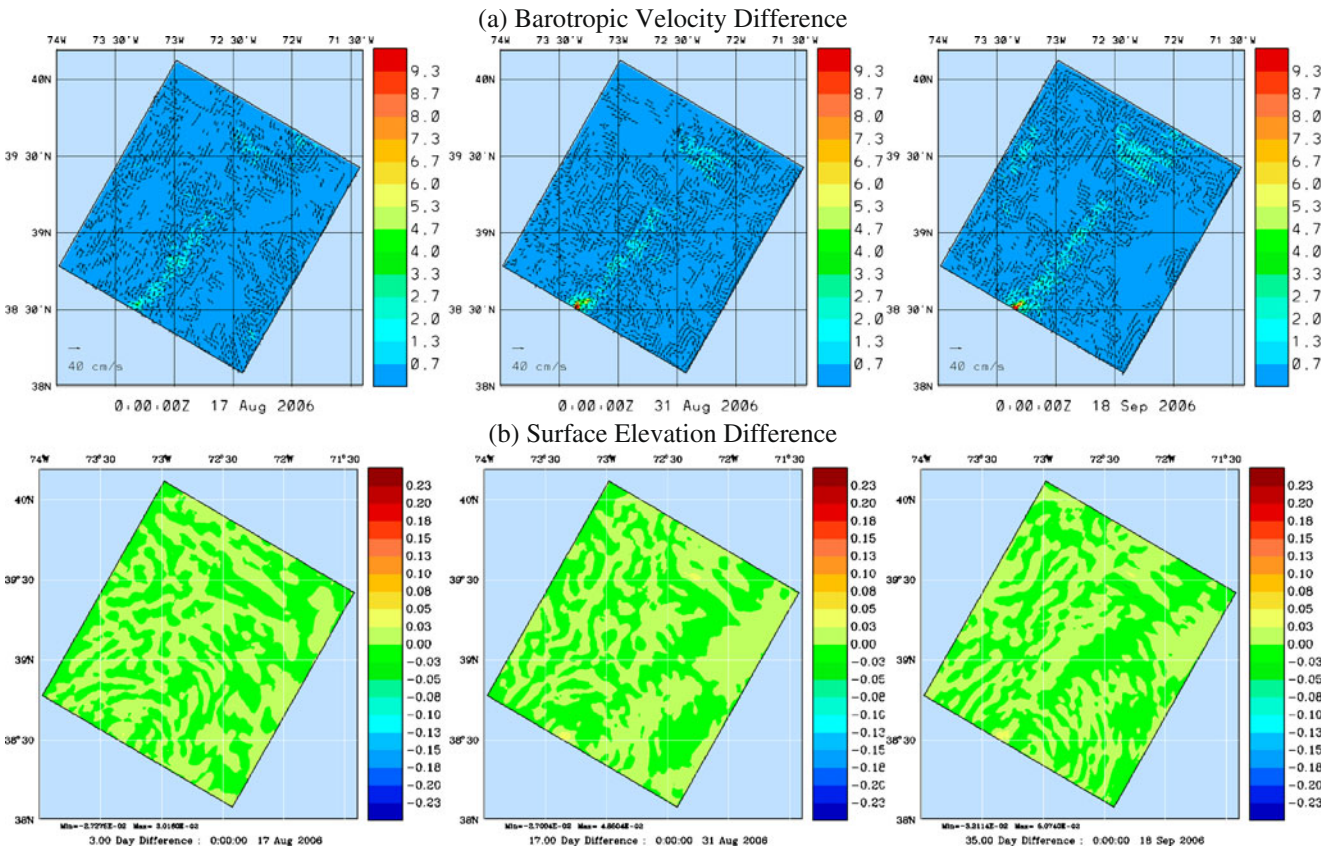

Fig. 9 Scheme 4: update $\mathbf{U}^n$ as a function of η^n. **a** Vector difference between barotropic velocity in coarse and fine domains plotted in the fine domain for 00Z on 17 Aug., 31 Aug., and 18 Sep. (overlain on magnitude of vector difference). **b** Difference between surface elevation in coarse and fine domains plotted in the fine domain for 00Z on 17 Aug., 31 Aug., and 18 Sep. Velocity and elevation differences generally small with intermittent misfits at the shelfbreak/southern boundary intersection

4.5 Scheme 5: pass $(H+\eta^n)\,\hat{\mathbf{U}}^{n+1}$ ("transport")

This is the fully implicit two-way nesting scheme which we presented in Section 3. This scheme improves upon scheme 4 by casting Eq. 47 in terms of transport instead of velocity. This brings Eq. 47 in line with Eqs. 45, 49, and 50 which were already written in terms of transports. Averaging and interpolating transports instead of velocities was chosen to enhance the long-term stability of the simulations by ensuring the consistency of the mass flux estimates between the coarse and fine domains.

Figure 10 shows vector differences between the fine and coarse estimates for **U**, η, using this fully implicit nesting scheme, at 3, 17, and 35 days into the coupled simulations. Differences between the estimates of **U** in the coarse and fine domains are generally less than 1 cm/s. Larger differences, between 1 and 4 cm/s, mainly occur at the shelf break and in Hudson Canyon, which are due to the superior ability of the fine domain to represent these topographic features. Peak differences for **U** again occur where the shelfbreak intersects the southern boundary, reaching O(1–10 cm/s). They are smaller than those of scheme 4 and show the same intermittent nature. It should also be noted that differences remain small before (Aug. 17), during (Aug. 31), and after (Sep. 18) the passage of tropical storm Ernesto. This indicates that the strength of the coupling of the coarse and fine solutions is not a function of the velocity magnitude. The differences between the estimates of η are generally bounded by ±3 cm, with peak values of around ±5 cm at the intersection of the shelfbreak and the southern boundary. The small improvement in the barotropic velocity coupled with the long-term advantages of maintaining consistent estimates of mass fluxes in the two domains led to the selection of this scheme as our fully implicit two-way nesting scheme (see Section 4.7.2). Note that at the end of each time step, a perfect nesting would not lead to zero differences between the coarse and fine estimates on the fine grid (differences

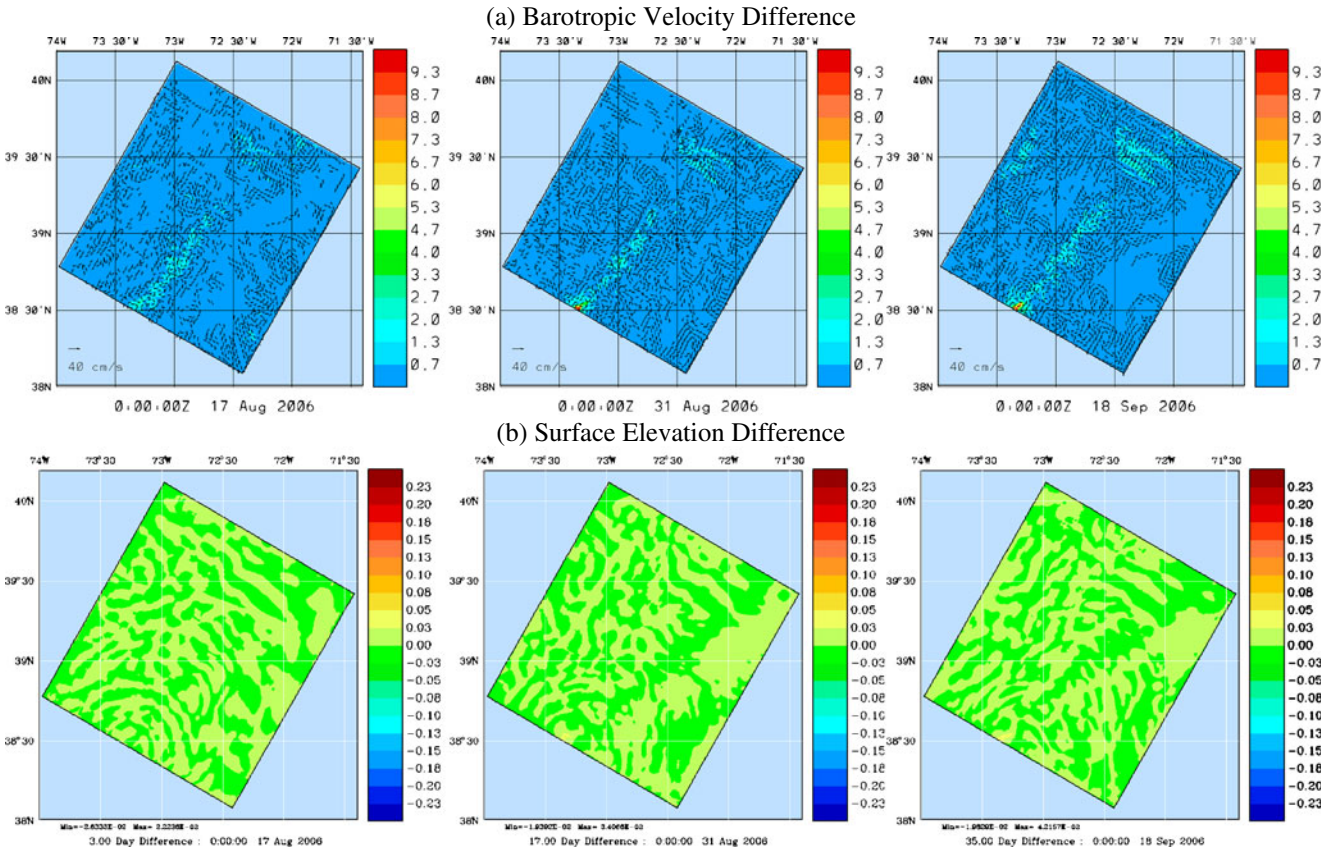

Fig. 10 Scheme 5: pass $(H+\eta^n)\,\hat{\mathbf{U}}^{n+1}$ ("Transport"). **a** Vector difference between barotropic velocity in coarse and fine domains plotted in the fine domain for 00Z on 17 Aug., 31 Aug., and 18 Sep. (overlain on magnitude of vector difference). **b** Difference between surface elevation in coarse and fine domains plotted in the fine domain for 00Z on 17 Aug., 31 Aug., and 18 Sep. Velocity and elevation differences small with intermittent misfits at the shelfbreak/southern boundary intersection

are only zero on the coarse grid). In perfect nesting, fine-grid differences vary at each time step due to dynamics, but they do not grow with the duration of integration.

4.6 Coarse-to-fine implicit nesting

In this section, we compare our fully implicit two-way nesting scheme (Section 3) to a more traditionally organized coarse-to-fine implicit two-way nesting scheme. We start by first designing the coarse-to-fine implicit scheme that most closely matches our fully implicit scheme.

1. Solve Eqs. 37–42 simultaneously in each domain for $(\mathbf{u}'^{n+1}\Delta z^{n+1}, \hat{\mathbf{U}}^{n+1}, T^{n+1}\Delta z^{n+1}, S^{n+1}\Delta z^{n+1})$
2. In the coarse domain, solve Eqs. 43 and 44 for η^{n+1}, $\mathbf{U}^{n+1}$, Δz^{n+1}, $\mathbf{u}'^{n+1}$, T^{n+1}, S^{n+1}.
3. Using piece-wise bi-cubic Bessel interpolation, $\mathcal{B}$, replace values in the fine grid boundary with values interpolated from the coarse grid

$$\phi^{n+1}_{i_{fb},j_{fb},k} = \mathcal{B}\left(\phi^{n+1}_{i_c,j_c,k}\right), \tag{51}$$

$$\mathbf{u}'^{n+1}_{i_{fb},j_{fb},k}\Delta z^{n+1}_{i_{fb},j_{fb},k} = \mathcal{B}\left(\mathbf{u}'^{n+1}_{i_c,j_c,k}\Delta z^{n+1}_{i_c,j_c,k}\right), \tag{52}$$

$$\mathbf{U}^{n+1}_{i_{fb},j_{fb},k} = \mathcal{B}\left[\left(H_{i_c,j_c} + \eta^{n+1}_{i_c,j_c}\right)\mathbf{U}^{n+1}_{i_c,j_c}\right] \times \frac{1}{H_{i_{fb},j_{fb}} + \eta^{n+1}_{i_{fb},j_{fb}}} \tag{53}$$

   where

   $\phi = T,\ S,\ \eta^{n+1}$.

4. In the fine domain, solve Eqs. 43 and 44 for η^{n+1}, $\mathbf{U}^{n+1}$, Δz^{n+1}, $\mathbf{u}'^{n+1}$, T^{n+1}, and S^{n+1}.
5. Replace values in the coarse domain at overlap nodes with the following averages from the fine domain values

$$\phi^{n+1}_{i_c,j_c,k}\Delta z^{n+1}_{i_c,j_c,k} = \frac{1}{\Delta\mathcal{A}_{i_c,j_c}}\sum_{j=j_{fc}-r_h}^{j_{fc}+r_h}\sum_{i=i_{fc}-r_h}^{i_{fc}+r_h}\phi^{n+1}_{i,j,k}\Delta\mathcal{V}^{n+1}_{i,j,k}, \tag{54}$$

$$\eta^{n}_{i_c,j_c} = \frac{1}{\Delta\mathcal{A}_{i_c,j_c}}\sum_{j=j_{fc}-r_h}^{j_{fc}+r_h}\sum_{i=i_{fc}-r_h}^{i_{fc}+r_h}\eta^{n}_{i,j}\Delta\mathcal{A}_{i,j} \tag{55}$$

   where

   $\phi = \mathbf{u}', T, S \quad ; \quad r_h = \lfloor \frac{r}{2} \rfloor$.

6. In the coarse domain, recompute $\mathbf{U}^n$ from Eq. 44 and updated η^n.

Figure 11 shows vector differences between the fine and coarse estimates for $\mathbf{U}$, at 3, 17, and 35 days into the coupled simulations. The top row shows results from the coarse-to-fine implicit scheme, the bottom row from our final fully implicit scheme. The coarse-to-fine implicit scheme leads to differences between the estimates of $\mathbf{U}$ in the coarse and fine domains which are generally between 1 and 2 cm/s, with peak values around 10 cm/s. These differences are organized on a smaller scale than in scheme 1. This improvement over scheme 1 is due to proper couplings maintained in the coarse-to-fine implicit scheme. However, when compared to our fully implicit scheme (bottom row), we see that the peak differences in the fully implicit scheme are bounded by 2 cm/s and most differences are less than 1 cm/s. The differences are also of smaller scales than those of the coarse-to-fine implicit scales. Overall, our final implicit scheme is much more consistent than a coarse-to-fine implicit scheme (see Section 4.7.2).

Note that here we are also testing the effects of smoothing the transition between the fine and coarse resolution topographies near the boundaries of the fine domain. In the two runs of Fig. 11, we used a six coarse-grid point transition (see "Appendix 2.1.1"). Comparing row (b) of Fig. 11 (with this topography transition) to row (a) of Fig. 10 (with no transition), we find that the intermittent spot of large velocity difference at the intersection of the shelfbreak with the southern boundary is absent in Fig. 11 (i.e., absent in the case of a smooth transition between the coarse and fine resolution topographies). This indicates that one of the factors driving those large isolated differences was the proximity of an artificial topography structure (the sharp coarse-fine transition at the boundary).

4.7 Error analysis

In Sections 4.1–4.5, we derived a series of nesting schemes and compared their performance in realistic simulations. Now, we complete a theoretical error analysis of the improvements among schemes. In general, the change between successive schemes is due to the use or nonuse of averaged values from the fine domain for the estimates in the coarse domain. In going from Section 4.1 to Section 4.2, we directly averaged the $\hat{\mathbf{U}}$ field not the $\mathcal{F}$ field. This is equivalent to saying that, in our $\hat{\mathbf{U}}$ estimate, we upgraded our coarse estimate of $\nabla\eta$ with averaged values from the fine domain (see Eq. 42). In going from Section 4.2 to Section 4.3, we explicitly upgrade our coarse estimate of η^n with averages from the fine domain. The change from Section 4.3 to Section 4.4 uses the upgraded η^n to improve the estimate of $\mathbf{U}^n$. Therefore, we present

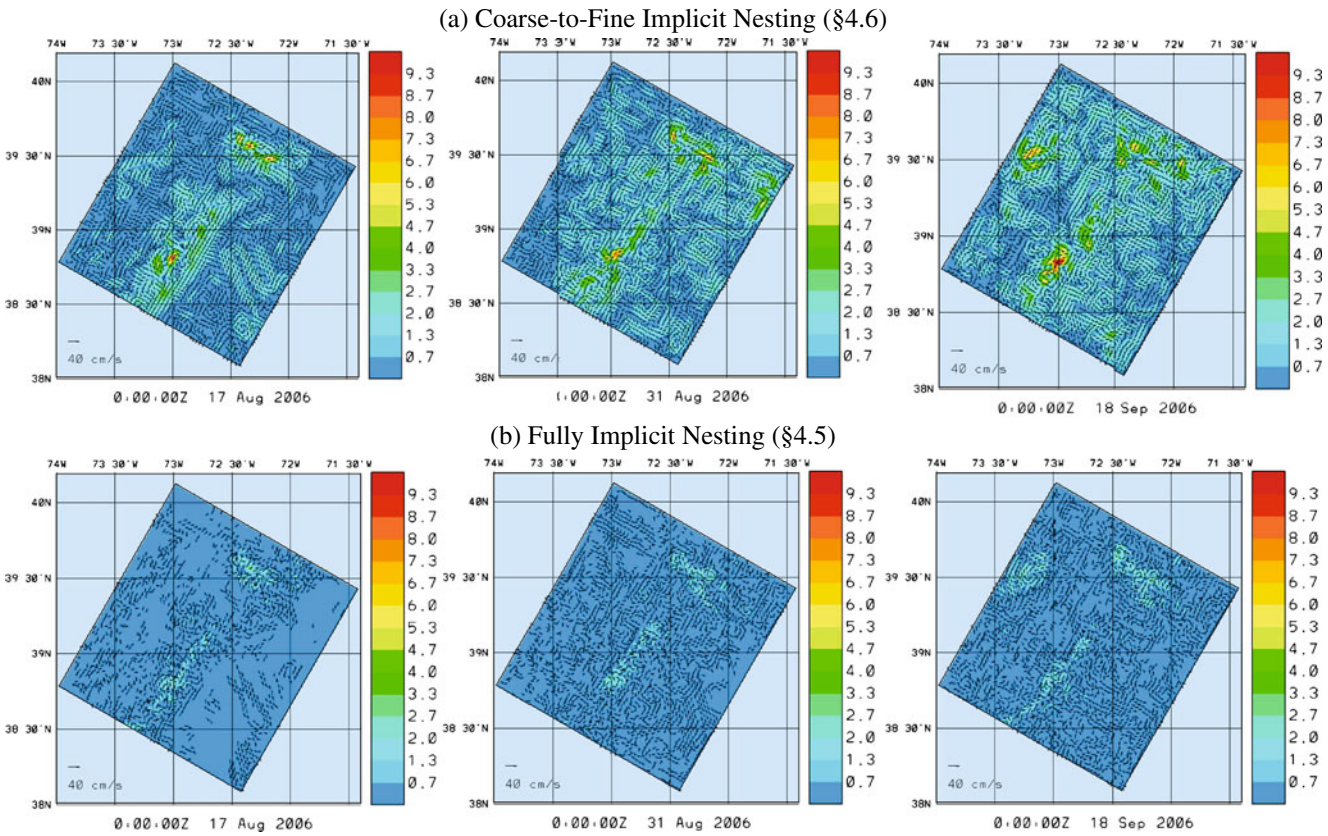

Fig. 11 Barotropic velocity differences: vector difference between barotropic velocity in coarse and fine domains plotted in the fine domain for 00Z on 17 Aug., 31 Aug., and 18 Sep. (overlain on magnitude of vector difference). **a** Differences for coarse-to-fine implicit nesting scheme. **b** Differences for fully implicit nesting scheme. Overall, our fully implicit scheme is much more consistent than a coarse-to-fine implicit scheme

the truncation error analysis for the averaging update in general and then apply this general analysis to the individual schemes.

4.7.1 General error analysis

The horizontal averaging operation for these fields is the central point approximation of a 2D integral. The error in this approximation can be easily shown to be of second order (Ferziger and Perić 1996):

$$\int_{\frac{-\Delta y}{2}}^{\frac{\Delta y}{2}} \int_{\frac{-\Delta x}{2}}^{\frac{\Delta x}{2}} \phi \, dxdy \equiv \left[\phi|_{(0,0)} + O\left(\Delta x^2\right) + O\left(\Delta y^2\right) \right] \Delta x \Delta y.$$

The r:1 fine-to-coarse averaging operation can then be written as

$$\begin{aligned}
\phi_{i_c,j_c} &= \frac{1}{\Delta x_c \Delta y_c} \sum_{j_f=j_{fc}-r_h}^{j_{fc}+r_h} \sum_{i_f=i_{fc}-r_h}^{i_{fc}+r_h} \int_{\frac{-\Delta y_f}{2}}^{\frac{\Delta y_f}{2}} \int_{\frac{-\Delta x_f}{2}}^{\frac{\Delta x_f}{2}} \phi \, dxdy \\
&= \frac{1}{\Delta x_c \Delta y_c} \sum_{j_f=j_{fc}-r_h}^{j_{fc}+r_h} \sum_{i_f=i_{fc}-r_h}^{i_{fc}+r_h} \left[\phi_{i_f,j_f} + O\left(\Delta x_f^2\right) \right. \\
&\quad \left. + O\left(\Delta y_f^2\right) \right] \Delta x_f \Delta y_f \\
&= \frac{1}{\Delta x_c \Delta y_c} \sum_{j_f=j_{fc}-r_h}^{j_{fc}+r_h} \sum_{i_f=i_{fc}-r_h}^{i_{fc}+r_h} \left[\phi_{i_f,j_f} + O\left(\Delta x_f^2\right) \right. \\
&\quad \left. + O\left(\Delta y_f^2\right) \right] \frac{1}{r^2} \Delta x_c \Delta y_c \\
&= \langle \phi \rangle_{i_{fc} \pm r_h, j_{fc} \pm r_h} + \left\langle O\left(\Delta x_f^2\right) \right\rangle_{i_{fc} \pm r_h, j_{fc} \pm r_h} \\
&\quad + \left\langle O\left(\Delta y_f^2\right) \right\rangle_{i_{fc} \pm r_h, j_{fc} \pm r_h}.
\end{aligned}$$

where $r_h = \lfloor r/2 \rfloor$ and $\langle \phi \rangle_{i_{fc} \pm r_h, j_{fc} \pm r_h}$ is the average value of ϕ over the $r \times r$ array of fine cells. From this we see that the estimate for ϕ_{i_c,j_c} as averaged on the fine grid is second order in the fine grid spacing, $O\left(\Delta x_f^2\right)$. The estimate for ϕ_{i_c,j_c} based on the coarse grid primitive equations is second order in the coarse grid, $O\left(\Delta x_c^2\right)$. Even for only a 3:1 ratio in the grid spacing, this equates to an order of magnitude smaller errors obtained by averaging the fine grid estimate. Furthermore, assuming that the time step is small enough to resolve the physical

processes (up to their second derivatives), then the signs of the averaged error terms will remain constant over many time steps, providing a bias on the scale of time stepping and a seed for larger-scale biases built up through nonlinear interactions.

As a side note, the above error analysis shows an unambiguous error reduction when using the averages of the fine grid values to replace the coarse grid estimates. This definitive statement is due to the fact that the averaging scheme and the discretization of the primitive equations (Section 2.4) are both second order. If one was to use higher-order methods to discretize the primitive equations, then the averaging scheme used should at least match the order of the discretization. The use of higher-order averaging schemes might require extra filtering of the smallest scales in the fine domain estimates to avoid aliasing (Debreu and Blayo 2008, Section 4.1).

4.7.2 The general error analysis applied to the specific schemes

Scheme 1, Section 4.1, and Fig. 6: In addition to having the same errors as the later schemes, this scheme uses an estimate of $\nabla\eta$ that is entirely based on the coarse domain fields to compute $\hat{\mathbf{U}}$. As shown above (Section 4.7.1), this maintains $O(\Delta x_c^2)$ in $\nabla\eta$ rather than $O(\Delta x_f^2)$ errors. Moreover, these are second-order errors in $\nabla\eta$; hence, the leading order error terms will be proportional to the third derivatives in η. These are more "singular" derivatives then the second-order derivatives in our PE scheme, explaining why they would have to be larger somewhere and spawn the dominant biases in Fig. 6. The fact that these errors are directly fed into $\hat{\mathbf{U}}$, which in turn feeds directly into $\mathbf{U}$, explains why these biases appear in the barotropic velocity.

Scheme 2, Section 4.2, and Fig. 7: This scheme updates $\hat{\mathbf{U}}$ in the coarse domain with averaged values from the fine domain, thereby reducing the error in $\nabla\eta$ in Eq. 42 but provides no feedback from the fine grid estimates of η to the coarse grid. Hence, the coarse domain errors in η remain everywhere $O(\Delta x_c^2)$ rather than having $O(\Delta x_f^2)$ in the overlap region. Given direct coupling of the $\hat{\mathbf{U}}$ estimates between the domains and the indirect coupling of $\mathbf{U}$ via Eq. 44, these errors only have the freedom to excite biases in η, as shown in Fig. 7.

Scheme 3, Section 4.3, and Fig. 8: This scheme updates η^n in the coarse domain with averaged values from the fine domain. These direct couplings of $\hat{\mathbf{U}}$ and η^n prevent the domain-wide biases seen in schemes 1 and 2 (Figs. 6 and 7). However, the coarse domain estimate of $\mathbf{U}^n$ is still based on the values of η^n that were available when the coarse domain computed $\mathbf{U}^n$ from Eq. 44; hence, the coarse domain errors in $\mathbf{U}^n$ remain everywhere $O(\Delta x_c^2)$ rather than having $O(\Delta x_f^2)$ errors in the overlap region. These errors can feed local instabilities, like those caused by the different coarse and fine representations of the shelfbreak topography across the southern boundary of the fine domain (Fig. 8).

Scheme 4, Section 4.4, and Fig. 9: This scheme updates $\mathbf{U}^n$ in the coarse domain with estimates of η^n which have been updated with averaged values from the fine domain. This means that each term in Eq. 44 now has errors of $O(\Delta x_f^2)$ in the overlap region rather than $O(\Delta x_c^2)$. As seen in Fig. 9, this produces a stable scheme with small scale errors.

Scheme 5, Section 4.5, and Fig. 10: This is the scheme we selected as the best. It updates the transport, i.e., the product $(H+\eta^n)\,\hat{\mathbf{U}}^{n+1}$, in the coarse domain with averages from the fine domain rather than updating the coarse estimate of $\hat{\mathbf{U}}^{n+1}$. In doing so, we add no new averaged values from the fine grid and hence make none of the error reductions described in Section 4.7.1. This is consistent with the observation that the resulting changes (Figs. 9 and 10) are relatively small. The advantage is that scheme 5 conserves transport at $O(\Delta x_f^2)$ from the fine-to-coarse grids while scheme 4 does not.

Coarse-to-Fine Implicit Scheme, Section 4.6, and Fig. 11: With this nesting scheme, we follow traditional coarse-to-fine implicit schemes. In doing so, we introduce two error sources of the type described in Section 4.7.1. First, when interpolating coarse domain values to the boundary of the fine domain, the coarse domain values used have errors of $O(\Delta x_c^2)$. In our fully implicit scheme, the coarse domain values in the overlap region have errors of $O(\Delta x_f^2)$. Second, we do not use fine grid averaged values of either $(H+\eta^n)\,\hat{\mathbf{U}}^{n+1}$ or $\hat{\mathbf{U}}^{n+1}$ to update the corresponding coarse grid values. This directly means that $\hat{\mathbf{U}}^{n+1}$ and $\mathbf{U}^{n+1}$ have errors of $O(\Delta x_c^2)$ in the overlap region rather than $O(\Delta x_f^2)$. Since $\hat{\mathbf{U}}^{n+1}$ is part of the forcing for Eq. 43, the effects of these larger errors can immediately spread outside of the overlap region due to the nonlocal nature of the Helmholtz operator (Eq. 43). Finally, we note that the computational cost of our fully implicit scheme 5 is pretty much equivalent to that of the more classic scheme 6. A major conclusion of our work is that fully

implicit (space and time) nesting schemes should be used when possible.

5 Examples

5.1 The middle Atlantic Bight

We continue the study of our fully implicit scheme (Section 3) and its use (Section 4.5) by now examining the consistency of the estimates of the 3D variables between the coarse and fine domains. In Fig. 12, we show the differences between the coarse and fine domain estimates of the temperature, salinity, and total velocity (T, S, $\mathbf{u}$), at a depth of 50 m. As with the barotropic fields, we see an excellent overall agreement between the estimates of the 3D fields in the two domains, in accord with their different resolutions. The temperature differences are mostly bounded by ±0.2°C, the salinity differences by ±0.02 PSU and the total velocity differences by ±1 cm/s. The larger differences occur in two main categories, the topography-driven dynamics (e.g., shelfbreak) and high-gradient dynamics (e.g., filaments). First, the fine domains better resolve the shelfbreak and the Hudson Canyon and the dynamics that these features generate. There, the differences can approach ±1.5°C for temperature, ±0.2 PSU for salinity and ±10 cm/s with intermittent peak spots around ±20 cm/s for velocity. Second, in the offshore regions where filamentation is taking place, larger differences are being generated at the edges of the filaments and eddies, which are better resolved in the fine domain. Note that these differences are not the same as the biases studies in Section 4. On the coarse grid, there would be no difference between the coarse and averaged fine grid solutions (coarse fields are replaced by averages of fine fields). On the fine grid, differences

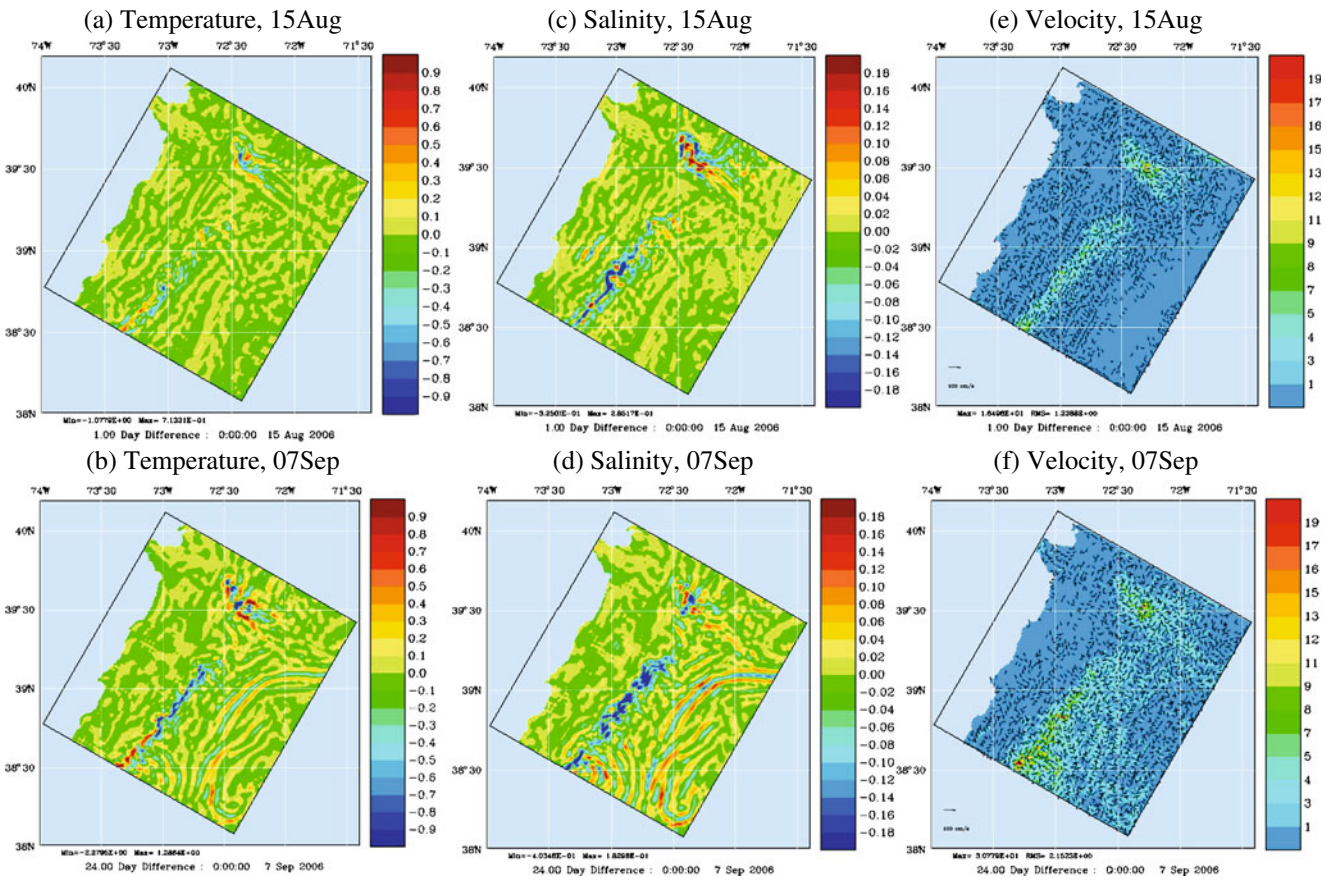

Fig. 12 Tracer and total velocity differences at 50 m in the fully implicit scheme, to illustrate baroclinic aspects and increased accuracy of finer nested domain. **a**, **b** Temperature differences. **c**, **d** Salinity differences. **e**, **f** Total velocity differences. **a**, **c**, **e** Differences at 1 day into the nested simulation. **b**, **d**, **f** Differences at 24 days into the nested simulation. Main differences occur at shelfbreak/canyon and offshore at edges of filament/eddies, all in regions where resolution is important

arise from the above dynamical and bathymetric reasons. Further, the coarse fields are bi-linearly interpolated to the fine grid, which cannot reproduce the real gradients on the fine grid.

In Fig. 13, we present a snap shot of the surface temperature fields at 11 Sep. 2006 overlaid with surface velocity vectors. Eleven days earlier, tropical storm Ernesto passed over the region, cooling the surface and advecting the shelfbreak front several kilometers offshore (Ernesto did not create any issues in nesting, see Section 4). In the relaxation which follows, filaments are being spun off of the shelfbreak front. These processes are well captured in this simulation. Additionally, Fig. 13 shows the continuity of the large-scale structures across the nesting boundary. No shocks or spurious waves are generated at the interface between the coarse and fine domains.

One of the most significant dynamical achievement of our new implicit nesting scheme is an increase in predictive capability. This is shown by comparing our estimates to independent acoustic Doppler current profiler (ADCP) data (T. Duda, personal communication) that were neither assimilated nor used in the initial conditions of the simulations. Results are illustrated in Fig. 14. The ADCP data from a mooring (SW30) are compared to velocity estimates from two different simulations. The first simulation (left panel) is the coarse 3-km resolution large domain run in "Stand Alone" mode, i.e., no nested subdomain. The second simulation (right panel) is the nested simulations using our new implicit scheme. The initial conditions, atmospheric and tidal forcings, assimilated data, and all model parameters are identical. The open boundary conditions for the two coarse 3-km resolution domains are also the same. The only difference is whether the 1-km resolution domain is nested in this 3 km or not. The results are dramatic. Simply including the high-resolution domain reduces the bias with respect to the mooring data from 12 to 2 cm/s and the RMS error from 15 to 8 cm/s. To assess the statistical significance of these improvements, we compute the standard deviation of subtidal signal (obtained by averaging the data with a $\pm$1-day window) about its mean, 5 cm/s, and the standard deviation of the tidal signal about the subtidal signal, 6 cm/s. Clearly the error reductions (10 and 7 cm/s) produced by the nesting are significant when compared to the variability in the data.

5.2 The Philippine archipelago

Our next realistic simulation results come from our research in the Philippine archipelago as part of the Philippines Straits Dynamics Experiment (PhilEx; Gordon 2009; Lermusiaux et al. 2009a). The goal of PhilEx was to enhance our understanding of physical and biogeochemical processes and features arising in and around straits and improve our capability to predict the spatial and temporal variability of these regions.

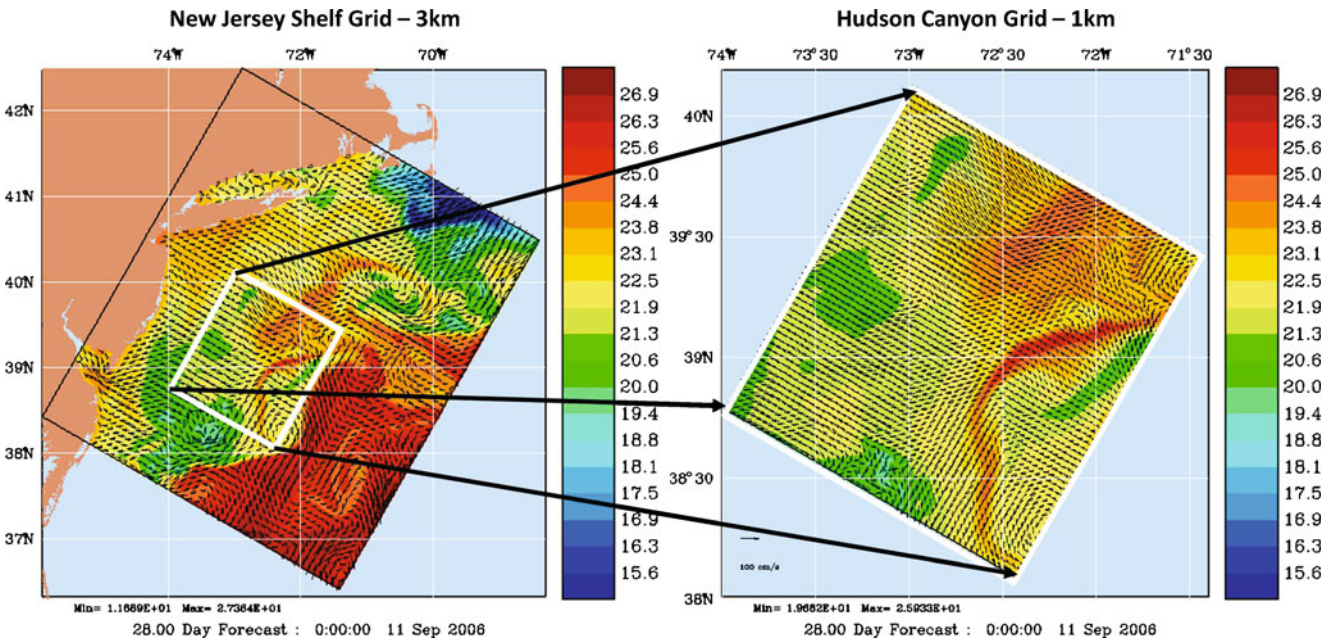

Fig. 13 Surface temperature, overlaid with surface velocity vectors, for 0000Z on 11 Sep. 2006 in the fully implicit two-way nested New Jersey shelf and Hudson Canyon domains

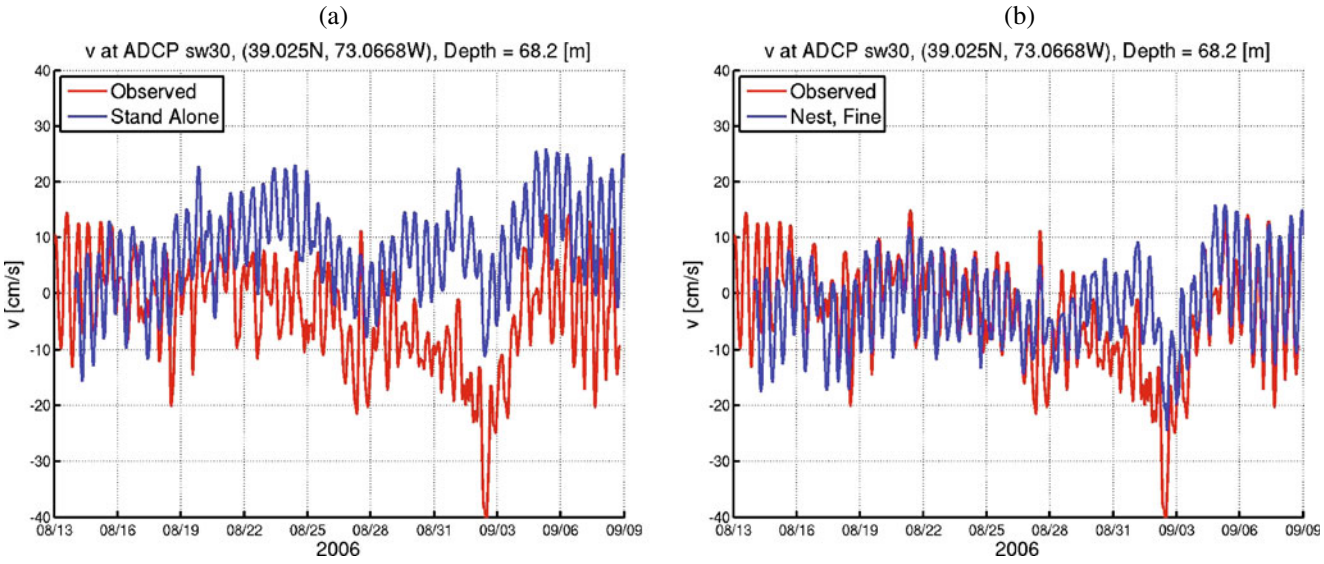

Fig. 14 Hourly meridional velocities (v) at 68 m depth at the location of mooring SW30, as measured by the moored ADCP (*red curves*) and as estimated by re-analysis simulations (*blue curves*) with atmospheric and barotropic tidal forcing. No mooring data are assimilated. **a** Comparing mooring velocities against velocity estimates from a 3-km simulation without nesting ("Stand Alone"). **b** Comparing mooring velocities against velocity estimates from a two-way nested simulation. Our new nesting scheme removes an $O(15$ cm/s) bias, as averaged during Aug. 22–Sep. 09. Notice that this bias reaches $O(30$ cm/s) during the Tropical Storm Ernesto

Here we used spherical coordinates and defined six two-way nested domains, in telescoping setups, ranging from a 3,267 × 3,429-km regional domain (with 27-km resolution) down to a pair of roughly 170 × 220-km strait domains with high (1-km) resolution (see Fig. 15). For physical, biogeochemical, and numerical parameter tuning and real-time forecasting, more than 1,000 simulations were run in this region, for three periods. The simulation shown here is for the Feb.–Mar. 2009 real-time experiment period, focusing on the 1,656 × 1,503 Philippine archipelago domain (9-km resolution) and the 552 × 519 Mindoro Strait domain (3-km resolution). Both domains have 70 vertical levels arranged in a double-σ configuration, optimized for the local steep bathymetry and depths of thermoclines/haloclines. Our bathymetry estimates merged profile data (C. Lee, personal communication) and ship data (Gordon and Tessler, personal communication) with V12.1 (2009) of the Smith and Sandwell (1997) topography. These simulations were initialized using SSH anomaly data (Colorado Center for Astrodynamics Research; Leben et al. 2002), climatological profiles (Locarnini et al. 2006), and our new mapping scheme (Agarwal and Lermusiaux 2010). Atmospheric forcing at the surface was obtained from Coupled Ocean Atmosphere Mesoscale Prediction System (COAMPS; wind stress) and Navy Operational Global Atmospheric Prediction System (NOGAPS; net heat flux, E-P) fields. For open boundary conditions (OBC), the transports from the HYCOM model were used. Our multiresolution tidal forcing was also used at the OBCs of our free-surface simulations (as well as in the initial conditions). SSH and SST are assimilated, but no in situ synoptic data are used, since one of the PhilEx goals was to evaluate if assimilating remotely sensed data in tuned models could capture some dynamics.

Figure 16 shows the surface velocity after 20 days of simulation. The Mindoro Strait domain (right panel) is used to resolve two main areas. The first is the Mindoro Strait which connects the South China Sea (northwest corner of the domain) to the Sulu Sea (southwest corner). The second is the Sibuyan Sea (interior of archipelago) which connects the Mindoro Strait to the Pacific Ocean via the San Bernardino Strait (12.5 N, 124.25 E). The higher resolution of the Mindoro Strait domain resolves the various pathways of the region. In the snapshot shown, the tides are favoring inflow from the Pacific. This inflow turns primarily northward along the island of Luzon. Even with two-way nesting, this pathway is poorly resolved in the Philippine archipelago domain (left panel). The Philippine archipelago domain provides the external forcing to the Mindoro Strait domain. Looking at the left panel, we again see the continuity of the flow across the boundary of the Mindoro Strait domain.

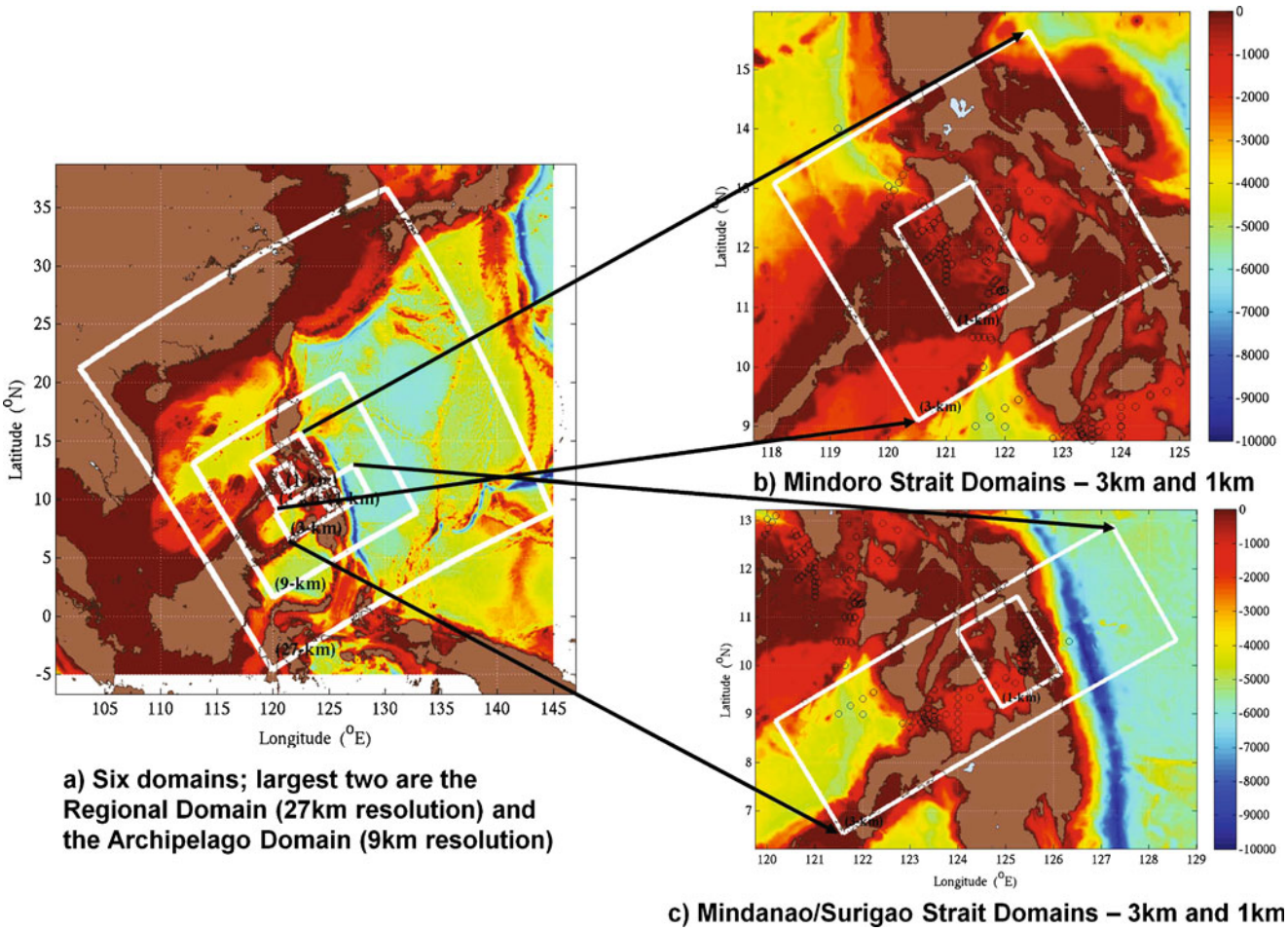

Fig. 15 Six spherical-grid domains in a telescoping zoom configuration for multiscale simulations in the Philippine archipelago

We also compared our nested fields to independent in situ data. As an example, we utilize the results of Tessler et al. (2010) who analyzed time series from moored ADCP that were deployed around the Panay Sill (11°16′ N, 121°55′ E) in the Mindoro strait system. As in our simulations, they found that the mean velocity showed a bottom intensified southward flow (lower 170 m) into the Sulu Sea. For the Feb.–Mar. 2009 time frame, Tessler et al. (2010) estimate a mean transport of 0.35 Sv southward into the Sulu Sea, with a standard deviation of 0.07 Sv. We compare this to a time-averaged transport of the bottom 170 m through a section from (11.305 N, 121.8003 E) to (11.3746 N, 121.9275 E) taken from our nested simulation. Our model estimate is a mean transport of 0.39 Sv into the Sulu Sea. Our ensemble of simulations gives a model standard deviation of 0.2 Sv, reflecting model parameter uncertainty. To show the impact of two-way nesting, we compare these runs to stand-alone coarse Archipelago runs, with an identical setup, but without the finer nested Mindoro domain. For this coarse-domain-only case, we find that the time-averaged net transport in the bottom 170 m is close to zero and in the central simulation, it is reversed (0.28 Sv northward out of the Sulu sea). This demonstrates the impact that resolution can have on the dynamical flow structure. Our multiresolution approach is likely important also for larger basin-scale and climate studies.

5.3 The Taiwan–Kuroshio region

Our last realistic simulation results come from our research in the Taiwan–Kuroshio region (see Liang et al. 2003, Fig. 1). This research was part of the Quantifying, Predicting, and Exploiting Uncertainty initiative (Gawarkiewicz 2008; Lermusiaux et al. 2009c), which aims to integrate coupled ocean-acoustic modeling, multidisciplinary data assimilation, and autonomous ocean platforms to improve prediction and reduce uncertainties. We defined a pair of nested domains, each with Cartesian coordinates and 70 levels. The

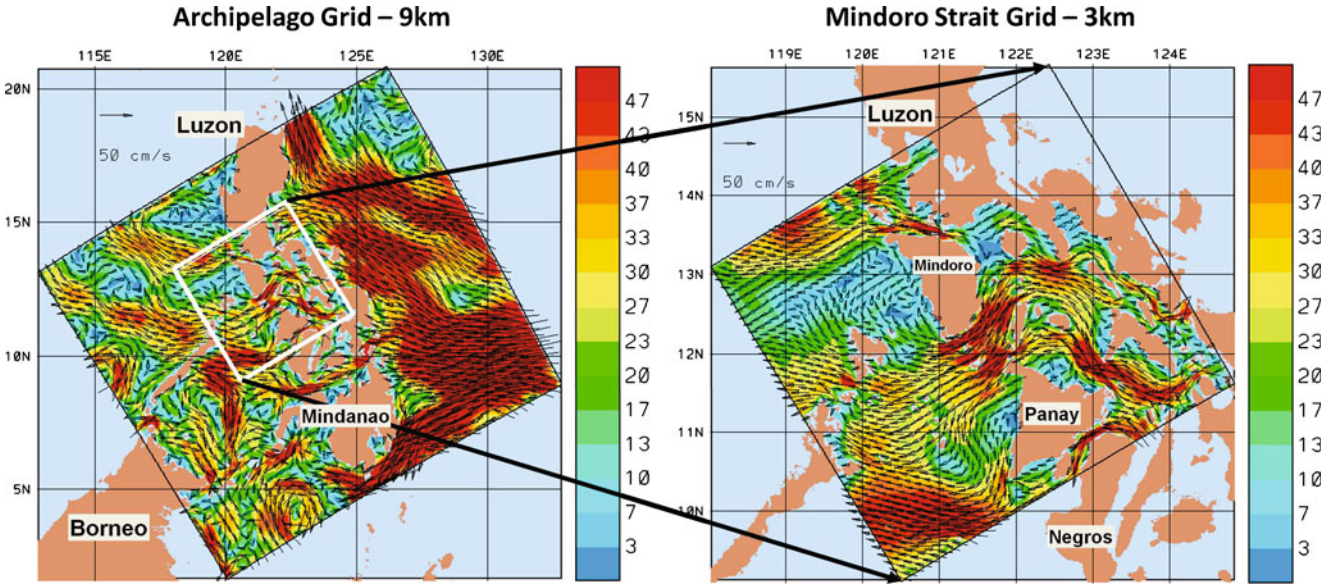

Fig. 16 Surface velocity at 0600Z on 22 Feb. 2009 in the Philippine archipelago, estimated by our new fully implicit two-way nesting. *Left panel*: the velocity in the 9-km Archipelago domain. *Right panel*: the surface velocity in the 3-km Mindoro Strait domain

larger domain spanned 1,125 × 1,035 km with a 4.5-km resolution. This domain was designed to maintain an accurate synoptic estimate of the Kuroshio and other regional influences on the intensive survey area. The smaller domain covered a 345 × 386-km area with 1.5-km resolution. This domain was designed to resolve the intensive survey, the recurring cold dome, and the Kuroshio–shelf interactions, especially in the region of the North Mien-Hwa canyon system. Our bathymetry estimate merges high-resolution data (B. Calder, personal communication) with V10.1 (2008) of the Smith and Sandwell (1997) topography. The initial and boundary fields were created using ship initialization surveys (OR2/OR3), Sea-Glider data and SST analyses, and a background constructed from high-resolution August WOA-05 climatology with deep Summer WOA-05 climatology profiles (Locarnini et al. 2006). The simulations were forced with a combination of COAMPS (wind stress) and NOGAPS (net heat flux, E-P) atmospheric fields along with barotropic tides. The barotropic tidal forcing was computed using the Logutov and Lermusiaux (2008) model, the regional high-resolution bathymetry, in situ tidal gages, and coarser-resolution tidal simulations from the Egbert global model (Egbert and Erofeeva 2002). During the IOP09 (Lermusiaux et al. 2009c), ensembles of ESSE simulations were used for uncertainty forecasting, over a 1-month period.

Figure 17 shows the relative vorticity computed at 50 m depth, 10 days (0000Z on 28 Aug. 2009) into 1-month-long realization of the ensemble of simulations that were carried out. This is a representative simulation for which the fine domain (right panel) captures the vortex generation of the Kuroshio passing over the I-Lan ridge (starting from Taiwan at 24.5 N, 121.9 E and extending to the southeast). A well-developed vortex wake is clearly visible trailing to the northeast off of Yonaguni Island (24.45 N, 123 E). Downstream of the I-Lan ridge, the vorticity field shows an eddy trapped between the Kuroshio and the shelf. The various off-shelf vorticity wakes generally follow the Kuroshio to the northeast, out of the domain. On the shelf, the interaction of tidal currents with topography produces a tight vorticity signal along the 50-m isobath just north of Taiwan. Across the mouth of the Taiwan strait, we find another (weaker) interaction of tidal currents and bathymetry, aligned north by northeast roughly along the 80-m isobath. In the coarse domain (left panel), the averaged versions of the small domain features are maintained. The vortex wakes streaming out of the small domain are smoothly continued in the external portions of the large domain. The nesting also maintains the wake off Yonaguni Island in the large domain even though the island itself is not explicitly represented in the large domain. Outside of the small domain, similar wakes, topographic generation, and eddies are present though of necessarily larger scale. We have also compared 2 to 3 days forecasts with in situ temperature and salinity data (Newhall et al. 2010). We find that the two-way nested simulations have RMSE and biases that are on .average 10% smaller than a stand-alone run (coarse domain alone, without

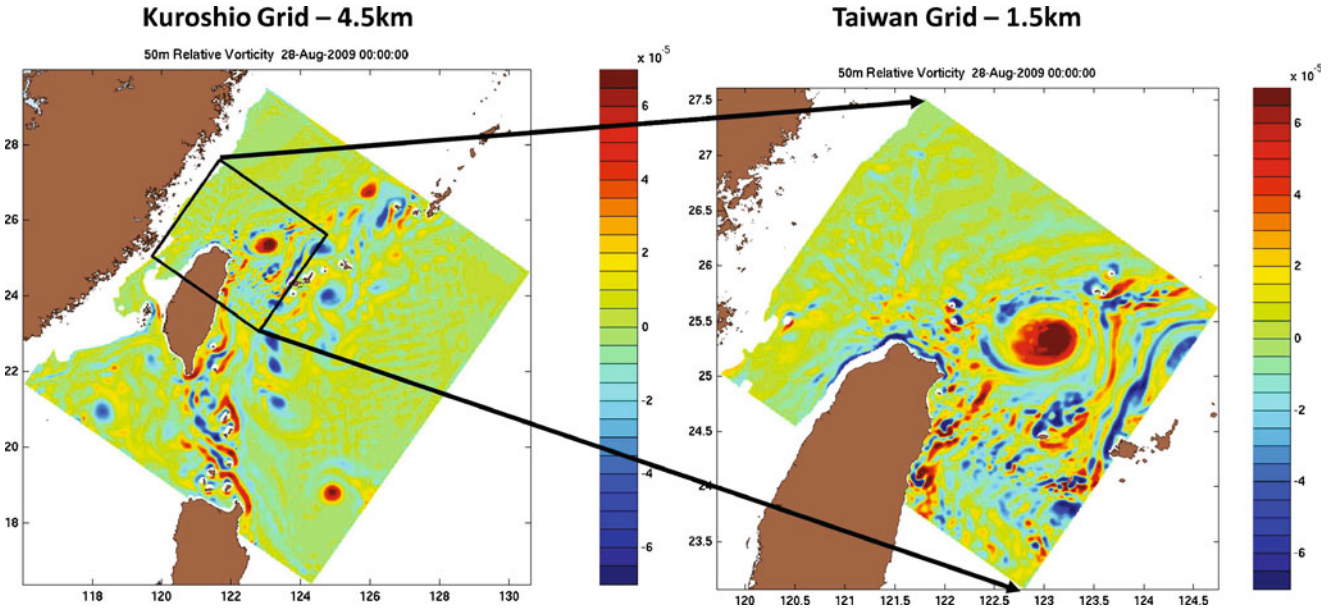

Fig. 17 Relative vorticity at 50 m in the Taiwan/Kuroshio region for 0000Z on 28 Aug. 2009, estimated by our new fully implicit two-way nesting. This is one of the ensemble simulations we have for the period 18 Aug.–10 Sep. 2009. Note the smaller scales maintained in the fine resolution, especially the topographic generation of vorticity as the Kuroshio crosses I-Lan ridge

the nested domain). Such higher-resolutions runs are also useful for internal tide predictions and acoustic simulations (Lermusiaux and Xu 2010).

6 Summary and conclusions

In this manuscript, we derived and applied conservative time-dependent structured discretizations and powerful two-way nesting (embedding) schemes for multiresolution, telescoping domains of primitive equation ocean models with a nonlinear free surface. The resulting schemes are suitable for realistic data-driven multiscale simulations over deep seas to very shallow coastal regions with strong tidal forcing. Starting from the primitive equations with a nonlinear free surface, cast into a control volume formulation, we introduced a second order temporal discretization, including a new time-splitting algorithm compatible with the nonlinear free-surface physics. We then derived a second-order spatial discretization that correctly accounts for the time variations in the finite volumes. It is coded for both spherical and Cartesian horizontal coordinates and for generalized vertical grids. We introduced the concepts of "implicit nesting" in space and time (exchange all of updated fields values as soon as they become available), "explicit nesting" (exchange coarse and fine domain fields only at the start of a discrete time integration or time step), "coarse-to-fine implicit" nesting (the coarse domain feeds the fine domain during its time step, usually boundary conditions, with no feedback from the fine domain), and "fine-to-coarse implicit" nesting (fine domain updates are fed to the coarse domain during its integration). We then derived a fully implicit two-way nesting scheme that is compatible with the time varying discretization of the nonlinear free-surface primitive equations. This scheme allowed us to use different parametrizations for the sub-gridscale physics in each of the nested domains. We compared this scheme to other two-way nesting schemes with modified interdomain couplings. We completed a theoretical truncation error analysis of these schemes, which confirmed and explained our simulation results. We have employed our new discretizations and nesting schemes with data assimilation in several ocean regions with multiscale dynamics around steep shelfbreaks, straits, or other complex geometries. Presently, such simulations were qualitatively and quantitatively evaluated in three regions: the middle Atlantic Bight off the east coast of the USA, the Philippine archipelago, and the Taiwan–Kuroshio region. Of course, we have used the schemes in other regions (e.g., Xu et al. 2008; Kaufman 2010).

Our comparisons of various two-way nesting schemes showed that for nesting with free surfaces, the most accurate schemes are those with strong implicit couplings among grids, especially for the velocity components. We showed selected results for five nesting schemes (including our fully implicit scheme) and a coarse-to-fine implicit scheme. Among the implicit schemes, providing more fine-to-coarse

feedback reduces biases and other discrepancies across grids and produces more stable simulations. For the simulations shown, the scheme with the least amount of fine grid feedback has differences between the barotropic velocity estimates of $O(10)$ cm/s, with the structures of the difference field organized on the (sub) mesoscale. Conversely, the schemes with more feedback keep these discrepancies smaller, $O(1)$ cm/s, intermittent and organized on smaller scales, with some larger differences, 1–4 cm/s along topographic features. Among these schemes, our fully implicit one has the smallest discrepancies, much smaller than 1 cm/s except near steep topography or strong dynamical gradients where differences reach 1 cm/s. This is because the finer grid is needed there to represent this variability. The coarse-to-fine implicit scheme is shown to have much greater discrepancies between the coarse and fine estimates, generally around 1–6 cm/s organized on intermediate scales.

With our theoretical truncation error analysis, we revealed the benefits of additional feedback from the fine-to-coarse domains. The leading error terms for each nesting scheme were determined and studied. We proved that coarse domain estimates which are made up from averages of fine domain estimates retain the truncation error of the fine grid. Even for a second-order scheme with only a 3:1 refinement, this equates to an order of magnitude reduction in the truncation error at these coarse domain points. A corollary of this analysis is that the improvement from the fine grid feedback can be guaranteed only if the feedback algorithm (in our case volume averaging) has at least the same order of accuracy as that of the overall discretization.

In our three realistic simulations, we resolved large domains with multiscale dynamics, including steep bathymetries and strong tidal flows in shallow seas. In each case, we found that without our new discrete PE model, or without nested grids, predictions do not match the ocean data. Specifically, in the middle Atlantic Bight, we compared nested estimates to moored data not assimilated in simulations. Using twin experiments between a 3-km resolution "stand-alone" domain and the same domain with a two-way nested 1-km resolution domain, we show that the addition of the nested subdomain removes large biases and RMSEs, $O(10–15)$ cm/s, in the model velocities when compared to the mooring data. The two-way nesting scheme is found especially needed during tropical storm Ernesto. It is also required for future studies of internal tide propagation. The application to the Philippine Archipelago region is mostly striking by the complexity of the geometry, with multiple islands and passages, and by the multiscale dynamics, from very strong tides in shallow areas to the North Equatorial Current in the Pacific. This complexity required novel schemes, from our multiscale objective analyses for such regions (Agarwal and Lermusiaux 2010) to the present time-dependent spatial discretizations and fully implicit two-way nesting. In this region, we have so far implemented six two-way nested domains, in a telescoping setup, covering four-grid resolutions. Without the present schemes, such multiscale simulations were not possible. In addition, comparisons with independent ADCP data show that nesting substantially improves mean transport estimates through straits. Such results are also significant for basin-scale and climate studies. The simulations for the Taiwan–Kuroshio region focused on uncertainty, and we showed one realization of an ensemble of two-way nested simulations. The region is characterized by strong interactions near the shelfbreaks involving the Kuroshio, mesoscale features, strong atmospheric events, rivers, strong tides, internal tides, and waves. Multiresolution two-way nesting or unstructured grids are required for such dynamics and also for their impacts on acoustics (Lermusiaux and Xu 2010; Xu and Lermusiaux 2010). Quantitative comparisons with independent data confirm increased forecast skill due to nesting.

Ongoing and future research includes a study of boundary conditions (e.g., Blayo and Debreu 2005; Oddo and Pinardi 2008) for nested domains and comparisons with our current radiation-allowing conditions. Finer scales should exit the finer domains freely, but this is challenging since they are not resolved in the coarser domains. We have also applied our schemes to coupled physical–biological dynamics in two-way nested domains in the Philippines archipelago (Lermusiaux et al. 2009b), and dynamical studies in various regions are underway. Also valuable would be to merge our nesting algorithms with refined time stepping in finer domains (Section 3) and compare the results to our present algorithm. Our implicit nesting scheme, after some modifications, could also be applied to other dynamics such as nonhydrostatic regimes as well as to multidynamics, e.g., two-way nesting of nonhydrostatic and hydrostatic models; sediment, wave, and coastal models (Warner et al. 2008); and coastal and storm surge models, with wetting and drying (Bunya et al. 2010; Tanaka et al. 2010). Other research directions are the comparison and merging of our nesting scheme with unstructured grid approaches (e.g., Deleersnijder and Lermusiaux 2008b; Ueckermann 2009; Ueckermann and Lermusiaux 2010; Kleptsova et al. 2010), multiscale feature initialization (Gangopadhyay et al. 2003), and multimodel fusion (Logutov 2007; Rixen et al. 2009). We also intend

to combine our schemes with uncertainty predictions based on dynamically orthogonal equations (Sapsis and Lermusiaux 2009). Even though it is only recently that realistic multiscale ocean modeling is becoming possible, the topic is so important for ocean science and societal applications that opportunities for our novel schemes are expected over a rich spectrum of needs.

Acknowledgements We are thankful to W.G. Leslie for his collaboration in the real-time experiments that formed the test-bed for this research and for his help in preparing the figures for this manuscript. We would like to thank C. Lozano and L. Lanerolle for many fruitful discussions during the development of our system. We thank O. Logutov for his tide research and discussions. We are very thankful to the reviewers for their careful and detailed reviews and their very useful suggestions. For our AWACS-SW06 effort, we thank G. Gawarkiewicz, P. Abbot and T. Duda for their ocean data and M. Taylor and J. Hare for their NMFS survey data. We also thank J. Evans, S. Glenn, and J. Wilkin for their real-time WRF atmospheric fluxes and the FNMOC teams for their own products. For our PhilEx effort, we thank A. Gordon, C. Villanoy, C. Lee, and J. Sprintall for their ocean data; H. Hurlburt and J. Metzger for large-scale boundary conditions; and H. Arango and J. Levin for discussions. We also thank B. Leben and CCAR for providing SSH anomaly data. For our QPE effort, we thank G. Gawarkiewicz, T. Duda, J. Sen, B. Cornuelle, J. Lynch, P. Niiler, L. Centurioni, C. Lee, R.-C. Lien, T. Sanford, L. Mayer, B. Calder, and Y.-T. Lin for data and discussions. We also thank J. Doyle, D. Marble, J. Nachimknin, and J. Cook as well as the FNMOC for providing us with atmospheric fluxes.

Appendix 1: Additional details of the conservative discrete equations

Our horizontal grids are structured Cartesian or spherical grids, either of which can be rotated from the standard geographic orientation (for more details, see Haley 1999, Section 2.2). To obtain the spherical representation of our Eqs. 37–44 and "Appendix 1.2", identify x with longitude and y with latitude. Then multiply all Δx terms by the radius of the earth (R_{earth}) times the cosine of the latitude and all Δy terms by R_{earth}. Our model also includes options for atmospheric forcing, assimilation, tidal forcing, and a river forcing that employs relaxation time constants which can be tuned to reproduce the desired mass, salt, and internal energy (heat) transports. All of these options are compatible with the two-way nesting scheme.

Appendix 1.1: Vertical grid

In Section 2.4, we introduced our vertical discretization, defining first a set of terrain-following depths for the undisturbed mean sea level, $z^{\text{MSL}}_{i,j,k}$. Here we present the details of $z^{\text{MSL}}_{i,j,k}$. We can currently employ five different schemes for defining these vertical levels, two of which are new:

(a) *σ-Coordinates* (Phillips 1957)

$$z^{\text{MSL}}_{i,j,k} = -\sigma_k H_{i,j} \tag{56}$$

where $0 \le \sigma_k \le 1$

(b) *Hybrid coordinates* (Spall and Robinson 1989)

$$z^{\text{MSL}}_{i,j,k} = \begin{cases} \tilde{z}_k & \text{if } k \le k_c \\ -h_c - \sigma_k \left(H_{i,j} - h_c\right) & \text{if } k > k_c \end{cases} \tag{57}$$

where $\tilde{z}_k$ are a set of constant depths and h_c is the sum of the top k_c flat level depths

(c) *Double σ-coordinates* (Lozano et al. 1994)

$$z^{\text{MSL}}_{i,j,k} = \begin{cases} -\sigma_k \tilde{f}_{i,j} & \text{if } k \le k_c \\ -\tilde{f}_{i,j} - (\sigma_k - 1)\left(H_{i,j} - \tilde{f}_{i,j}\right) & \text{if } k > k_c \end{cases} \tag{58}$$

$$\tilde{f}_{i,j} = \frac{z_{c_1} + z_{c_2}}{2} + \frac{z_{c_2} - z_{c_1}}{2} \times \tanh\left[\frac{2\alpha}{z_{c_2} - z_{c_1}}\left(H_{i,j} - h_{ref}\right)\right]$$

$$\sigma_k \in \begin{cases} [0,\,1] \text{ if } k \le k_c \\ [1,\,2] \text{ if } k > k_c \end{cases}$$

where $\tilde{f}_{i,j}$ is the variable interface depth between the upper and lower σ-systems, z_{c_1} and z_{c_2} are the shallow and deep bounds for $\tilde{f}_{i,j}$, h_{ref} is the reference topographic depth at which the hyperbolic tangent term changes sign, and α is a nondimensional slope parameter ($||\nabla \tilde{f}|| \le \alpha ||\nabla H||$).

(d) *Multi-σ-coordinates* This new system is a generalization of the double σ system in which, for P σ-systems, we define $P+1$ nonintersecting interface surfaces. Then the depths are found from

$$z^{\text{MSL}}_{i,j,k} = -\tilde{f}^{p-1}_{i,j} - (\sigma_k - p + 1)\,\tilde{f}^{p}_{i,j} \quad \text{for} \quad k_{p-1} < k \le k_p \tag{59}$$

$$\tilde{f}^{0}_{i,j} = 0\,; \quad \tilde{f}^{P}_{i,j} = H_{i,j}$$

$$\sigma_k \in [(p-1),\ p] \quad \text{for} \quad k_{p-1} < k \le k_p$$

The intermediate interfaces are free to be chosen from many criteria, including key σ_θ surfaces (e.g., top of mean thermocline) or large mean vertical gradients.

(e) *General coordinates* For this new system, we provide a 3D field of level thicknesses, $\Delta z^{\mathrm{MSL}}_{i,j,k}$, under the constraint

$$\sum_{k=1}^{K} \Delta z^{\mathrm{MSL}}_{i,j,k} = H_{i,j}\,.$$

The unperturbed levels are then found from

$$z^{\mathrm{MSL}}_{i,j,k} = \begin{cases} \dfrac{-1}{2}\Delta z^{\mathrm{MSL}}_{i,j,1} & \text{if } k = 1 \\ z^{\mathrm{MSL}}_{i,j,k-1} - \dfrac{1}{2}\left(\Delta z^{\mathrm{MSL}}_{i,j,k-1} + \Delta z^{\mathrm{MSL}}_{i,j,k}\right) & \text{if } k > 1 \end{cases} \tag{60}$$

Note that our new general coordinate scheme contains schemes (a–d) as special cases. Hence, schemes (a–d) are now implemented by specifying $\Delta z^{\mathrm{MSL}}_{i,j,k}$ outside the model, according to their respective rules, and using the resulting $\Delta z^{\mathrm{MSL}}_{i,j,k}$ as input to the general coordinate scheme.

Appendix 1.2: Fluxes through boundaries of computational cells

To complete the conservative spatial discretizations of Section 2.4, we first establish some notation. Values taken at the centers of tracer volumes have integer indices, e.g., $T_{i,j,k}$, while values taken at the centers of velocity volumes have odd-half integer indices, e.g., $\mathbf{u}_{i+\frac{1}{2},j+\frac{1}{2},k}$. In the vertical, values taken at either the centers of tracer or velocity volumes have integer indices while those at the tops or bottoms of the computational volumes have odd-half integer indices, e.g., $\omega_{i,j,k+\frac{1}{2}}$. Using these rules, we define the following averaging and differencing operators:

$$\begin{aligned}
\langle\phi\rangle^{x}_{i,j,k} &= \frac{1}{2}\left(\phi_{i+\frac{1}{2},j,k} + \phi_{i-\frac{1}{2},j,k}\right) \\
\langle\phi\rangle^{y}_{i,j,k} &= \frac{1}{2}\left(\phi_{i,j+\frac{1}{2},k} + \phi_{i,j-\frac{1}{2},k}\right) \\
\langle\phi\rangle^{z}_{i,j,k} &= \frac{1}{2}\left(\phi_{i,j,k+\frac{1}{2}} + \phi_{i,j,k-\frac{1}{2}}\right) \\
\delta^{x}(\phi)_{i,j,k} &= \phi_{i+\frac{1}{2},j,k} - \phi_{i-\frac{1}{2},j,k} \\
\delta^{y}(\phi)_{i,j,k} &= \phi_{i,j+\frac{1}{2},k} - \phi_{i,j-\frac{1}{2},k} \\
\delta^{z}(\phi)_{i,j,k} &= \phi_{i,j,k-\frac{1}{2}} - \phi_{i,j,k+\frac{1}{2}}\,.
\end{aligned}$$

Note that in the above, i and j increase with increasing x and y while k increases with decreasing depth (negative below sea level).

Now we can define the fluxes through the sides of the computational cells. We start with the "flux velocities" evaluated at the centers of the sides. Following Dukowicz and Smith (1994, Appendix E), we define the integrated flows through the "east" and "north" lateral walls of the tracer volumes as

$$\upsilon^{n}_{i+\frac{1}{2},j,k} = \Delta y_j \left\langle \Delta z^n u' + \frac{1}{2}\Delta z^{n-1}\left(U^n + U^{n-1}\right)\right\rangle^{y}_{i+\frac{1}{2},j,k},$$

$$\upsilon^{n}_{i,j+\frac{1}{2},k} = \Delta x_i \left\langle \Delta z^n v' + \frac{1}{2}\Delta z^{n-1}\left(V^n + V^{n-1}\right)\right\rangle^{x}_{i,j+\frac{1}{2},k},$$

while at the velocity boxes, we define the integrated flows through the "east" and "north" lateral walls as

$$\upsilon^{n}_{i+1,j+\frac{1}{2},k} = \Delta y_{j+\frac{1}{2}} \left\langle\left\langle\left\langle \Delta z^n u' + \frac{1}{2}\Delta z^{n-1}\left(U^n + U^{n-1}\right)\right\rangle^{y}\right\rangle^{x}\right\rangle^{y}_{i+1,j+\frac{1}{2},k},$$

$$\upsilon^{n}_{i+\frac{1}{2},j+1,k} = \Delta x_{i+\frac{1}{2}} \left\langle\left\langle\left\langle \Delta z^n v' + \frac{1}{2}\Delta z^{n-1}\left(V^n + V^{n-1}\right)\right\rangle^{x}\right\rangle^{y}\right\rangle^{x}_{i+\frac{1}{2},j+1,k}.$$

These particular spatial averagings are chosen to match the discrete transport constraint (Eq. 65 in "Appendix 1.4"). The new aspect here is the temporal evaluations. The baroclinic velocity components are evaluated at time n. However, the timings for the barotropic components are, again, chosen to match the transport constraint (Eq. 65). Also note that these timings assume $\theta = 1$. To get the corresponding flows through the "west"("south") lateral walls, simply decrement $i(j)$ by one.

To evaluate the fluxes through the tops of the computational volumes, we use the above definitions in Eq. 37. At tracer volumes this yields

$$\delta^{z}(\omega^n)_{i,j,k}\Delta x_i \Delta y_j + \delta^{x}(\upsilon^n)_{i,j,k} + \delta^{y}(\upsilon^n)_{i,j,k} + \frac{\Delta\mathcal{V}^{n}_{i,j,k}}{H_{i,j} + \eta^{n}_{i,j}}\frac{\delta(\eta_{i,j})^{n,n-2}}{\tau} = 0 \tag{61}$$

while at velocity volumes, we get

$$\delta^{z}(\omega^n)_{i+\frac{1}{2},j+\frac{1}{2},k}\Delta x_{i+\frac{1}{2}}\Delta y_{j+\frac{1}{2}} + \delta^{x}(\upsilon^n)_{i+\frac{1}{2},j+\frac{1}{2},k} + \delta^{y}(\upsilon^n)_{i+\frac{1}{2},j+\frac{1}{2},k} + \frac{\Delta\mathcal{V}^{n}_{i+\frac{1}{2},j+\frac{1}{2},k}}{H_{i,j} + \langle\langle\eta^n\rangle^{x}\rangle^{y}_{i+\frac{1}{2},j+\frac{1}{2}}}\frac{\delta\left(\langle\langle\eta\rangle^{x}\rangle^{y}_{i+\frac{1}{2},j+\frac{1}{2}}\right)^{n,n-2}}{\tau} = 0\,. \tag{62}$$

Using these definitions of the fluxes through the boundaries of the computational volumes, we can now simply write the discrete advection operator as

$$\breve{\Gamma}(\phi)^n_{i,j,k} = \delta^x \left(\langle\phi^n\rangle^x \upsilon^n\right)_{i,j,k} + \delta^y \left(\langle\phi^n\rangle^y \nu^n\right)_{i,j,k} + \delta^z \left(\langle\phi^n\rangle^z \omega^n\right)_{i,j,k} \Delta x_i \Delta y_j \,.$$

This formulation is valid for both tracer and velocity computational volumes, with the understanding that for velocity volumes the i, j indices are shifted by one half.

We have evaluated the pressure force term, $-\frac{1}{\rho_0}\int_{\mathcal{S}^n} p^n_h \hat{n}_h \cdot d\boldsymbol{\mathcal{A}}$, both by directly discretizing the integrals of pressure along the cell walls (including the horizontal contributions from the sloping cell tops and bottoms) and by interpolating the pressure to the corresponding velocity depths and evaluating the differential gradient. Both give similar results, but the integral evaluation is conservative and produces less noise in the resulting velocities (especially near sloping bottoms).

Appendix 1.3: Open boundary conditions

For $\mathbf{u}'$, T, S, and η, the application of boundary conditions is straightforward. Our options (see Haley et al. 2009; Lermusiaux 1997) include using values based on data, applying radiation conditions (Orlanski 1976; Spall and Robinson 1989), or, following Perkins et al. (1997), using radiation conditions to correct the provided values. For nested subdomains, we have first used the interpolated values directly or with Perkins et al. (1997) corrections. Some other promising options we have explored with nested subdomains include using the coarse grid values in a narrow buffer zone around the fine domain, which reduces discontinuities. Another important multiscale conservative boundary condition option is to feedback the averages of the fluxes across the boundary walls shared with the large domain (Fig. 3). These include the advective fluxes of momentum and tracers, the pressure force, and the diffusive fluxes of momentum and tracers.

We still need an additional boundary condition for $\overline{\hat{\mathcal{F}}^{n,n-1}}$ since we are unable to directly evaluate Eq. 42 at the boundaries. To derive this boundary condition, we recast Eq. 42 in the form of Eq. 24 and solve for $\overline{\hat{\mathcal{F}}^{n,n-1}}$:

$$\overline{\hat{\mathcal{F}}^{n,n-1}} = \frac{\delta(\mathbf{U})}{\tau} + f\hat{k} \times \mathbf{U}^\alpha + g\nabla\eta^\alpha_{i,j} \,. \tag{63}$$

Now, the right-hand side of Eq. 63 is made up entirely of quantities that can be directly evaluated at the boundary of the velocity grid. For the free surface, we have found that it is more stable to rewrite Eq. 63 in terms of transports:

$$\overline{\hat{\mathcal{F}}^{n,n-1}} = \frac{1}{H+\eta^n}\left\{\frac{\delta\left[(H+\eta)\,\mathbf{U}\right]}{\tau} + f\hat{k}\times\left[(H+\eta)\,\mathbf{U}\right]^\alpha\right\} + g\nabla\eta^\alpha_{i,j} \,. \tag{64}$$

Note: when evaluating Eq. 64, only values at time t_{n+1} are taken from the provided fields (or nesting interpolations). The fields at times t_n and t_{n-1} are both already in memory and in primitive equation balance. They are combined with the t_{n+1} fields to evaluate (Eq. 64).

Applying Perkins et al. (1997) boundary conditions: Following the algorithm of Perkins et al. (1997), corrections to the provided values (and nesting interpolation values) are obtained by applying the Orlanski radiation algorithm to the difference between the PE model values and these provided values and using these differences to correct the boundary values.

For the barotropic transport, however, this is only done for the tangential component to the boundary. The correction to the normal component is derived from the correction obtained for the surface elevation, $\Delta\eta$, and the barotropic continuity equation

$$\frac{\partial\Delta\eta}{\partial t} + \nabla\cdot\left[(H+\eta)\Delta\mathbf{U}\right] = 0 \,.$$

Appendix 1.4: Maintaining the vertically integrated conservation of mass

To see how the free-surface algorithm maintains the vertically integrated conservation of mass, start from Eq. 44, multiply by $\theta(H+\eta^n)$, and take the divergence of the result to get

$$\nabla\cdot\left[\left(H+\eta^n\right)\theta\mathbf{U}^{n+1}\right] = \nabla\cdot\left[\left(H+\eta^n\right)\theta\hat{\mathbf{U}}^{n+1}\right] - \alpha\theta g\tau\nabla\cdot\left[\left(H+\eta^n\right)\nabla\delta\eta\right] + \theta\nabla\cdot\left(\mathbf{u}'^n\big|_\eta\,\frac{\delta\eta}{\tau}\right) \,.$$

Substitute for the right-hand side of the above equation from Eq. 43 and rearrange to obtain

$$\frac{\delta\eta}{\tau} + \frac{1}{2}\nabla\cdot\left[\left(H+\eta^n\right)\left(\theta\mathbf{U}^{n+1} + \mathbf{U}^n + (1-\theta)\mathbf{U}^{n-1}\right)\right] = 0. \tag{65}$$

Equation 65 represents the discrete form of the barotropic continuity enforced by the free-surface algorithm. Imbalances in Eq. 65 produce unrealistic vertical velocities via Eqs. 61 and 62.

However, as illustrated by the above derivation, Eq. 65 is only satisfied to the same degree that Eqs. 43 and 44 are satisfied. This places restrictions on the valid avenues for nesting. For example, we can safely replace the coarse domain estimates of $(H+\eta)\hat{\mathbf{U}}^{n+1}$ with averages from the fine domain without disturbing (Eq. 65). Moving this exchange one step later in the algorithm and trying to average $(H+\eta)[\theta\hat{\mathbf{U}}^{n+1}+\mathbf{U}^{n}+(1-\theta)\mathbf{U}^{n-1}]$ would violate Eq. 43, in the sense that we would not be able to make the last substitution leading to Eq. 65 and hence we would violate Eq. 65. Similar to the bias argument of Section 4.7, these fields and their errors will persist over many time steps, leading to unsustainable vertical velocities.

Appendix 2: Additional details on running our fully implicit two-way nested free-surface PE model

Appendix 2.1: Setting up domains

Appendix 2.1.1: Topography

There are two main issues when defining topographies for nested simulations. The first is that the finer resolution grid can support finer topography scales, including sharper gradients. The bathymetry on the finer grid is not an interpolation of the coarser grid bathymetry, but the coarser grid bathymetry is a coarse-control-volume average of the finer grid bathymetry. The refinement in topographic scales can lead to abrupt artificial discontinuities in the topography where the fine and coarse domains meet. This can be exacerbated by conditioning the topography (Haley and Lozano 2001) to control the hydrostatic consistency condition (Haney 1991). For a given value of the hydrostatic consistency factor (roughly proportional to $\frac{d\mathbf{x}\times\nabla h}{h}$), the finer resolution domain can support steeper bathymetric features (e.g., shelfbreak). To ensure a smooth transition, we define a band of points around the outer edge of a fine domain (e.g., a band from the boundary to 6 points inside the boundary, see also Penven et al. 2006). In this band, we replace the fine grid topography with a blend of the coarse and fine grid topographies:

$$h_{\text{blend}} = \alpha h_{\text{fine}} + (1-\alpha)h_{\text{coarse}} \tag{66}$$

where α varies from zero at the boundary to one at the inner edge of the band (e.g., 6 points).

The second issue comes about from the nesting algorithm itself. As mentioned in Section 3, we force the undisturbed vertical grid , $z^{\text{MSL}}_{i,j,k}$, to satisfy the nesting rules of Eqs. 46 and 48. To ensure that the topographies in nested domains satisfy Eq. 46 and 48 and the blending Eq. 66 ,we usually follows these steps:

1. Apply the nesting constraints on the unconditioned topographies. Starting from the smallest domain, average the fine grid topographies on the successively larger topographies according to Eq. 46. Then starting from the coarsest domain, interpolate the topographies to the boundaries of the successively smaller domains according to Eq. 48.
2. Starting from the largest domain, apply the conditioning. After the largest domain is conditioned, apply the blending Eq. 66 to the second largest. Condition that domain and repeat the blending-conditioning cycle with the successively smaller domains.
3. Reapply the nesting constraints on the conditioned topography. Repeat step 1.

Appendix 2.1.2: Land masking

The first constraint for masking occurs at the boundaries of the finer domains. Considering any two nested domains, we want continuity of the masks across the domain boundary. In other words, a coastline that crosses the boundary of the fine domain should not have a jump or jog at the boundary of the fine domain. Enforcing this consistency, along with boundary constraints on the topography ("Appendix 2.1.1"), enforces consistent estimates of the areas of the lateral boundaries of the fine domain as measured in both the coarse and fine grids.

The second constraint is to have a certain degree of consistency in defining land and sea in the interior of the fine domain. This is a less exact statement because the fine domain supports a more detailed resolution of the land/sea boundary than the coarse domain. Because of the superior resolution, we take the view that the land mask in the interior of the fine domain is "more correct" than the coarse domain mask. Since we use collocated grids, this provides us with a simple algorithm for resetting the coarse mask. For each coarse grid point fully supported by fine grid points, we count how many of the supporting fine grid points are land and how many are sea. If at least one half the fine grid points are sea, the coarse grid point is marked as sea; otherwise, it is masked as land.

Our general procedure is to first define the land mask for the largest (coarsest) domain. Then use that mask to define a crude first guess for the mask in the fine domain. We then reset the interior nodes of the fine mask to better resolve the coasts (leaving a narrow band around the exterior untouched to ensure

continuity through the boundary). If we have more then two domains, we use the current domain to initialize the mask for the next finest domain and repeat. When we finish the mask in the smallest (finest) domain, we use that mask to reset the mask in the next coarser domain, using the above sea/land counting procedure. We then examine the modified mask in that next coarser domain to eliminate any spurious artifacts that may have been created (e.g., a narrow mouth of a bay may have been closed leaving an isolated "lake" that we do not need to maintain). We repeat with the next coarser domain and so on until we get back to the coarsest domain.

Appendix 2.2: Initialization

Our common situation is to estimate the best initial synoptic state from temperature and salinity data (in situ, climatologies, satellite, etc.) but little or no direct velocity data. Our initialization scheme for this situation is described next, focusing mainly on the nesting considerations, first briefly for the rigid-lid procedures and then the extensions for initializations with a free surface.

Appendix 2.2.1: Rigid lid

Our procedures for rigid lid initializations in nested grids are based on (e.g., Haley et al. 2009). Starting from temperature and salinity data, climatologies, etc., we create 3D estimates of temperature and salinity, often using objective analyses (Carter and Robinson 1987; Agarwal and Lermusiaux 2010). From these 3D temperature and salinity estimates, we construct density (Eq. 6) and the hydrostatic pressure (Eq. 9). We then estimate the total velocity using the rigid lid geostrophic relation

$$f\hat{k} \times (\mathbf{u} - \mathbf{u}_{\text{ref}}) = \frac{1}{\rho_0}\nabla p_h = \frac{g}{\rho_0}\int_{Z_{\text{ref}}}^{z} \nabla\rho \, d\zeta$$

where Z_{ref} is a suitably chosen reference level, which can be a "level of no motion", $\mathbf{u}_{\text{ref}}$ is the absolute velocity at that depth, and we have interchanged the horizontal gradient with the vertical integral. When evaluating $\nabla\rho$ at a particular depth, if any of the ρ values used for the gradient would be below topography, we set $\nabla\rho$ to zero. To enforce no penetration of land, we find a stream function, ψ, which satisfies $\nabla^2\psi = \nabla \times \mathbf{u}$ with ψ set to be constant along coasts. From this ψ, we recompute the velocity. We decompose this velocity into barotropic and baroclinic parts (Eq. 8). The baroclinic portion is fine as is, but barotropic velocities at this stage generally do not satisfy the nondivergence of transport. To enforce this, we define a transport stream function $\hat{k} \times \nabla\Psi = H\mathbf{U}$ and fit it to our estimated barotropic velocities via the Poisson equation

$$\nabla \times \left(\frac{\hat{k}}{H} \times \nabla\Psi\right) = \nabla \times \mathbf{U} \, .$$

We derive Dirichlet boundary conditions for the above by first noting that the tangential derivative of Ψ to the boundary equals the normal component of transport, $H\mathbf{U}$, through the boundary. We then integrate this relation around the boundary to obtain the Dirichlet values. For domains with islands, we also need to provide constant values for Ψ along the island coasts. We do this in a two-stage process in which we first compute Ψ assuming all the islands are open ocean. We then use that initial guess to derive constant island values that minimize the relative interisland transports using fast marching methods (Agarwal 2009).

Nesting considerations For nesting the initial temperature, salinity, other tracers, and baroclinic velocity, we can directly enforce some conservation constraints by averaging estimates from finer to coarser grids. For the transport stream function, we go to the additional step of generating the Dirichlet boundary values for the Poisson equation in the fine domain by interpolating the stream function values from the coarse domain. This ensures that the same constant of integration is used for both domains and that the net flows through the fine domain are consistent in both the coarse and fine grids. For island values, if the island is represented in both the coarse and fine domains, the coarse domain value is used. If the island is only in the fine domain, then the procedure of the preceding paragraph is used.

Appendix 2.2.2: Free surface

The starting point for the free-surface initialization scheme is the above rigid-lid initialization. We start by explicitly computing the final, rigid-lid barotropic velocities from

$$\mathbf{U} = \frac{\hat{k}}{H} \times \nabla\Psi \, .$$

We next create an equation for the initial surface elevation. We start from the geostrophic approximation with the full pressure

$$f\hat{k} \times \mathbf{u} = g\nabla\eta + \frac{1}{\rho_0}\nabla p_h \, .$$

Integrating this equation in the vertical from $-H$ to 0 and isolating η results in

$$gH\nabla\eta = f\hat{k} \times H\mathbf{U} - \frac{1}{\rho_0}\int_{-H}^{0} \nabla p_h dz \, . \tag{67}$$

Finally, we take the divergence of Eq. 67 to get

$$\nabla \cdot (gH\nabla\eta) = \nabla \times (fH\mathbf{U}) - \frac{1}{\rho_0}\nabla \cdot \left(\int_{-H}^{0} \nabla p_h dz \right) . \tag{68}$$

To generate Dirichlet boundary values for Eq. 68, we integrate the tangential components of Eq. 67 around the boundary. Because the coastal boundary condition is zero normal derivative, no special action needs to be taken for islands.

Once an initial value for η is constructed, then, by Eq. 35, the initial depths are recomputed. The tracers (temperature, salinity, etc.) and baroclinic velocity are re-interpolated to these new initial depths. Finally, the barotropic velocities from the rigid-lid calculation are rescaled to preserve the transports:

$$\mathbf{U}_{\text{free surface}} = \frac{H}{H+\eta}\mathbf{U}_{\text{rigid lid}} .$$

Nesting considerations These are the same as for the rigid-lid case. The additional detail is that now we also interpolate the coarse grid estimate of η to generate Dirichlet boundary values for solving Eq. 68 in the fine domain.

Appendix 2.3: Tidal forcing

Appendix 2.3.1: Constructing the tidal forcing

When adding tidal forcing to our simulations, our underlying assumption is that our regional domains are small enough so that the tidal forcing through the lateral boundaries completely dominates the local body force effects. To model these lateral forcings, we employ the linearized barotropic tidal model (Logutov 2008; Logutov and Lermusiaux 2008). We use a shallow water spectral model and generate 2D fields for the amplitude and phase of tidal surface elevation and the barotropic tidal velocity. We dynamically balance these barotropic tidal fields with our best available topographic and coastal data along with the best exterior barotropic tidal fields (e.g., Egbert and Erofeeva 2002). Once we have constructed our tidal fields for the desired modes, we can simply evaluate them for any time.

The above procedures can provide tidal fields on different grids than used by our PEs. For example, Logutov (2008) and Logutov and Lermusiaux (2008) are formulated on a C-grid, instead of the B-grid being used here. In particular, this means that tidal fields interpolated from these grids will not, in general, exactly satisfy the same discrete continuity as in our grid. Our experience shows that satisfying the same discrete continuity leads to more robust solutions. To enforce this constraint, we solve the constrained minimization problem

$$\mathcal{J} = \int \{\alpha\eta_1^*\eta_1 + \theta\beta\mathbf{U}_1^* \cdot \mathbf{U}_1 + \phi\beta\mathbf{U}^* \cdot \mathbf{U} + \lambda\Re[i\omega\eta + \nabla\cdot(H\mathbf{U})] + \gamma\Im[i\omega\eta + \nabla\cdot(H\mathbf{U})]\} d\mathcal{V}$$

where η_0, $\mathbf{U}_0$ as the complex tidal surface elevation and barotropic tidal velocity interpolated from the original grid and η_1, $\mathbf{U}_1$ are the additive "correction" complex tidal surface elevation and barotropic tidal velocity that minimize $\mathcal{J}$,

$$\eta = \eta_0 + \eta_1 ; \quad \mathbf{U} = \mathbf{U}_0 + \mathbf{U}_1$$

α and β are the weights (including nondimensionalizing factors), λ and γ are the Lagrange multipliers, the superscript * indicate complex conjugation, $\Re$ and $\Im$ refer to the real and imaginary parts, and θ, ϕ are penalty parameters to inhibit unreasonably large total velocities. Using the calculus of variations, the above minimization is equivalent to solving the following system of equations

$$\omega\eta_1 - \nabla\cdot\left[\frac{H^2}{(\theta+\phi)\omega\beta}\nabla(\alpha\eta_1)\right] = -\omega\eta_0 + i\nabla\cdot\left(\frac{\theta}{\theta+\phi}H\mathbf{U}_0\right)$$

$$\alpha|_{\text{open boundary}} = 0$$

$$\left(i\omega\eta_1 + \sqrt{gH}\frac{\partial\eta_1}{\partial n}\right)\Big|_{\text{open boundary}} = 0$$

$$\mathbf{U}_1 = -\frac{\phi}{\theta+\phi}\mathbf{U}_0 - i\frac{H}{(\theta+\phi)\omega\beta}\nabla(\alpha\eta_1)$$

Note that the radiation boundary condition does not come from the variations but is a useful addition we are free to make after obtaining $\alpha = 0$ from the variations.

Appendix 2.3.2: Applying the tidal forcing

We use the barotropic tides both for initialization and for boundary forcing. For the surface elevation, we simply superimpose the tidal surface elevation with the subtidal elevation estimated in "Appendix 2.2.2". For initialization, this superposition is done over the entire area before the final vertical interpolation of tracers and $\mathbf{u}'$. For lateral forcing, this is done at run time in the PE model at the exterior boundaries (and also along two bands inside these boundaries for Perkins et al.

(1997) boundary conditions). The resulting boundary values are used to generate Dirichlet boundary conditions for Eq. 43. A similar procedure is used for the barotropic velocities with two notable differences. First, the superposition is performed to preserve transport:

$$\left(H+\eta_{\text{superimposed}}\right)\mathbf{U}_{\text{superimposed}} = \left(H+\eta_{\text{subtidal}}\right)\mathbf{U}_{\text{subtidal}} + H\mathbf{U}_{\text{tidal}}\,.$$

Note that the tidal velocity is only multiplied by the undisturbed water depth. This is because we are using a linearized tidal model. The second difference is that the run-time boundary values of the barotropic velocity are used for Eq. 64, not directly applied to the final barotropic velocities.

Nesting considerations For initialization, the process is as for the unnested case. The superpositions described above are done for the initial conditions of each domain. For the lateral forcing, however, the barotropic tidal fields are only applied at the boundaries of the coarsest domain. The reason being that applying the barotropic tidal fields at the boundary of the coarsest domain can produce the full tidal response in the interior, and hence, the barotropic tidal fields are unnecessary for the nested subdomains.

Appendix 2.4: Solving the equation for the surface elevation

Equation 43, with Dirichlet boundary conditions, represents an elliptic system of equations for the surface elevation, η. To numerically solve this system, we use a preconditioned conjugate gradient solver for sparse matrices (e.g., SPARSKIT; Saad 2009). A typical convergence test for such an iterative solver is an integrated measure of the reduction in the norm of the residual over all points. Specifically, if r is the residual of the current solver iteration and r_0 is the residual of the initial guess, the convergence test is

$$\|r\| \le \tau_r \|r_0\| + \tau_a$$

where τ_r is the relative tolerance and τ_a is the absolute tolerance. In practice, we tend to use very small values (10^{-12} and 10^{-25}, respectively) to ensure a tight convergence. Using the results of "Appendix 3", we also supplement this global constraint with the following point-wise constraint:

$$\left|\frac{\partial \delta\eta^k}{\partial x} - \frac{\partial \delta\eta^{k-1}}{\partial x}\right| \le \tau_r^g \left|\frac{\partial \delta\eta^k}{\partial x}\right| + \tau_a\,;$$

$$\left|\frac{\partial \delta\eta^k}{\partial y} - \frac{\partial \delta\eta^{k-1}}{\partial y}\right| \le \tau_r^g \left|\frac{\partial \delta\eta^k}{\partial y}\right| + \tau_a$$

$$\left|\delta\eta^k - \delta\eta^{k-1}\right| \le \tau_r^g \left|\delta\eta^k\right| + \tau_a$$

where the superscript k refers to the iteration number and τ_r^g is the relative tolerance for the gradient test (typically around 10^{-8}). Here we test on both $\delta\eta$ and its gradients to ensure the relative convergence of the barotropic velocities (Eq. 44).

Since we have discretized our equations on the B-grid, both Eq. 43 and, especially, Eq. 68 possess a well-known checkerboard mode in their null spaces (Deleersnijder and Campin 1995; le Roux et al. 2005; Wubs et al. 2006). For realistic geometries, we found that applying a Shapiro filter (Shapiro 1970) to the solution was sufficient to suppress the noise while maintaining the correct physical features. The one case where this approach failed was in creating an initialization for an idealized flow in a periodic channel. The lack of Dirichlet boundary values in that case and corresponding lack of structure they would have imposed allowed the checkerboard mode to suppress all other structures. To control this error, the matrix in Eq. 68 was augmented with a Laplacian filter (Deleersnijder and Campin 1995; Wickett 1999) to prevent the appearance of this mode. Again, this filter was only needed for the idealized periodic channel flow.

Appendix 3: Review of rigid-lid nesting algorithm

One of our first nesting schemes (see Section 4.1) for the nonlinear free-surface version of the PEs was built by analogy with our two-way nesting scheme (e.g., Spall and Holland 1991; Fox and Maskell 1995; Sloan 1996) for the HOPS rigid-lid formulation of the PEs. We have used this scheme in a variety of dynamical situations (Robinson et al. 2002; Leslie et al. 2008; Haley et al. 2009) and we present it next.

We start with the rigid-lid PEs:

$$\nabla\cdot\mathbf{u} + \frac{\partial w}{\partial z} = 0\,,$$

$$\frac{\partial \mathbf{u}'}{\partial t} + f\hat{k}\times\mathbf{u}' = \widehat{\mathcal{F}} - \overline{\widehat{\mathcal{F}}}\,,$$

$$\nabla\times\left(\frac{\hat{k}}{H}\times\nabla\delta\Psi\right) + \alpha_{\text{RL}}\nabla\times\left[f\,\hat{k}\times\left(\frac{\hat{k}}{H}\times\nabla\delta\Psi\right)\right] = \nabla\times\overline{\widehat{\mathcal{F}}}\,,$$

$$p_h = -\int_z^0 \rho g\, d\zeta\,,$$

$$\frac{DT}{Dt} = F^T ,$$

$$\frac{DS}{Dt} = F^S ,$$

$$\rho = \rho(z, T, S)$$

where

$$\mathbf{u} = \mathbf{u}' + \mathbf{U} ; \quad \mathbf{U} = \frac{1}{H}\int_{-H}^{0} \mathbf{u}\, dz ; \quad H\mathbf{U} = \hat{k} \times \nabla\Psi ,$$

$$\widehat{\mathcal{F}} = -\frac{1}{\rho_0}\nabla p_h - \mathbf{\Gamma}(\mathbf{u}) + \mathbf{F} ; \quad \overline{\widehat{\mathcal{F}}} = \frac{1}{H}\int_{-H}^{0} \widehat{\mathcal{F}}\, dz$$

$$\mathbf{\Gamma}(\mathbf{u}) = \begin{pmatrix} \Gamma(u) \\ \Gamma(v) \end{pmatrix} ; \quad \Gamma(\phi) = u\frac{\partial\phi}{\partial x} + v\frac{\partial\phi}{\partial y} + w\frac{\partial\phi}{\partial z}.$$

Here we have introduced the transport stream function, Ψ, to replace the barotropic velocity as a state variable. Using this notation, we write our rigid-lid nesting algorithm as

1. Solve the rigid-lid PEs simultaneously in each domain for $(\mathbf{u}'^{n+1}, T^{n+1}, S^{n+1})$
2. Replace values in the coarse domain at overlap nodes with the following averages from the fine domain values

$$\phi^{n+1}_{i_c,j_c,k} = \frac{1}{\Delta\mathcal{V}_{i_c,j_c}} \sum_{j=j_{fc}-r_h}^{j_{fc}+r_h} \sum_{i=i_{fc}-r_h}^{i_{fc}+r_h} \phi^{n+1}_{i,j,k}\Delta\mathcal{V}_{i,j,k} , \tag{69}$$

$$\left(\nabla\times\overline{\widehat{\mathcal{F}}}\right)_{i_c,j_c} = \frac{1}{\Delta\mathcal{A}_{i_c,j_c}} \sum_{j=j_{fc}-r_h}^{j_{fc}+r_h} \sum_{i=i_{fc}-r_h}^{i_{fc}+r_h} \left(\nabla\times\overline{\widehat{\mathcal{F}}}\right)_{i,j} \Delta\mathcal{A}_{i,j} \tag{70}$$

where r is the ratio of the resolution of the coarse grid to fine grid,

$$\phi = \mathbf{u}', T, S ; \quad \Delta\mathcal{V}_{i,j,k} = \Delta x_{i,j}\Delta y_{i,j}\Delta z_{i,j,k} ;$$

$$\Delta\mathcal{A}_{i,j} = \Delta x_{i,j}\Delta y_{i,j} ; \quad r_h = \lfloor r/2 \rfloor .$$

3. In the coarse domain, solve the rigid-lid barotropic momentum equation for Ψ^{n+1}.
4. Using piece-wise bi-cubic Bessel interpolation, $\mathcal{B}$, replace values in the fine grid boundary with values interpolated from the coarse grid

$$\phi^{n+1}_{i_{fb},j_{fb},k} = \mathcal{B}\left(\phi^{n+1}_{i_c,j_c,k}\right) ,$$

$$\mathbf{u}'^{n+1}_{i_{fb},j_{fb},k}\Delta z_{i_{fb},j_{fb},k} = \mathcal{B}\left(\mathbf{u}'^{n+1}_{i_c,j_c,k}\Delta z_{i_c,j_c,k}\right) ,$$

where

$$\phi = T, S, \Psi .$$

Note that the interpolation of baroclinic velocity is written in terms of transport rather than velocity. This is done to preserve the baroclinicity of $\mathbf{u}'$ in the fine domain.

5. In the fine domain, solve the barotropic momentum equation for Ψ^{n+1}.

We found that this scheme maintains consistency between the estimates on the coarse and fine grids provided that a sufficiently stringent convergence criterion is used when solving for Ψ^{n+1}. Only using a test on the integral of the residuals did not always maintain the consistency. Such a test is global in nature and can give different convergence results in the coarse and fine domains. Instead, we found the best results occurred when we supplemented the residual tests with the following point-wise test:

$$\left|\frac{\frac{\partial\Psi^{k+1}}{\partial x} - \frac{\partial\Psi^{k}}{\partial x}}{\frac{\partial\Psi^{k}}{\partial x}}\right| \le \epsilon ; \quad \left|\frac{\frac{\partial\Psi^{k+1}}{\partial y} - \frac{\partial\Psi^{k}}{\partial y}}{\frac{\partial\Psi^{k}}{\partial y}}\right| \le \epsilon$$

where here the superscripts k refer to the iteration count in the solver and ϵ is the relative tolerance. We test the derivative of Ψ instead of Ψ so that the convergence is on the relative change in the transport, a physically more meaningful quantity than Ψ itself. This test ensures that in both domains the iterative solution proceeds at least until the specified tolerance is reached at every point, thereby ensuring at least that level of consistency between the solutions in the coarse and fine domains.

References

Agarwal A (2009) Statistical field estimation and scale estimation for complex coastal regions and archipelagos. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA

Agarwal A, Lermusiaux PFJ (2010) Statistical field estimation for complex coastal regions and archipelagos. Ocean Model. Sub Judice

Arakawa A, Lamb VR (1977) Computational design of the basic dynamical processes of the UCLA general circulation model. Methods Comput Phys 17:173–265

Arango HG, Ezer T, Shchepetkin AF (2010) Terrain-following ocean modeling system (TOMS). http://www.myroms.org/index.php?model=toms

Barth A, Alvera-Azcárate A, Rixen M, Beckers JM (2005) Two-way nested model of mesoscale circulation features in the Ligurian Sea. Prog Oceanogr 66(2–4):171–189

Barth A, Alvera-Azcárate A, Beckers JM, Rixen M, Vandenbulcke L (2007) Multigrid state vector for data

assimilation in a two-way nested model of the Ligurian Sea. J Mar Syst 65(1–4):41–59
Beckers JM (1991) Application of the GHER 3D general circulation model to the western Mediterranean. J Mar Syst 1(4):315–332
Beckers JM, Brasseur P, Nihoul JCJ (1997) Circulation of the western Mediterranean: from global to regional scales. Deep-Sea Res II 44(3–4):531–549
Berntsen J, Svendsen E (1999) Using the SKAGEX dataset for evaluation of ocean model skills. J Mar Syst 18(4): 313–331
Blayo E, Debreu L (2005) Revisiting open boundary conditions from the point of view of characteristic variables. Ocean Model 9(3):231–252
Bleck R (2002) An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Model 4(1):55–88
Blumberg AF, Mellor GL (1987) A description of a three-dimensional coastal ocean circulation model. In: Heaps NS (ed) Three-dimensional coastal ocean models. American Geophysical Union, Washington, DC, pp 1–16
Bryan K (1969) A numerical method for the study of the circulation of the world ocean. J Comput Phys 4(3):347–376
Bunya S, Dietrich JC, Westerink JJ, Ebersole BA, Smith JM, Atkinson JH, Jensen R, Resio DT, Luettich RA, Dawson C, Cardone VJ, Cox AT, Powell MD, Westerink HJ, Roberts HJ (2010) A high-resolution coupled riverine flow, tide, wind, wind wave, and storm surge model for southern Louisiana and Mississippi. Part I: model development and validation. Mon Weather Rev 138(2):345–377
Cailleau S, Fedorenko V, Barnier B, Blayo E, Debreu L (2008) Comparison of different numerical methods used to handle the open boundary of a regional ocean circulation model of the Bay of Biscay. Ocean Model 25(1–2):1–16
Carnes MR, Fox DN, Rhodes RC, Smedstad OM (1996) Data assimilation in a north Pacific Ocean monitoring and prediction system. In: Malanotte-Rizzoli P (ed) Modern approaches to data assimilation in ocean modeling. Elsevier, Amsterdam, pp 319–345
Carter EF, Robinson AR (1987) Analysis models for the estimation of oceanic fields. J Atmos Ocean Technol 4(1):49–74
Chao Y, Li Z, Farrara J, McWilliams JC, Bellingham J, Capet X, Chavez F, Choi JK, Davis R, Doyle J, Fratantoni DM, Li P, Marchesiello P, Moline MA, Paduan J, Ramp S (2009) Development, implementation and evaluation of a data-assimilative ocean forecasting system off the central California coast. Deep-Sea Res Part 2 Top Stud Oceanogr 56(3–5):100–126
Chapman NR, Lynch JF (2010) Special issue on the 2006 shallow water experiment. IEEE J Oceanic Eng 35(1):1–2
Chassignet EP, Malanotte-Rizzoli P (eds) (2000) Ocean circulation model evaluation experiments for the North Atlantic basin. Dyn Atmos Ocean 32(3–4)
Chassignet EP, Arango H, Dietrich D, Ezer T, Ghil M, Haidvogel DB, Ma CC, Mehra A, Paiva AM, Sirkes Z (2000) DAMEE-NAB: the base experiments. Dyn Atmos Ocean 32(3–4):155–183
Chassignet EP, Smith L, Halliwell G, Bleck R (2003) North Atlantic simulation with the hybrid coordinate ocean model (HYCOM): impact of the vertical coordinate choice, reference density, and thermobaricity. J Phys Oceanogr 33(12):2504–2526
Chassignet EP, Wallcraft A, Halliwell G, Bleck R, McDonald M, Ryan E, Srinivasan A (2009) HYCOM consortium for data assimilative modeling. http://www.hycom.org/
Courant R, Friedrichs K, Lewy H (1928) Über die partiellen differenzengleichungen der mathematischen physik. Math Ann 100(1):32–74
Cushman-Roisin B, Beckers JM (2010) Introduction to geophysical fluid dynamics: physical and numerical aspects. Academic, New York
Debreu L, Blayo E (2008) Two-way embedding algorithms: a review. Ocean Dyn 58(5–6):415–428
Deleersnijder E, Campin JM (1995) On the computation of the barotropic mode of a free-surface world ocean model. Ann Geophys 13(6):675–688
Deleersnijder E, Lermusiaux PFJ (2008a) Multi-scale modeling: nested grid and unstructured mesh approaches. Ocean Dyn 58(5–6):335–336. doi:10.1007/s10236-008-0170-5
Deleersnijder E, Lermusiaux PFJ (eds) (2008b) Multi-scale modeling: nested grid and unstructured mesh approaches. Ocean dynamics, vol 58. Springer, Berlin, pp 335–336
Dukowicz JK, Smith RD (1994) Implicit free-surface method for the Bryan–Cox–Semtner ocean model. J Geophys Res 99(C4):7991–8014
Dutay JC, Bullister JL, Doney SC, Orr JC, Najjar R, Caldeira K, Campin JM, Drange H, Follows M, Gao Y (2002) Evaluation of ocean model ventilation with CFC-11: comparison of 13 global ocean models. Ocean Model 4(2):89–120
Egbert GD, Erofeeva SY (2002) Efficient inverse modeling of barotropic ocean tides. J Atmos Ocean Technol 19(2): 183–204
Estournel C, Auclair F, Lux M, Nguyen C, Marsaleix P (2009) Scale oriented embedded modeling in the north-western Mediterranean in the frame of MFSTEP. Ocean Sci 5(2): 73–90
Ezer T, Arango H, Shchepetkin AF (2002) Developments in terrain-following ocean models: intercomparisons of numerical aspects. Ocean Model 4(3–4):249–267
Ferziger JH, Perić M (1996) Computational methods for fluid dynamics. Springer, Berlin
Fox AD, Maskell SJ (1995) Two-way interactive nesting of primitive equation ocean models with topography. J Phys Oceanogr 25(12):2977–2996
Gangopadhyay A, Robinson AR, Haley PJ, Leslie WG, Lozano CJ, Bisagni JJ, Yu Z (2003) Feature-oriented regional modeling and simulations in the Gulf of Maine and Georges Bank. Cont Shelf Res 23(3–4):317–353
Gawarkiewicz G (2008) Quantifying, predicting, exploiting uncertainty (QPE), 2008 pilot experiment; Aug. 22nd–Sep. 11th, 2008. Tech. rep., Woods Hole Oceanographic Institute, Woods Hole
Gent PR, Bryan FO, Danabasoglu G, Doney SC, Holland WR, Large WG, McWilliams JC (1998) The NCAR climate system model global ocean component. J Climate 11(6): 1287–1306
Ginis I, Richardson RA, Rothstein LM (1998) Design of a multiply nested primitive equation ocean model. Mon Weather Rev 126(4):1054–1079
Gordon AL (2009) Regional cruise intensive observational period 2009. Tech. rep., Lamont Doherty Earth Observatory. http://mseas.mit.edu/Sea_exercises/Philex_IOP09/RIOP09_LEG2_rept.pdf
Griffies SM, Böning C, Bryan FO, Chassignet EP, Gerdes R, Hasumi H, Hirst A, Treguier AM, Webb D (2000) Developments in ocean climate modelling. Ocean Model 2(3–4): 123–192
Griffies SM, Harrison J, Pacanowski RC, Rosati T (2007) Ocean modelling with MOM. CLIVAR Exchanges 12(3): 3–5

Griffies SM, Schmidt M, Herzfeld M (2010) Elements of MOM4p1. GFDL Ocean Group Technical Report 6. Geophysical Fluid Dynamics Laboratory, Princeton. http://www.gfdl.noaa.gov/cms-filesystem-action/model_development/ocean/mom-guide4p1.pdf

Guo X, Hukuda H, Miyazawa Y, Yamagata T (2003) A triply nested ocean model for simulating the Kuroshio—roles of horizontal resolution of JEBAR. J Phys Oceanogr 33(1):146–169

Haidvogel DB, Arango HG, Hedstrom K, Beckmann A, Malanotte-Rizzoli P, Shchepetkin AF (2000) Model evaluation experiments in the North Atlantic Basin: simulations in nonlinear terrain-following coordinates. Dyn Atmos Ocean 32(3–4):239–281

Haley PJ Jr (1999) GRIDS user's guide. Harvard Open Ocean Model Reports 54. Harvard University, Cambridge. http://mseas.mit.edu/archive/HOPS/Grids/grids.ps.gz

Haley PJ Jr, Lozano CJ (2001) COND_TOPO: topography conditioning in MATLAB. http://mseas.mit.edu/archive/HOPS/Cond_Topo/cond_topo.ps.gz

Haley PJ Jr, Lermusiaux PFJ, Leslie WG, Robinson AR (1999) Harvard ocean prediction system (HOPS). http://oceans.deas.harvard.edu/HOPS/

Haley PJ Jr, Lermusiaux PFJ, Robinson AR, Leslie WG, Logoutov O, Cossarini G, Liang XS, Moreno P, Ramp SR, Doyle JD, Bellingham J, Chavez F, Johnston S (2009) Forecasting and reanalysis in the Monterey Bay/California current region for the autonomous ocean sampling network-II experiment. Deep-Sea Res II 56(3–5):127–148. doi:10.1016/j.dsr2.2008.08.010

Haney RL (1991) On the pressure gradient force over steep topography in sigma coordinate ocean models. J Phys Oceanogr 21(4):610–619

Kaufman M (2010) Upwelling dynamics off Monterey Bay: heat flux and temperature variability, and their sensitivities. Senior thesis

Kleptsova O, Stelling GS, Pietrzak JD (2010) On conservative advection schemes in an unstructured grid finite volume three-dimensional hydrostatic model. Ocean Dyn (in press, this special issue)

Lam FP, Haley PJ Jr, Janmaat J, Lermusiaux PFJ, Leslie WG, Schouten MW, te Raa L, Rixen M (2009) At-sea real-time coupled four-dimensional oceanographic and acoustic forecasts during battlespace preparation 2007. J Mar Syst 78:306–320. doi:10.1016/j.jmarsys.2009.01.029

le Roux DY, Sène A, Rostand V, Hanert E (2005) On some spurious mode issues in shallow-water models using a linear algebra approach. Ocean Model 10(1–2):83–94

Leben RR, Born GH, Engebreth BR (2002) Operational altimeter data processing for mesoscale monitoring. Marine Geodesy 25(1–2):3–18

Lermusiaux PFJ (1997) Error subspace data assimilation methods for ocean field estimation: theory, validation and applications. Ph.D. thesis, Harvard University, Cambridge

Lermusiaux PFJ (1999) Data assimilation via error subspace statistical estimation. Part II: middle Atlantic Bight shelfbreak front simulations and ESSE validation. Mont Weather Rev 127(8):1408–1432

Lermusiaux PFJ (2002) On the mapping of multivariate geophysical fields: sensitivity to size, scales and dynamics. J Atmos Ocean Technol 19(10):1602–1637

Lermusiaux PFJ (2006) Uncertainty estimation and prediction for interdisciplinary ocean dynamics. J Comput Phys 217(1):176–199. (special issue of on "Uncertainty quantification". J Glimm, G Karniadakis, eds)

Lermusiaux PFJ (2007) Adaptive sampling, adaptive data assimilation and adaptive modeling. Physica D 230:172–196. (special issue on "Mathematical issues and challenges in data assimilation for geophysical systems: interdisciplinary perspectives". Christopher KRT Jones and Kayo Ide, eds)

Lermusiaux PFJ, Xu J (2010) Coupled ocean-acoustic prediction of transmission loss in a continental shelfbreak region: predictive skill, uncertainty quantification and dynamical sensitivities. IEEE Trans J Ocean Eng. doi:10.1109/JOE.2010.2068611

Lermusiaux PFJ, Anderson DGM, Lozano CJ (2000) On the mapping of multivariate geophysical fields: error and variability subspace estimates. Q J R Meteorol Soc 126(565):1387–1429

Lermusiaux PFJ, Robinson AR, Haley PJ, Leslie WG (2002) Filtering and smoothing via error subspace statistical estimation. In: Advanced interdisciplinary data assimilation. The OCEANS 2002 MTS/IEEE. Holland Publications, pp 795–802

Lermusiaux PFJ, Haley PJ Jr, Leslie WG, Logoutov O, Robinson AR (2006) Autonomous wide aperture cluster for surveillance (AWACS): adaptive sampling and search using predictive models with coupled data assimilation and feedback—Harvard page. http://mseas.mit.edu/archive/AWACS/index_AWACS.html

Lermusiaux PFJ, Haley PJ Jr, Leslie WG, Logutov O (2009a) Philippines Strait Dynamics Experiment—MSEAS Home Page. http://mseas.mit.edu/Research/Straits/

Lermusiaux PFJ, Haley PJ Jr, Leslie WG, Logutov O, Agarwal A, Burton L (2009b) MSEAS PHILEX IOP09 Real-Time Home Page. http://mseas/Sea_exercises/Philex_IOP09/

Lermusiaux PFJ, Haley PJ Jr, Leslie WG, Logutov O, Xu J, Agarwal A (2009c) MSEAS QPE IOP 2009 Real-Time Results. http://mseas/Sea_exercises/QPE_IOP09/index_iop2009.html

Leslie WG, Robinson AR, Haley P, Logoutov O, Moreno P, Lermusiaux PFJ, Coehlo E (2008) Verification and training of real-time forecasting of multi-scale ocean dynamics for maritime rapid environmental assessment. J Mar Syst 69(1–2):3–16

Liang WD, Tang TY, Yang YJ, Ko MT, Chuang WS (2003) Upper-ocean currents around Taiwan. Deep-Sea Res Part 2 50(6–7):1085–1105

Lin Y-T, Newhall AE, Duda TF, Lermusiaux PFJ, Haley PJ (2010) Merging multiple-partial-depth data time series using objective empirical orthogonal function fitting. IEEE Trans J Ocean Eng. doi:10.1109/JOE.2010.2052875

Locarnini RA, Mishonov AV, Antonov JI, Boyer TP, Garcia HE (2006) NOAA NODC WOA05. http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NODC/.WOA05/

Logutov OG (2007) Multi-model fusion and uncertainty estimation for ocean prediction. Ph.D. thesis, Harvard University, Cambridge

Logutov OG (2008) A multigrid methodology for assimilation of measurements into regional tidal models. Ocean Dyn 58(5–6):441–460

Logutov OG, Lermusiaux PFJ (2008) Inverse barotropic tidal estimation for regional ocean applications. Ocean Model 25(1–2):17–34

Lozano CJ, Haley PJ, Arango HG, Sloan NQ, Robinson AR (1994) Harvard coastal/deep water primitive equation model. Harvard Open Ocean Model Reports 52. Harvard University, Cambridge

Lynch DR, Davies AM (eds) (1995) Quantitative skill assessment for coastal ocean models, coastal and estuarine studies, vol 47. American Geophysical Union, Washington, DC

Lynch DR, Naimie CE, Ip JT, Lewis CV, Werner FE, Luettich R, Blanton BO, Quinlan J, McGillicuddy DJ Jr, Ledwell JR, Churchill J, Kosnyrev V, Davis CS, Gallager SM, Ashjian CJ, Lough RG, Manning J, Flagg CN, Hannah CG, Groman RC (2001) Real-time data assimilative modeling on Georges Bank. Oceanography 14(1):65–77

Maderich V, Heling R, Bezhenar R, Brovchenko I, Jenner H, Koshebutskyy V, Kuschan A, Terletska K (2008) Development and application of 3D numerical model THREETOX to the prediction of cooling water transport and mixing in the inland and coastal waters. Hydrol Process 22(7):1000–1013

Marshall J, Hill C, Perelman L, Adcroft A (1997) Hydrostatic, quasi-hydrostatic, and nonhydrostatic ocean modeling. J Geophys Res 102(C3):5733–5752

Martin V (2003) Méthodes de décomposition de domaine de type relaxation dondes pour des équations de locéanographie. Ph.D. thesis, Université Paris, Paris

Martin V (2004) An optimized Schwarz waveform relaxation method for the unsteady convection diffusion equation in two dimensions. Comput Fluids 33(5–6):829–837

Mason E, Molemaker J, Shchepetkin AF, Colas F, McWilliams JC, Sangrà P (2010) Procedures for offline grid nesting in regional ocean models. Ocean Model 35(1–2):1–15. doi:10.1016/j.ocemod.2010.05.007

Mellor GL (2004) Users guide for a three-dimensional, primitive equation, numerical ocean model. Program in atmospheric and oceanic sciences, Princeton University. http://www.aos.princeton.edu/WWWPUBLIC/htdocs.pom/FTPbackup/usersguide0604.pdf

Mooers CNK (ed) (1999) Coastal ocean prediction, coastal and estuarine studies, vol 56. American Geophysical Union, Washington, DC

MSEAS Group (2010) The Multidisciplinary Simulation, Estimation, and Assimilation Systems (http://mseas.mit.edu/. http://mseas.mit.edu/codes). Reports in Ocean Science and Engineering 6, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge

Newhall AE, Gawarkiewicz GG, Lynch JF, Duda TF, McPhee NM, Bahr FB, Marquette CD, Lin YT, Jan S, Wang J, Chen CF, Chiu LYS, Yang YJ, Wei RC, Emerson C, Morton D, Abbot T, Abbot P, Calder B, Mayer L, Lermusiaux PFJ (2010) Acoustics and oceanographic observations collected during the QPE experiment by research vessels OR1, OR2 and OR3 in the east China Sea in the summer of 2009. WHOI Technical Report WHOI-2010-06. Woods Hole Oceanographic Institution, Woods Hole

NOAA (2006) National Geophysical Data Center, NGDC coastal relief model. http://www.ngdc.noaa.gov/mgg/coastal/coastal.html

Oddo P, Pinardi N (2008) Lateral open boundary conditions for nested limited area models: a scale selective approach. Ocean Model 20(2):134–156

Onken R, Robinson AR, Lermusiaux PFJ, Haley PJ Jr, Anderson LA (2003) Data-driven simulations of synoptic circulation and transports in the Tunisia–Sardinia–Sicily region. J Geophys Res 108(C9):8123–8136

Onken R, Álvarez A, Fernández V, Vizoso G, Basterretxea G, Tintoré J, Haley P Jr, Nacini E (2008) A forecast experiment in the Balearic Sea. J Mar Syst 71(1–2):79–98

Orlanski I (1976) A simple boundary condition for unbounded hyperbolic flows. J Comput Phys 21(3):251–269

Penven P, Debreu L, Marchesiello P, McWilliams JC (2006) Evaluation and application of the ROMS 1-way embedding procedure to the central California upwelling system. Ocean Model 12(1–2):157–187. doi:10.1016/j.ocemod.2005.05.002

Perkins AL, Smedstad LF, Blake DW, Heburn GW, Wallcraft AJ (1997) A new nested boundary condition for a primitive equation ocean model. J Geophys Res 102(C2):3483–3500

Phillips NA (1957) A coordinate system having some special advantages for numerical forecasting. J Meteorol 14(2):184–185

Pinardi N, Woods JD (eds) (2002) Ocean forecasting: conceptual basis and applications. Springer, Berlin

Pinardi N, Auclair F, Cesarini C, Demirov E, Fonda-Umani S, Giani M, Montanari G, Oddo P, Tonani M, Zavatarelli M (2002) Toward marine environmental prediction in the Mediterranean Sea coastal areas: a monitoring approach. In: Pinardi N, Woods JD (eds) Ocean forecasting: conceptual basis and applications. Springer, Berlin, pp 339–373

Rixen M, Le Gac JC, Hermand JP, Peggion G, Coelho E (2009) Super-ensemble forecasts and resulting acoustic sensitivities in shallow waters. J Mar Syst 78:S290–S305

Robinson AR (1999) Forecasting and simulating coastal ocean processes and variabilities with the Harvard Ocean Prediction System. In: Mooers CNK (ed) Coastal ocean prediction, coastal and estuarine studies, vol 56. American Geophysical Union, Washington, DC, pp 77–100

Robinson AR, Bellingham JG, Chryssostomidis C, Dickey TD, Levine E, Patrikalakis N, Porter DL, Rothschild BJ, Schmidt H, Sherman K, Holliday DV, Atwood DK, the LOOPS Group (1999) Realtime forecasting of the multidisciplinary coastal ocean with the littoral ocean observing and predicting system (loops). In: Third conference on coastal atmospheric and oceanic prediction and processes. American Meteorological Society, Boston

Robinson AR, Rothschild BJ, Leslie WG, Bisagni JJ, Borges MF, Brown WS, Cai D, Fortier P, Gangopadhyay A, Haley PJ Jr, Kim HS, Lanerolle L, Lermusiaux PFJ, Lozano CJ, Miller MG, Strout G, Sundermeyer MA (2001) The development and demonstration of an advanced fisheries management information system. In: 17th international conference on interactive information and processing systems for meteorology, oceanography and hydrology. American Meteorological Society, Boston, p 6

Robinson AR, Haley PJ, Lermusiaux PFJ, Leslie WG (2002) Predictive skill, predictive capability and predictability in ocean forecasting. In: Proceedings of the "The OCEANS 2002 MTS/IEEE" conference. Holland Publications, pp 787–794

Saad Y (2009) SPARSKIT: a basic tool-kit for sparse matrix computations. http://www-users.cs.umn.edu/~saad/software/SPARSKIT/sparskit.html

Sapsis TP, Lermusiaux PFJ (2009) Dynamically orthogonal field equations for continuous stochastic dynamical systems. Physica D 238:2347–2360. doi:10.1016/j.physd.2009.09.017

Sarkisya AS, Ivanov VF (1971) Joint effect of baroclinicity and bottom relief as an important factor in dynamics of sea currents. Izv Akad Nauk SSSR Fiz Atmos Okeana 7(2):173–188

Semtner AJ Jr (2000) Very high-resolution estimates of global ocean circulation, suitable for carbon-cycle modeling. In: Wigley TML, Schimel DS (eds) The carbon cycle. Cambridge University Press, Cambridge, pp 212–228

Shapiro R (1970) Smoothing, filtering and boundary effects. Rev Geophys Space Phys 8(2):359–387

Shchepetkin AF, McWilliams JC (2005) The regional ocean modeling system (roms): A split-explicit, free-surface, topography following coordinates ocean model. Ocean Model 9(4):347–404

Shen CY, Evans TE (2004) A free-surface hydrodynamic model for density-stratified flow in the weakly to strongly non-hydrostatic regime. J Comput Phys 200(2):695–717

Signell RP, Jenter HL, Blumberg AF (2000) Predicting the physical effects of relocating Boston's sewage outfall. J Estuar Coast Shelf Sci 50(1):59–72

Sloan NQ III (1996) Dynamics of a shelf-slope front: process studies and data-driven simulations in the middle Atlantic Bight. Ph.D. thesis, Harvard University, Cambridge

Smith RD, Dukowicz JK, Malone RC (1992) Parallel ocean general circulation modeling. Physica D 60(1–4):38–61

Smith WHF, Sandwell DT (1997) Global seafloor topography from satellite altimetry and ship depth soundings. Science 277(5334):1957–1962

Spall MA, Holland WR (1991) A nested primitive equation model for oceanic applications. Math Comput Simul 21(2):205–220

Spall MA, Robinson AR (1989) A new open ocean, hybrid coordinate primitive equation model. Math Comput Simul 31(3):241–269

Stammer D, Chassignet EP (2000) Ocean state estimation and prediction in support of oceanographic research. Oceanography 13(2):51–56

de Szoeke RA, Springer SR, Oxilia DM (2000) Orthobaric density: a thermodynamic variable for ocean circulation studies. J Phys Oceanogr 30(11):2830–2852

Tanaka S, Bunya S, Westerink JJ, Dawson C, Luettich RA (2010) Scalability of an unstructured grid continuous Galerkin based hurricane storm surge model. J Sci Comput. doi:10.1007/s10915-010-9402-1

Tessler ZD, Gordon AL, Pratt LJ, Sprintall J (2010) Transport and dynamics of the Panay Sill overflow in the Philippine seas. J Phys Oceanogr. doi:10.1175/2010JPO4395.1

Tian RC, Lermusiaux PFJ, McCarthy JJ, Robinson AR (2004) A generalized prognostic model of marine biogeochemical-ecosystem dynamics: structure, parametrization and adaptive modeling. Harvard Reports in Physical/Interdisciplinary Ocean Science 67, Harvard University, Cambridge

Ueckermann MP (2009) Towards next generation ocean models: novel discontinuous Galerkin schemes for 2D unsteady biogeochemical models. Master's thesis, Massachusetts Institute of Technology, Cambridge

Ueckermann MP, Lermusiaux PFJ (2010) High order schemes for 2D unsteady biogeochemical models. Ocean Dyn (this special issue). doi:10.1007/s10236-010-0351-x

Wallcraft AJ, Kara AB, Hurlburt HE, Rochford PA (2003) The NRL layered global ocean model (NLOM) with an embedded mixed layer submodel: formulation and tuning. J Atmos Ocean Technol 20(11):1601–1615

Warner JC, Sherwood CR, Signell RP, Harris CK, Arango HG (2008) Development of a three-dimensional, regional, coupled wave, current, and sediment-transport model. Comput Geosci 34(10):1284–1306

Webb DJ, de Cuevas BA, Coward A (1998) The first main run of the OCCAM global ocean model. Internal Document 34, Southampton Oceanography Centre. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.17.5811&rep=rep1&type=pdf

WHOI (2006) Shallow water experiment 2006. http://acoustics.whoi.edu/sw06/

Wickett ME (1999) A reduced grid method for a parallel global ocean general circulation model. Ph.D. thesis, University of California, Davis, Davis

Wubs FW, de Niet AC, Dijkstra HA (2006) The performance of implicit ocean models on B- and C-grids. J Comput Phys 211(1):210–228

Xu J, Lermusiaux PFJ (2010) Three dimensional sound propagation in the north Mien Hua Canyon using coupled 4D ocean and 3D acoustics models. JASA Express Lett. Sub Judice

Xu J, Lermusiaux PFJ, Haley PJ Jr, Leslie WG, Logutov OG (2008) Spatial and temporal variations in acoustic propagation during the PLUSNet07 exercise in Dabob Bay. In: Proceedings of meetings on acoustics (POMA), vol 4. 155th Meeting Acoustical Society of America. doi:10.1121/1.2988093

Ocean Dynamics (2010) 60:1539–1558
DOI 10.1007/s10236-010-0356-5

A finite element sea ice model of the Canadian Arctic Archipelago

Arjen D. Terwisscha van Scheltinga · Paul G. Myers · Julie D. Pietrzak

Received: 29 April 2010 / Accepted: 26 October 2010 / Published online: 11 November 2010

Abstract The Canadian Arctic Archipelago (CAA) is a complex area formed by narrow straits and islands in the Arctic. It is an important pathway for freshwater and sea-ice transport from the Arctic Ocean to the Labrador Sea and ultimately to the Atlantic Ocean. The narrow straits are often crudely represented in coupled sea-ice–ocean models, leading to a misrepresentation of transports through these straits. Unstructured meshes are an alternative in modelling this complex region, since they are able to capture the complex geometry of the CAA. This provides higher resolution in the flow field and allows for more accurate transports (but not necessarily better modelling). In this paper, a finite element sea-ice model of the Arctic region is described and used to estimate the sea-ice fluxes through the CAA. The model is a dynamic–thermodynamic sea-ice model with elastic–viscous–plastic rheology and is coupled to a slab ocean, where the temperature and salinity are restored to climatology, with no velocities and surface elevation. The model is spun-up from 1973 to 1978 with NCEP/NARR reanalysis data. From 1979 to 2007, the model is forced by NCEP/DoE reanalysis data. The large scale sea-ice characteristics show good agreement with observations. The total sea-ice area agrees very well with observations and shows a sensitivity to the Arctic oscillation (AO). For 1998–2002, we find estimates for the sea-ice volume and area fluxes through Admunsen Gulf, McClure Strait and the Queen Elizabeth Islands that compare well with observation and are slightly better than estimates from other models. For Nares Strait, we find that the fluxes are much lower than observed, due to the missing effect of topographic steering on the atmospheric forcing fields. The 1979–2007 fluxes show large seasonal and interannual variability driven primarily by variability in the ice velocity field and a sensitivity to the AO and other large-scale atmospheric variability, which suggests that accurate atmospheric forcing might be crucial to modelling the CAA.

Keywords Ocean modelling · Unstructured meshes · Sea ice · Arctic Ocean · Canadian Arctic Archipelago · Freshwater flux

Responsible Editor: Pierre Lermusiaux

A. D. Terwisscha van Scheltinga (✉) · P. G. Myers
Department of Atmospheric and Earth Sciences,
University of Alberta, Edmonton, AB, Canada
e-mail: terwissc@ualberta.ca

J. D. Pietrzak
Faculteit CTIG, Technische Universiteit Delft,
Delft, The Netherlands

1 Introduction

During the last decade, rapid changes have been observed in the extent of sea ice in the Arctic (Stroeve et al. 2005, 2007). In addition, rapid sea-ice thinning (Rothrock et al. 1999; Wadhams and Davis 2000) and loss of perennial sea ice (Comiso 2002) are cause for concern. Sea ice is a key component of the high-latitude climate (Serreze et al. 2007), and due to the Arctic amplification of global climate changes (Flato and Boer 2001; Johannessen 2004), it serves as an excellent indicator of global change. The high sensitivity of the Arctic sea ice to climate changes has been attributed to the ice-albedo feedback (Ebert and Curry 1993). Due to brine rejection during sea ice formation, sea

ice also plays an important role in deep convection in high-latitude oceans. Therefore, sea ice also has an important role to play in the global thermohaline circulation (e.g. Aagaard and Carmack 1989). Recent comparisons between the observed climate and that simulated by coupled general circulation models indicate that our understanding of sea ice and its implementation in climate models needs to be improved (e.g. Zhang and Walsh 2006; Lefebvre and Goosse 2008).

The sea-ice circulation in the Arctic Ocean consists of two main features: the anti-cyclonic Beaufort Gyre in the western Arctic, where sea ice can be trapped for over a decade (Hibler 1980), and the large Transpolar Drift that transports sea ice from the region north of Siberia to the Atlantic Ocean through Fram Strait (Kwok et al. 2004), the main exit for sea ice from the Arctic Ocean. These large-scale flows are mainly driven by the wind and ocean currents. The ocean currents in the Arctic Ocean are influenced by the thermohaline circulation, transporting surface heat to the Arctic Ocean from the Atlantic Ocean, and a sea level difference between the Atlantic and Pacific which sets up a net transport into the Arctic through Bering Strait.

In addition, the sea ice is sensitive to large-scale interannual variability in the atmosphere. Rigor et al. (2002) found that summertime sea-ice concentration correlates well with the Arctic oscillation (AO) index of the previous winter. This response of the sea ice to this long-term atmospheric variability is due to the dynamical influence of the AO on the wintertime sea-ice thickness, due to the imprint of the AO on the surface air temperature anomalies (Rigor et al. 2000). This modifies sea-ice advection during years with a positive AO index: a decrease in ice advection from the western Arctic into the eastern Arctic, an increase in ice advection away from the coast of the East Siberian and Laptev Seas and a small increase of ice transport out of the Arctic through Fram Strait (Rigor et al. 2002). These changes suggest that part of the observed thinning of sea ice in the Arctic Ocean can be attributed to large-scale interannual atmospheric variability.

The complex area of islands and straits in the Canadian Arctic Archipelago (CAA; see Fig. 1a) is an important pathway of sea ice and watermass transport from the Arctic to the North Atlantic (Kwok et al. 2004; Kwok 2006). The narrow straits in the CAA have a large impact on the ice circulation. Due to high concentrations and low temperatures, ice bridges can form, which prevent ice flow for several months. Later, ice can suddenly be released. This leads to complex dynamical behavior, and as a consequence, strong seasonal and interannual variability has been observed in the ice characteristics (Melling 2002; Kwok 2006).

A net flow from the Arctic Ocean to the CAA is observed (Kwok 2006), as are ice transports from the Arctic and the CAA to Baffin Bay (Kwok et al. 2004; Serreze et al. 2006; Agnew et al. 2008). In addition, a recirculation of the sea ice to the Arctic Ocean from the Queen Elizabeth Islands (QEI) is observed (Kwok 2006). In most of the northern gates of the CAA, sea ice is landfast and acts as a buffer for Arctic Ocean–CAA sea-ice exchange (Kwok 2006). Therefore, most of the ice exported to Baffin Bay from the CAA consists of ice formed within the CAA.

The flux of freshwater through the CAA plays an important role in the local freshwater budget of the Arctic and North Atlantic (Dickson et al. 2007) and may play an important role in convective processes in the Labrador Sea (Goosse et al. 1997). The contribution of sea ice on freshwater fluxes is not clearly understood. The impact of sea ice on the polynya formation in Northern Baffin Bay and the stratification in Baffin Bay also remains unclear but might be responsible for many sea-ice models over estimating the summer sea-ice extent in Baffin Bay (e.g. Timmermann et al. 2005; Vancoppenolle et al. 2009). Furthermore, observations in the region are difficult to obtain and therefore scarce, especially ice thickness measurements (Comiso and Steffen 2008).

One of the major issues in modelling the CAA is the problem of representing the complex geometry of its many islands and straits. Unstructured meshes have the advantage of accurately representing these complex geometries and allow for spatial variation in the mesh resolution (e.g. Adcroft and Marshall 1998; Legrand et al. 2006). This allows for more accurate transports, though not necessarily better modelling. In the past decade, many studies have investigated the use of unstructured meshes in ocean (e.g. LeRoux et al. 2000; Danilov et al. 2004; Pietrzak et al. 2005; White et al. 2008) and sea-ice modelling (e.g. Schulkes et al. 1998; Yakolev 2003; Wang and Ikeda 2004; Timmermann et al. 2009). Recently, a study into data assimilation of ice drift data using an unstructured mesh model was performed by Rollenhagen et al. (2009). These studies have been focused on the circulation in the Arctic or Antarctic. Only a few studies have used unstructured meshes to model the CAA. The mean circulation in the region was diagnosed by Kliem and Greenberg (2003). A study into the past (1950–2005) and future (2045–2060) circulation in the CAA was performed by the structured model of Sou and Flato (2009), while Lietaer et al. (2008) investigated the effect of resolution in the narrow straits in the CAA using an unstructured mesh model. The finite-volume unstructured model FVCOM (Chen and Beardsley 2003; Chen et al. 2006) has

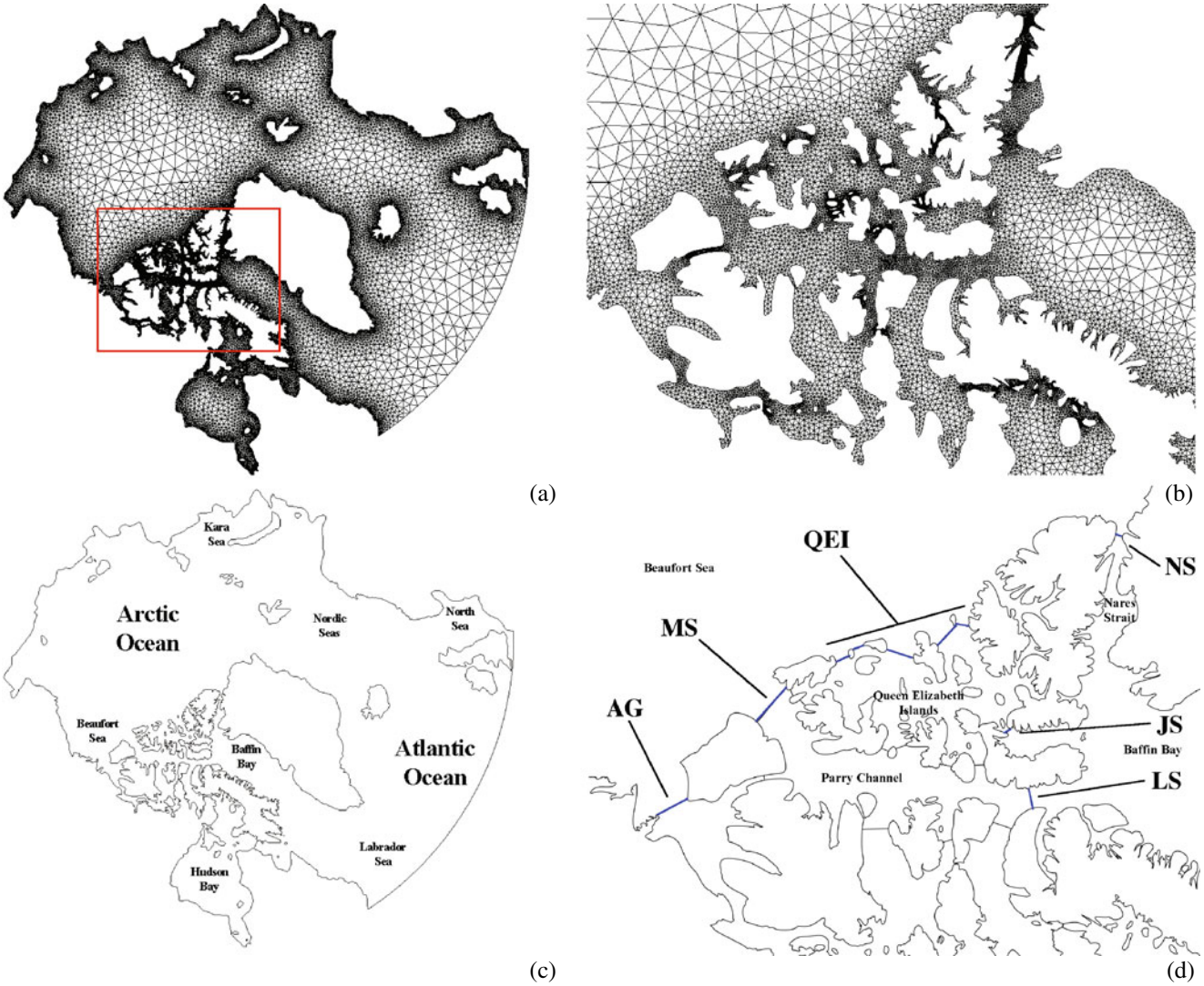

Fig. 1 Domain and mesh used: **a** Arctic Ocean, *red box* indicates area of shown in **b** and **d**; **b** zoom in on the Canadian Arctic Archipelago; **c** location of the bodies of water discussed in this paper and **d** location of the transects where the sea-ice volume and area fluxes are calculated: Admunsen Gulf (*AG*), McClure Straits (*MS*), the Queen Elizabeth Islands (*QEI*), Nares Strait (*NS*), Lancaster Sound (*LS*) and Jones Sound (*JS*)

recently been configured for the Arctic region and used for tidal simulation (Chen et al. 2009).

One approach to unstructured modelling is the finite-element method (FEM). It is an alternative to finite-difference methods commonly used in ocean modelling, since FEM easily provides conservation of energy, natural treatment of boundaries and the flexibility of spatial discretization (Le Provost 1986). However, mass is not conserved unless a discontinuous Galerkin or non-conforming formulation is used. Since the 1970s, finite elements have been used for modelling sea ice. Crack propagation in sea ice was investigated by (Mukherji 1973), within the framework of the Arctic Ice Dynamics Joint Experiment project. Other studies followed (e.g. Becker 1976; Sodhi and Hiber 1980; Thomson et al. 1988). Recently, two FEM models were introduced: a sea ice model coupled to a simple slab ocean (Lietaer et al. 2008) and FESOM, a coupled sea-ice–ocean model (Timmermann et al. 2009). For this study, we use a version of the latter.

Our long-term objective is to investigate the dynamics behind the recent sea-ice decline, freshwater fluxes (both solid and liquid) from the Arctic Ocean through the CAA to the Labrador Sea and their influence on convection in the Labrador Sea. We also we wish to study questions of modelling of sea-ice dynamics, thermodynamics and interactions with the atmosphere and ocean. As a step in this direction,

the approach we adopt in this paper is to start with a relatively simple model and then in later studies, to increase the model complexity. Therefore, following Lietaer et al. (2008), instead of running a fully coupled (atmosphere–)sea-ice–ocean model, we run only the sea-ice model with a slab ocean. Lietaer et al. (2008) suggest that this approximation is sufficient for studying the sea-ice cover and freshwater flux through the CAA. This slab ocean is a very simplified version of the ocean component of FESOM, rewritten specifically for this study. The model used here differs from the study by Lietaer et al. (2008), by using a different element formulation and different atmospheric forcing data. Furthermore, our resolution of the mesh is much more refined than the mesh used by Lietaer et al. (2008), allowing us to more accurately represent fluxes through the CAA. The sea-ice dynamics use the Hibler (1979) dynamics, which is only valid for larger spatial scales. The modelling of the sea ice does therefore not necessarily improve with higher resolution, but it does result in more accurate ice transports through the CAA.

This model allows us to carry out detailed studies into the different processes taking place in the Arctic, as well as to investigate the limitations of our model formulations and implementations. In this paper, we focus on (a) validation of this model with respect to observations and other studies by comparing large-scale sea-ice characteristics and (b) the estimation of sea-ice volume and area fluxes through the CAA. The model results will be compared to the results of Lietaer et al. (2008) and Sou and Flato (2009). The model used for the latter study differs from the model used here by using a finite-difference approach and has coarse resolution. In addition, it has a full 3D ocean. In comparison with both, our analysis also includes an investigation to the influence of the Arctic oscillation on the sea-ice cover and the time-varying properties of the volume ice flux through the CAA. The model presented here is the first step in this approach and serves as a baseline against which we can evaluate future improvements. It also allows us to make comparisons against more complex models.

This paper is organized as follows: In Section 2, we briefly describe the model, its numerical implementation, initialization, and atmospheric forcing, followed by a validation of the model in Section 3. In Section 4, the estimates of the ice volume and area fluxes are presented and compared to those found by other studies. In addition, we present time series of the volume flux for the period 1979–2007 in selected straits and discuss the sources of their variability and connection to the Arctic oscillation. We end with conclusions and a brief discussion in Section 5.

2 Model description

The model we use is FESOM (Timmermann et al. 2009), which is a finite element coupled sea-ice–ocean model, consisting of an ocean component (FEOM; Danilov et al. 2004) and a sea-ice component. Since large parts of FESOM have already been described in Timmermann et al. (2009), we only give a brief overview of the model and concentrate on modifications to the model made for this study.

2.1 Ocean component

In FESOM, the ocean component (FEOM) is based on a model introduced by Danilov et al. (2004). It is a hydrostatic ocean circulation model that solves the primitive equations on a sphere. Here, however, we consider the ocean to be a simple slab ocean with a constant mixed layer depth H_{ml} of 30 m, no velocities and a flat surface. The ocean temperature T_w and salinity S_w are computed by the following equations:

$$\rho_w C_{pw} H_{\mathrm{ml}} \frac{\partial T_w}{\partial t} = Q_H + R_T \tag{1}$$

$$H_{\mathrm{ml}} \frac{\partial S_w}{\partial t} = Q_S + R_S \tag{2}$$

where ρ_w and C_{pw} are the density and heat capacity of sea water, Q_H and Q_S are the heat and freshwater flux and R_T and R_S are restoring terms. These terms restore the ocean temperature and salinity towards a climatological temperature and salinity of the mixed layer (T_{ml} and S_{ml}):

$$R_T = \rho_w C_{pw} H_{\mathrm{ml}} \gamma_t (T_{\mathrm{ml}} - T_w) \tag{3}$$

$$R_S = H_{\mathrm{ml}} \gamma_t (S_{\mathrm{ml}} - S_w) \tag{4}$$

where γ_t is a relaxation constant. As in Lietaer et al. (2008), we chose a value of $6 \times 10^{-8} s^{-1}$, corresponding to a relaxation timescale of 190 days. The values for T_{ml} and S_{ml} are taken from the Polar Science Center Hydrographic Climatology (PHC3, updated from Steele et al. 2001). We use the monthly climatology, linearly interpolated to get daily values. Below the sea-ice T_w is set to the freezing temperature of seawater T_f as calculated from S_w using the nonlinear equation of Millero (1978). The approach chosen here follows Lietaer et al. (2008), with the distinction that the ocean layer is characterized by both T_w and S_w, instead of only T_w.

2.2 Ice model

For the sea-ice model, a dynamic–thermodynamic model is used. The dynamic part follows Hibler (1979) and includes a momentum equation and continuity equation for the sea-ice thickness, sea-ice concentration and snow thickness. Since we have a slab ocean, the ocean–ice drag and ocean elevation terms in the momentum equation are neglected, and the only momentum input is through drag by the winds. The model offers both viscous–plastic rheology (Hibler 1979) and elastic–viscous–plastic rheology (Hunke and Dukovicz 1997), but only the latter is used here with an internal time step of 96 s. A prognostic model for the snow is used (Owens and Lemke 1990) and accounts for snow–ice conversion due to flooding (Leppäranta 1983; Fischer 1995). The zero-layer approach of Semtner (1976) is used for diffusion of heat in thick sea ice, which assumes a linear temperature profile in the ice neglecting heat storage. The Parkinson and Washington (2007) thermodynamics are used.

We use the parameter values for the sea ice as given by Timmermann et al. (2009), except for the numerical diffusion coefficient K. The solutions are sensitive to the numerical diffusion. A lower value of $K = 1{,}000\ \mathrm{m/s^2}$ was chosen here to reduce the necessary smoothing of the solution (Timmermann et al. 2009). This sets the numerical diffusion as small as possible, while the model remains numerically stable.

2.3 Numerical implementation

The equations for the sea ice are discretized on an unstructured triangular mesh using a finite-element method. The mesh and domain are shown in Fig. 1a, with a zoom in on the CAA (Fig. 1b). The spatial extent of the domain differs from those used by Lietaer et al. (2008) and Rollenhagen et al. (2009). The former covers the entire Arctic north of 50° N, while the latter only covers the Arctic ocean, Nordic Seas and the Northern Atlantic Ocean. Like Lietaer et al. (2008), our domain covers the Arctic north of 50° N but excludes the Bering Sea and the Baltic Sea. Following Rollenhagen et al. (2009), both the boundary through the Bering Strait and the North Atlantic are closed. Our setup also differs from the domain by Sou and Flato (2009), who used a regional model of the CAA, forced at the lateral boundaries by data provided by a model for the whole Arctic.

The mesh was produced following the approach of Lambrechts et al. (2008) and using the GMSH mesh generator (Geuzaine and Remacle 2009). A sufficient resolution in the narrow straits in the CAA is obtained by taking into account the width of the channels when determining the size of the triangular elements. More specifically, the size of the elements is determined by dividing the width of the channel by the necessary number of elements across the strait. This procedure allows us to refine the background mesh, where the size of the elements is determined solely by the proximity to the nearest shoreline (see Lambrechts et al. 2008). Note that since we have a slab ocean without bathymetry, no refinement near steep bathymetry or shallow seas is necessary. The high resolution in the CAA allows us to more accurately capture the fluxes through its straits. The coastline data used by the mesh generator was taken from the Global Self-consistent Hierarchical High-resolution Shoreline database (Wessel and Smith 1996), which has a resolution of 50 m. This should be sufficient to accurately capture the straits in the CAA.

The resolution in the open Arctic Ocean far from the coast is 200 km, smoothly decreasing to 10 km at the coast. Due to the resolution in the central Arctic, some short-term and small-scale processes might be neglected. In analyzing the results, most attention is paid to the climatological monthly means, which are dominated by large-scale spatial variations in the Arctic Ocean. Hence, this coarse resolution is sufficient for the Arctic Ocean. In the CAA, the resolution varies with the width of the straits, such that each cross section is discretized by at least eight elements. This results in some straits having a resolution of a few hundred meters, which is sufficient for estimation of fluxes. The resolution used here in the CAA is higher than for Lietaer et al. (2008) and Rollenhagen et al. (2009) and the structured model of Sou and Flato (2009).

The ice dynamics used here are described by Hibler (1979) and assume the ice to be a 2D plastic continuum. Strictly, this assumption is only valid for scales of 10 km and larger. Since we use a resolution much smaller than this, one can question if this is still the appropriate dynamics. At the resolution we use in the CAA, small-scale processes are dominated by individual ice floes and the sea ice is more granular. Recently, some progress has been made to account for smaller-scale processes in the dynamics, for example, by describing the ice as a granular material (Tremblay and Mysak 1997) or as a collection of diamond-shaped ice floes (Wilchinsky and Feltham 2006). To our knowledge, these more appropriate formulations have only been applied in idealized cases or small regions. For the application described here, there is not yet an alternative available to the Hibler-type dynamics. Though the modelling does not necessarily improve with resolution, the higher resolution here does allow us to obtain more accurate ice transports.

Following the standard FEM approach, piecewise polynomial approximation of the variables have to be chosen. Here the $P_1 - P_1$ element is used to obtain a system of equations, and backward Euler time stepping is used for the sea-ice equations. This scheme has low dispersion and conserves the total tracer but is not strictly monotonic so that Laplacian diffusion with a diffusivity coefficient (here $K = 1{,}000$ m/s^2) is necessary to control the smoothness of the solution (Timmermann et al. 2009). No stabilization is used in the sea ice model. The scheme differs from Lietaer et al. (2008), where a non-conforming element ($P_1^{NC} - P_1$) was used and advection was computed via a first-order finite-volume upwind-weighted scheme. Though numerically diffusive, it is monotonic and conservative, and no Laplacian diffusion is necessary to control the smoothness of the solution. This will lead to (small) differences in model behavior.

A splitting technique is used for the continuity equations for ice thickness, ice concentration and snow thickness. First, the continuity equations are solved with thermodynamic sources and sinks on the right-hand side set to zero. This is followed by an update of these variables using a vertical thermodynamic equation at each node. In contrast, Lietaer et al. (2008) use a Newton–Rhapson procedure to solve the continuity equations. Here domain decomposition is used to run the model in parallel mode, and eight processors are used. Solution of the system of equations is obtained using the PETSc solver (Balay et al. 1997, 2008, 2009), specifically the Bi-CGSTAB methods with a block Jacobi pre-conditioner.

The wide range of element sizes, ranging from 100 to 200 m to more than 100 km, might lead to an ill-conditioned stiffness matrix and thus to an ill-conditioned problem. We found that the small amount of numerical diffusion combined with strict criteria to ensure the solutions have converged, does circumvent ill-conditioning. In addition, the problem is 2D and the time-step has been chosen much smaller than time scales of the variability seen, such that any ill-conditioning is bound to be small. Though the $P_1 - P_1$ element might not be the best choice for the element from a numerical point of view, it's implementation is straightforward. In addition, the code is fast and efficient. A typical 30-year run takes 1.5–2 days on eight processors.

2.4 Initialization and atmospheric forcing

The sea ice and slab ocean are initialized on January 1st, 1973. For the slab ocean climatological values are used. Following Timmermann et al. (2005, 2009), sea ice in regions where the ocean temperature is below -1°C an initial sea-ice thickness of 3 m, initial snow thickness of 10 cm and a concentration of 0.9 is prescribed.

The model is then spun-up for 6 years (to December 31st, 1978) using NCEP/NCAR reanalysis atmospheric forcing fields (Kalnay et al. 1996). From 1979 to 2007, the model is forced with the NCEP/DoE reanalysis II (Kanamitsu et al. 2002). The analysis of the model will be limited to this period. The atmospheric fields used are the wind at 10 m, the air temperature and specific humidity at 2 m, the atmospheric pressure, precipitation and evaporation at the surface and the total cloudiness. These fields were interpolated to the mesh using a bicubic interpolation. Interpolation is required because these data are on a structured grid and for most of the domain used here it has a much coarser resolution than used in our mesh. Ideally, higher-resolution data would be preferable. However, those data are not available. Both Lietaer et al. (2008) and Sou and Flato (2009) use the NCEP/NCAR reanalysis to force their model. Here we have chosen to use the NCEP/DoE reanalysis II because it has improved physics and assimilation compared to NCEP/NCAR (Kanamitsu et al. 2002).

3 Validation

Before focusing our attention on the CAA and the fluxes through its straits, we first look at the large-scale ice concentration, ice thickness and mean ice drift climatology in order to validate the model.

3.1 Sea-ice concentration

Figures 2 and 3 compare the 1979–2007 mean ice concentration fields derived from SMMR (1979–August 1987) and SSM/I (September 1987–2007) data sets (Comiso 1999) with our model results. The SMMR and SSM/I data have been interpolated on our mesh, where observational data are available, i.e. in a small region around the North Pole that cannot be observed by the satellites, the so-called pole hole. For SMMR this hole is north of 84° N, while for SSM/I it is slightly it smaller: 87° N. To account for the different sizes of the pole hole, in the remainder of this section, we use only data south of 84° N, i.e. SSM/I data and model data in the region of the SMMR pole hole are not used. To indicate this in Figs. 2a, b and 3a, b, the pole hole has been masked out.

The observational data show that the ice concentrations are high for March, with values over 90% for most of the domain (Fig. 2). The simulated mean field agrees

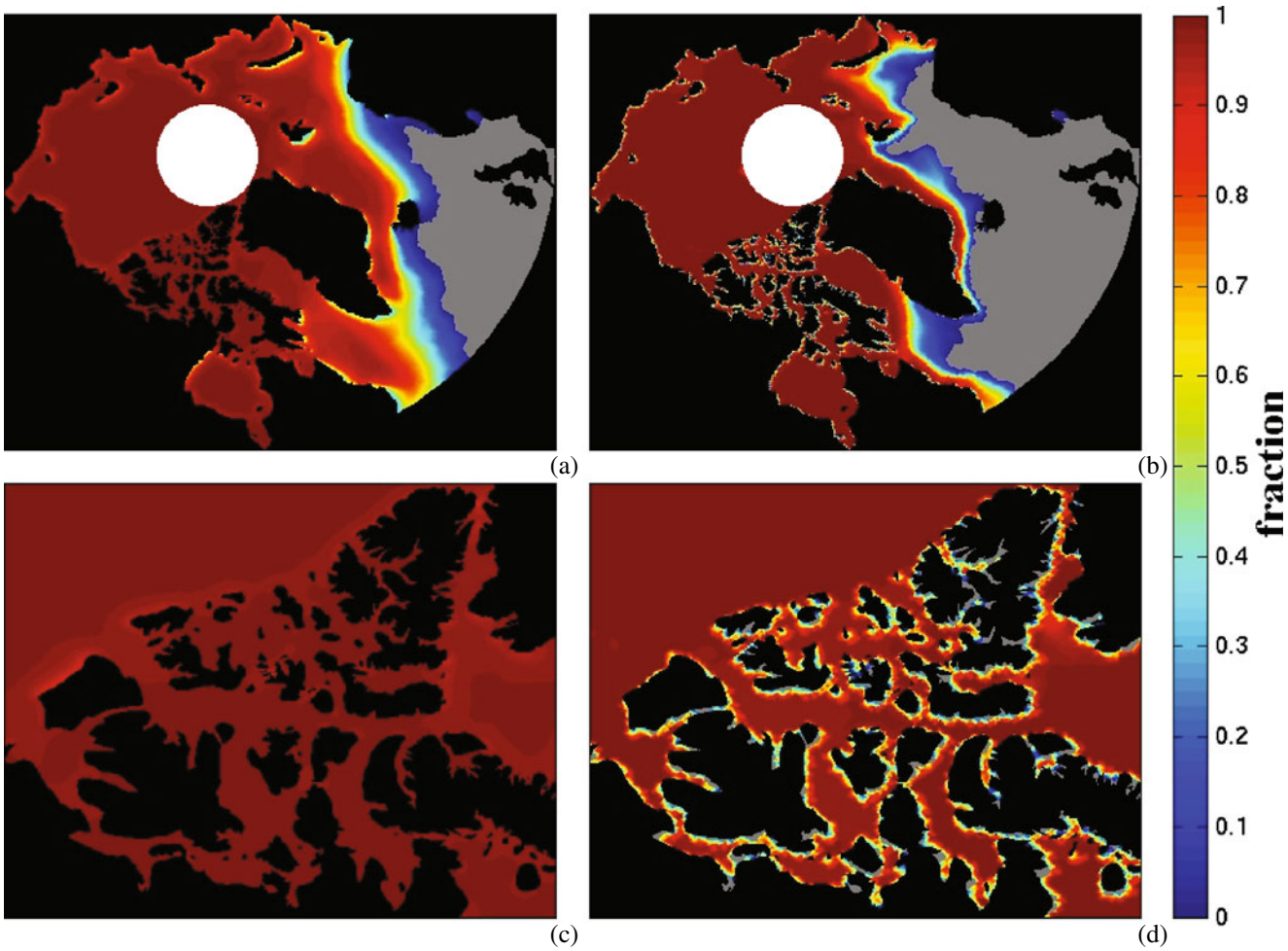

Fig. 2 The mean 1979–2007 March sea-ice concentration. **a** Model result, entire domain with the pole hole masked (in *white*); **b** interpolated Comiso (1999) data, entire domain with the pole hole masked (in *white*); **c** model result, CAA and **d** interpolated Comiso (1999) data, CAA. Note: the low sea-ice concentration near the coastlines in **d** are due to the interpolation

well with the observational mean for the ice cover in the central Arctic. However, the ice extends too far south into the Labrador Sea, Norwegian Sea and the North Atlantic. Since we use a slab ocean, heat that is brought in by the Gulf Stream is not represented in our model. In addition, the PHC climatology has a resolution of $1° \times 1°$, which is too coarse to accurately represent the important ocean currents that transport heat to the Arctic. As a result, oceanic heat advection is missing, resulting in the over estimation of the sea-ice cover. The overextension of the ice was also found in the results of Lietaer et al. (2008) but to a smaller degree, and their mean had a more sharply defined ice edge. This difference is due to the different schemes used. Here $P_1 - P_1$ is used with backward Euler time stepping for the advection, which makes Laplacian diffusion necessary to control the smoothness of the solution. In contrast, Lietaer et al. (2008) use a $P_1^{NC} - P_1$ element with an upwind scheme for the advection. Though this is numerically dispersive, it is less dispersive than the scheme used here. Within the framework used here and in Lietaer et al. (2008), this is hard to mitigate, since it is an artifact of the model and setup. A full 3D ocean might improve this; however, one should keep in mind that currents and therefore also heat advection are influenced greatly by resolution, parameter values and configuration. As such, some improvement is to be expected, but might be limited.

As seen in Fig. 3, only the central Arctic remains covered by sea ice in summer. In our simulation, the sea-ice extent agrees well with the observed extent. Only in the Kara Sea and Baffin Bay is there an overestimation with ice extending too far south. This is a common feature of most sea-ice models (Timmermann et al. 2005; Vancoppenolle et al. 2009), probably associated with missing oceanic advection. The concentration values are a bit lower than those observed in the Arctic Ocean but agree well in the CAA. The mean

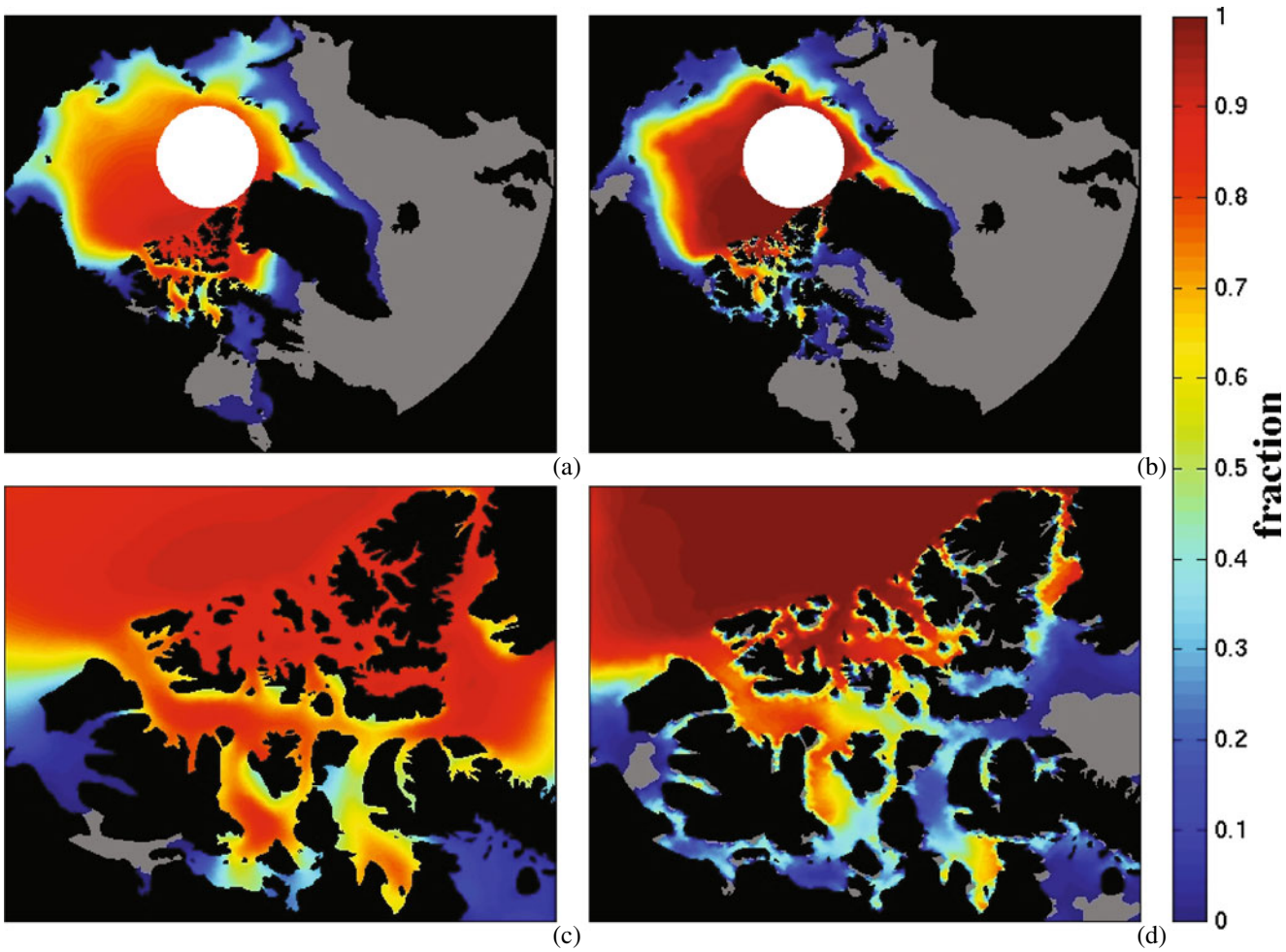

Fig. 3 The mean 1979–2007 September sea-ice concentration. **a** Model result, entire domain with the pole hole masked (in *white*); **b** interpolated Comiso (1999) data, entire domain with the pole hole masked (in *white*); **c** model result, CAA and **d** interpolated Comiso (1999) data, CAA

sea-ice extent found for September agrees well with those obtained by Lietaer et al. (2008). Our mean concentration is a little lower and spatial gradients in the concentration are smaller leading to less small-scale variation. These small differences can be attributed to differences in element formulation, treatment of ice advection and numerical diffusion.

3.2 Total sea-ice area and anomaly

Figure 4 shows the total sea-ice area, defined as the surface integral of the sea-ice concentration and the total sea-ice anomaly for the period 1979–2007. The anomaly has been calculated by subtracting for each month the 1979–2007 mean for that month. In calculating the total sea-ice area model values, data from the SSM/I sensor and model north of 84° N are not used to account for the different coverage around the pole.

As can be seen from Fig. 4a, our model systematically overestimates the total sea-ice area in winter due to missing oceanic heat advection, while the estimates for summer are much closer to the observations. Despite the large differences between our model results and the observations, the variability seen in the total sea-ice anomaly shows good agreement (Fig. 4b). The correlation between both curves is 0.86. In particular, the timing of the upward and downward peaks agree well, although peaks in the model are a bit bigger.

Three distinct periods can be distinguished in the total ice volume and area: (a) 1979–1988, where the anomaly for both model and observations is lower than for later periods; followed by (b) 1988–1998, where the anomaly is much higher for both and finally (c) 1999–2007, where a clear downward trend is present for both the model and the observational record. The shift seen for both model and the observational record in 1988

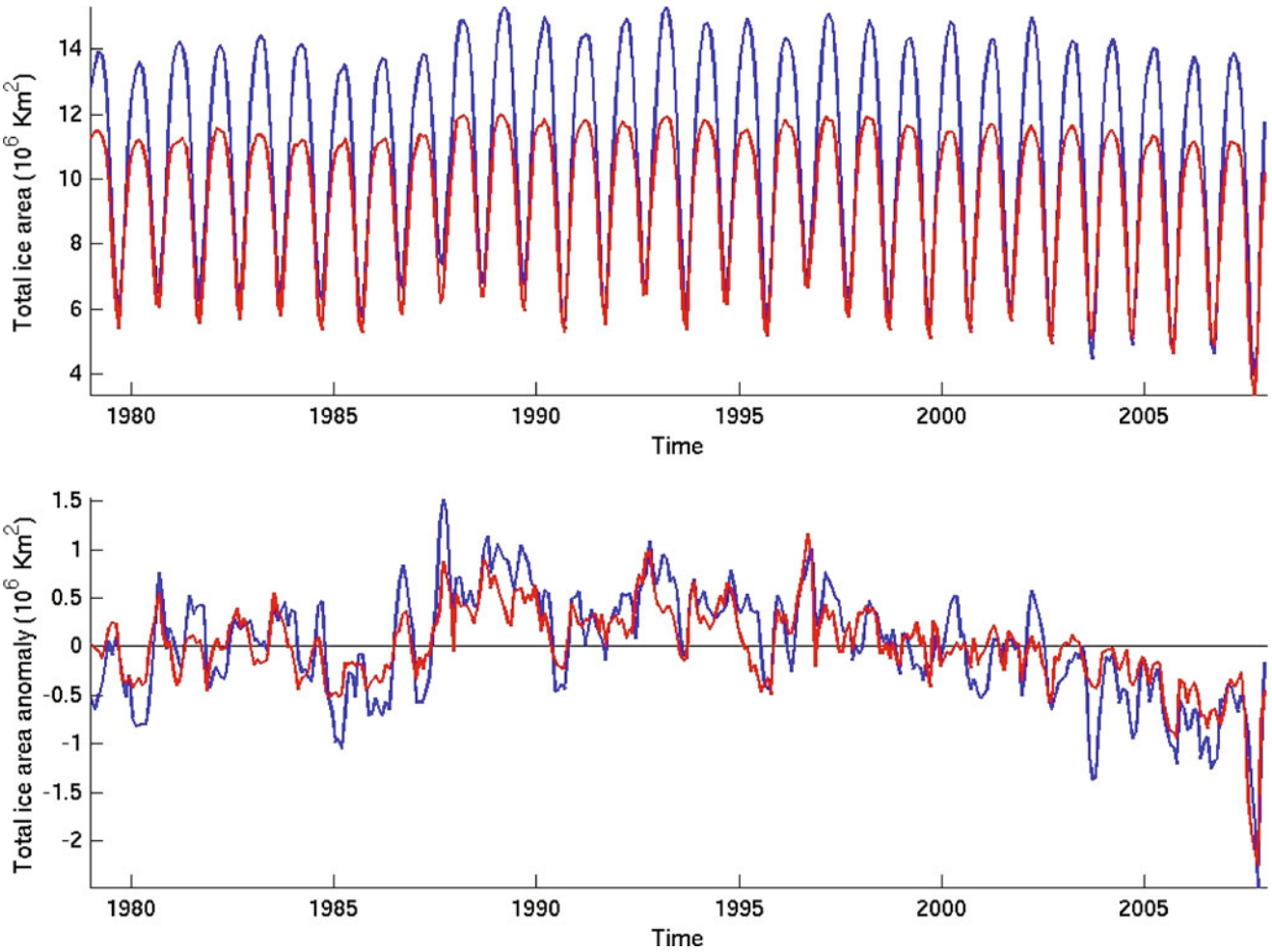

Fig. 4 Total sea-ice area (*top*) and anomaly (*bottom*) as calculated from interpolated Comiso (1999) data (*red*) and model (*blue*)

corresponds well with the changes observed in the AO index (Rigor et al. 2002, see Fig. 5). In this figure, the total sea-ice volume and the AO index are shown, as well as their mean over the periods 1979–1987, 1988–1998 and 1999–2007. The change in the ice cover in 1988 corresponds well with the change seen in the AO from a neutral to a positive mode. For 1998, the connection between AO and ice cover is less clear. The AO moves from the positive back to neutral in 1995, while the downward trend in the sea-ice area anomaly starts in 1999 (Fig. 4).

As such, the first period with lower values for the anomaly corresponds with a neutral AO mode, while the second period (1988–1998) corresponds with a positive AO phase. During a positive AO, ice export through Fram Strait is enhanced, especially multi-year ice, due to a strengthening of the Transpolar Drift and a weakening of the Beaufort Gyre. In Fig. 4, it can be seen that over the second period the total ice volume decreases. This is consistent with the enhanced export. Since a large part of the enhanced export consists of multi-year ice, the fraction of first-year ice increases. This leads to reduced sea-ice extent in summer, since first-year ice is thinner. Figure 4 shows the opposite: an increased ice extent for most of this period including summers. A decrease in total ice volume, coupled with a small increase in ice, results in an ice cover that is decreasing in thickness on average. Our model does not distinguish between first-year and multi-year ice; enhanced export of ice will not have a preference for older or thicker ice. On the contrary, more ice is exported, but it is of an average thickness, since no category of ice is preferred. Instead of a decreasing ice extent, the ice is redistributed. Combined with the model tendency to overestimate the ice extent, an increase in the ice extent is seen. However, during two summers, the ice area anomaly is smaller than on average during this period. This might also explain why reaction to the ice

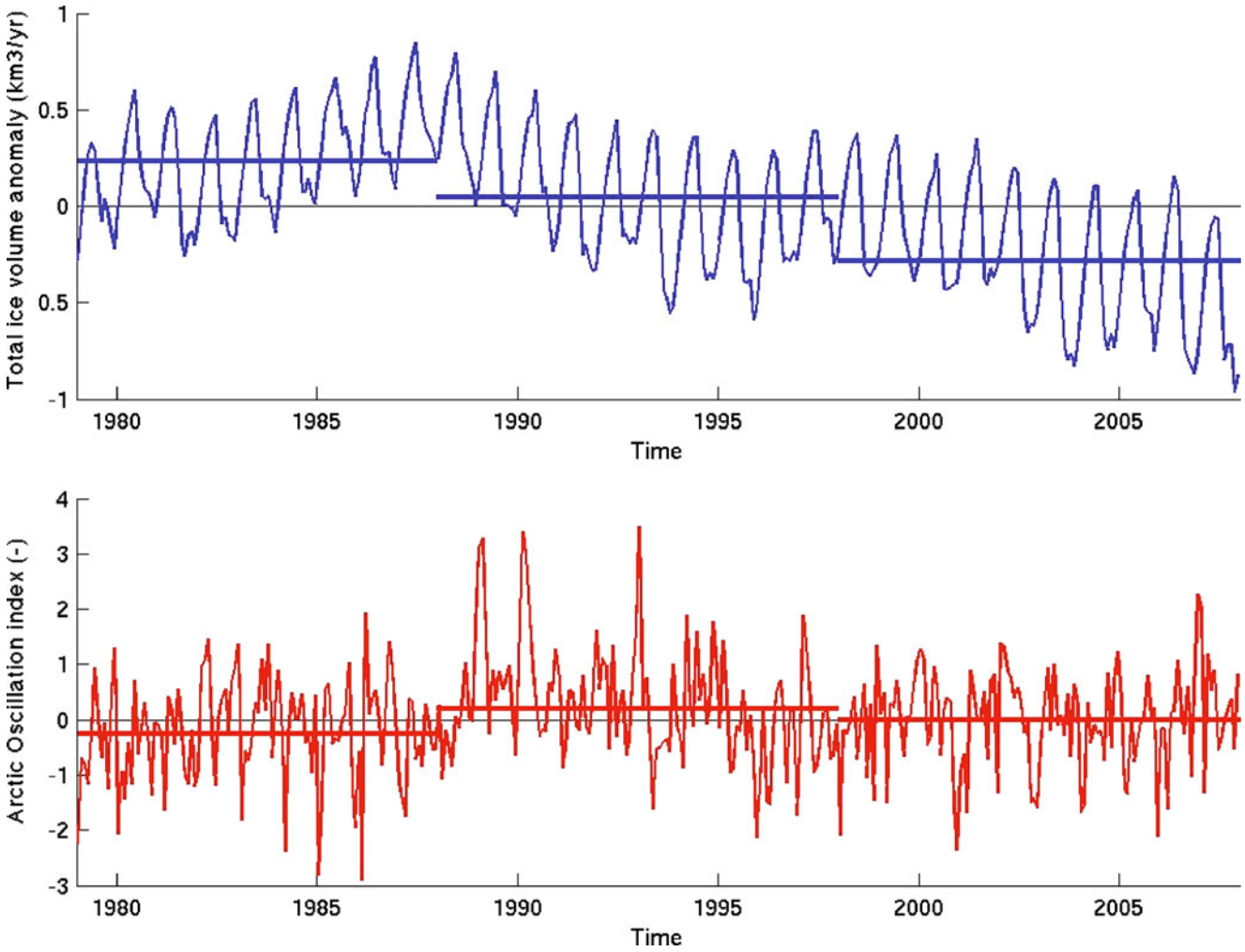

Fig. 5 Total sea-ice volume anomaly (*top*) and the monthly AO index (*bottom*). *Horizontal lines* indicate the mean over 1979–1987, 1988–1998 and 1999–2007

cover to the change of the AO back to a neutral mode is weak and slow.

3.3 Sea-ice thickness

Another important aspect of sea ice is its thickness. In Fig. 6, the 1979–2007 mean thickness is shown. For March, on average the thickness is about 3 m or higher in large parts of the Arctic with a decrease towards the sea-ice edge, which agrees well with Bourke and Garrett (1987). North of the CAA and in the CAA, the average increases to over 4 m and even 5 m in Nares Strait and north of the Queen Elizabeth Islands. This corresponds with those found in observations: 3 to 4 m was found by Melling (2000) for the Queen Elizabeth Islands, 2.5 m found for McClure Strait (Bourke and Garrett 1987; Agnew et al. 2008) and 2–6 m in Nares Strait (Kwok 2006). For summer, Fig. 6b, the picture is different. Due to melt during summer, the thickness is about 2 m lower on average than in winter. Still the highest values of thickness are found near and in the CAA.

3.4 Sea-ice drift

Finally, the 1979–2007 mean of the sea-ice drift is shown in Fig. 7. Important features of the mean ice circulation and transports, the Beaufort Gyre and the Transpolar Drift, are present in our model run. Furthermore, sea ice is transported out of the Arctic Ocean, through Fram Strait, to the Atlantic Ocean and through the CAA and Baffin Bay towards the Labrador Sea.

The magnitude of the ice drift is small in the Arctic Ocean with values of around 2 cm/s. Hibler (1980) using a low-resolution finite-difference model found a comparable value. This indicates that our FEM approach is

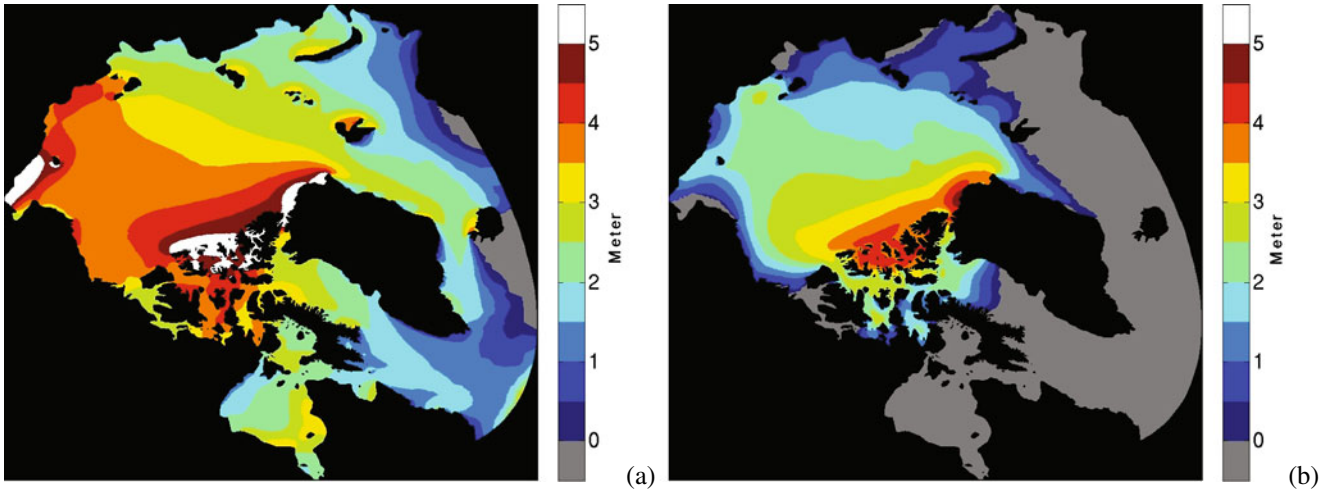

Fig. 6 The 1979–2007 mean sea-ice thickness: **a** March; **b** September

consistent with the Hibler (1980) implementation of the ice dynamics. The magnitude of ice drift and the patterns of flow seen here agree well with Fowler (2003), who estimated the ice drift from satellite observations.

Where the ice concentration is on average low, the ice drift is much faster. This can be seen in regions east of Greenland and in the Labrador Sea. Though the ice concentrations in these regions are not necessarily

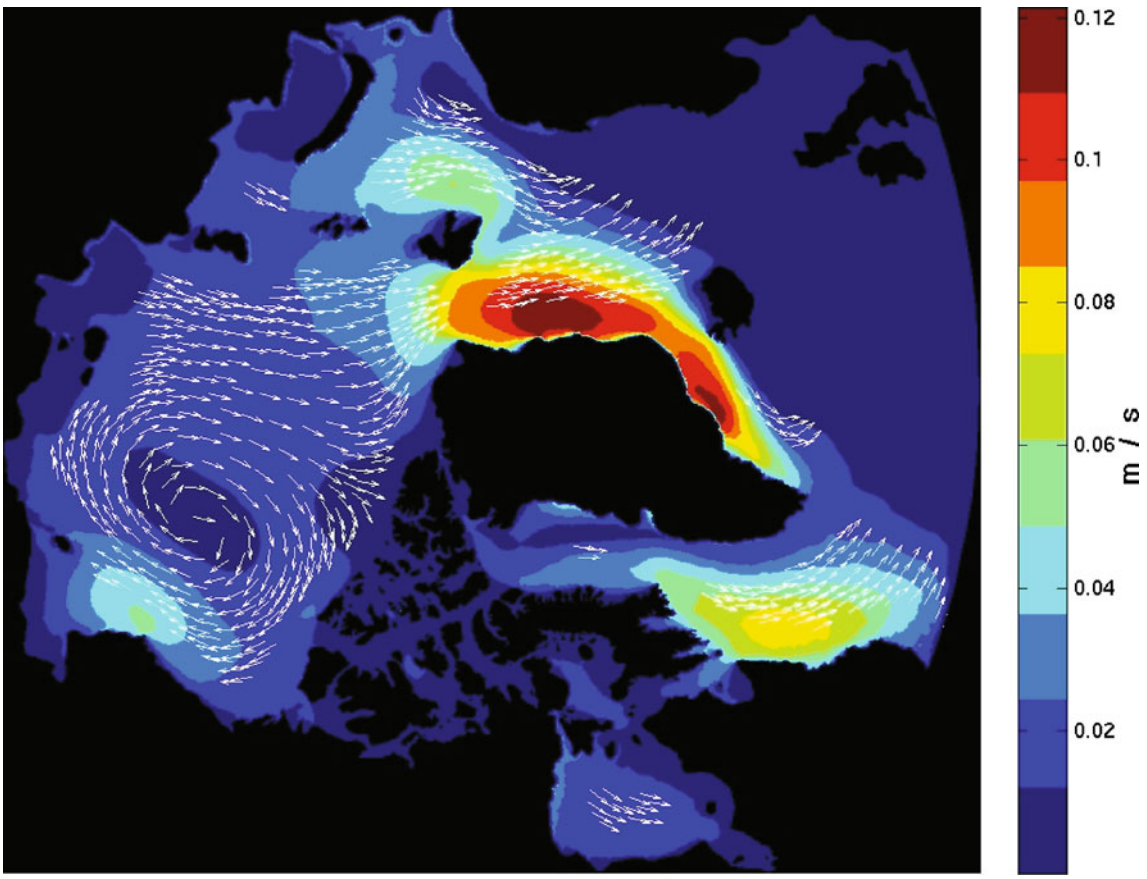

Fig. 7 The 1979–2007 mean of the sea-ice drift over all the months. *Filled contours* represent the ice drift magnitude, while *arrow* represent direction. For viewing purposes, vector were not drawn near coasts

small, in fact they can be quite large, the mean shown in Fig. 7 is a long-term mean. It's values are biased towards the summers with low ice concentrations and relatively fast ice drift.

Within the CAA, the magnitude of the drift is small, due to the high sea-ice concentration and thickness in winter. During these periods, the narrow straits in the CAA act as an effective barrier against sea-ice transport from the Arctic Ocean to the Labrador Sea.

4 Sea-ice fluxes through the CAA

We now focus our attention on the CAA and more specifically on the sea-ice volume and area fluxes through selected straits. Figure 1a shows a geometric representation of the CAA, with the transects along which the sea-ice fluxes have been calculated. We focus our attention on the major pathways of sea ice into the CAA (Admunsen Gulf, McClure Strait and the into the Queen Elizabeth Island), out of the CAA towards Baffin Bay (Lancaster Sound and Jones Sound) and from the Arctic Ocean to Baffin Bay through Nares Strait.

4.1 Comparison 1998–2002

For the period 1998 to 2002, we compare the sea-ice volume and area flux, with several studies. Table 1 summarizes this comparison. In addition, mean ice thicknesses were compared.

Kwok (2006) describes and quantifies the sea-ice transport from the Arctic Ocean to the CAA, based on 5 years of RADARSAT ice motion data. Ice volume fluxes are estimated from the ice area fluxes and assuming a representative ice thickness. This study shows a net mean ice flux into the CAA through Admunsen Gulf and McClure Strait and a net outflow to the Arctic Ocean through the QEI, but large seasonal as well as interannual variability is observed in the fluxes.

A finite difference, regional, coupled sea-ice–ocean model was used by Sou and Flato (2009) to model the flow and variability for the period 1950–2005. The model was forced with NCEP/NCAR reanalysis data and run on a nested grid with a 0.2° resolution. Their sea-ice component uses two categories of thickness and a viscous–plastic rheology. The area fluxes through the Admunsen Gulf and the QEI estimated by this model compare reasonably well with those obtained by Kwok (2006). For McClure Strait, they, however, find mean flow in the opposite direction with respect to what Kwok (2006) found.

Table 1 Comparison of sea-ice volume (in cubic kilometres per year) and area flux (in $10^3 \times km^2\ year^{-1}$) and mean sea-ice thickness (in metres through AG, MS, the gates located in the QEI and NS)

	Kwok (2005, 2006)			Sou and Flato (2009)			Lietaer et al. (2008)			FESOM		
	Volume	Area	Thickness	Volume	Area	Thickness	Volume	Area	Thickness	Volume	Area	Thickness
AG[a]	−85	−85	1.0	−97	−130	1.1	−87	−70	1.5	−89.3	−52.7	1.3
MS[a]	−80	−20	4.0	54	2	2.0	−32	−19	2.6	−41.6	−19.6	2.5
QEI[a]	27	8	3.4	72	11	3.6	131	33	4.2	21.1	3.5	3.9
Net	−138	−97		29	−117		12	−56		−109.8	−68.8	
NS[b]	130	33	4	0.1	−0.1	3.2	–	−3	–	−19.3	−4.3	5.6

See Fig. 1a for location. Note: Positive sign indicate northward transport
AG Admunsen Gulf, *MS* McClure Strait, *QEI* Queen Elizabeth Islands, *NS* Nares Straits
[a]Kwok (2006)
[b]Kwok (2005)

Lietaer et al. (2008) does a slightly better job for Admunsen Gulf and finds good agreement with Kwok (2006) for the ice area flux through McClure Strait. For this strait, however, a lower volume flux was found, mainly due to the lower mean ice thickness when compared to the representative ice thickness used by Kwok (2006). The use of a representative thickness to calculate the ice volume fluxes in Kwok (2006) was questioned by Sou and Flato (2009), arguing that in a region of high thickness and drift variability, a constant thickness might lead to inaccurate estimates for the ice volume flux. Both volume and area flux found by Lietaer et al. (2008) are over estimated for the Queen Elizabeth Islands by a factor of about 4.

Here we find good agreement with Kwok (2006) for the volume flux for Admunsen Gulf and the Queen Elizabeth Islands. Additionally, the mean thickness compares well with the representative thickness used by Kwok (2006); however, the area fluxes are lower. For McClure Strait, the area flux compares well, although the volume flux is half of the observed due to the low mean thickness.

Compared with Lietaer et al. (2008), here we find slightly lower mean thicknesses. For Admunsen Gulf, the volume and area fluxes are comparable, but for McClure Strait and the Queen Elizabeth Islands, our model performs better, when compared with Kwok (2006). Especially for the Queen Elizabeth Islands, our results are much closer to the observed fluxes. Compared to Sou and Flato (2009) and Lietaer et al. (2008), the net fluxes into the CAA compare very well with those observed by Kwok (2006). Both these studies found a small net flow out of the CAA.

For Nares Strait, both Lietaer et al. (2008) and Sou and Flato (2009) found that the net area ice flux through this region was too small when compared with observations. Lietaer et al. (2008) found a mean area ice flux of $-3{,}000$ km^2 year^{-1}, and Sou and Flato (2009) found an area flux of an order of magnitude lower. Here we find an area flux of $-4{,}300$ km^2 year^{-1}, which is higher than the estimate found by Lietaer et al. (2008), but still much lower than the area flux estimated by Kwok (2005). In addition, our mean thickness in this strait in 5.6 m for this period, which is well over the 4-m representative thickness used by Kwok (2005). Sou and Flato (2009) attributed their low estimate to the wind forcing, which is missing the effect of topographic steering on local wind patterns. Although we use different atmospheric forcing (NCEP/DoE reanalysis instead of NCEP/NCAR reanalysis), the resolution is still the same and is inadequate to capture the strong along-strait winds known to dominate the region, a limitation most models have, since high resolution is not available for long periods of time for this region. The low ice flux through this strait can be attributed to the same deficiencies in the atmospheric forcing. In addition, Lietaer et al. (2008) also suggested that for sea-ice only models, the missing ocean currents and spatial resolution might also have a big influence on the ice fluxes through Nares Strait. The spatial resolution of our mesh in Nares Strait and surrounding areas is more refined than the mesh used by (Lietaer et al. 2008), leading to slightly higher ice area flux estimates since we resolve fluxes better in Nares Straits. The consistent underestimation of the ice volume and ice area flux through Nares Strait by several models suggests that atmospheric forcing is of critical importance for modelling this region.

4.2 Comparison 2002–2007

A recent study by Agnew et al. (2008) estimated the sea-ice flux between the CAA and the Arctic Ocean and Baffin Bay for the period September 2002 to June 2007 using enhanced resolution Advanced Microwave Scanning Radiometer imagery (Long and Daum 1999). This study did not include the summer months July and August. The estimated mean ice volume and area fluxes and mean thickness are shown in Table 2. In order to compare the results with their data (Table 2) shows the results for the model, including and excluding the months of July and August. The study of Agnew et al. (2008) will be compared with the model excluding the two months. To investigate the influence of July and August, the fluxes as estimated by the model with and without these months will be compared.

For both Admunsen Gulf and McClure Strait, the model overestimates the volume ice flux by a factor of 3 when July and August are excluded from the analysis, when compared to Agnew et al. (2008). The ice area flux is also a factor of 4 too high, while the mean thickness compares well. Agnew et al. (2008) found no significant volume or ice flux through the Queen Elizabeth Islands and Jones Sound and suggested that this was due to missing the summer months July and August in the analysis. For Jones Sound, the model shows small ice volume and area fluxes, which is consistent with (Agnew et al. 2008). For the Queen Elizabeth Islands, we find a volume flux of -9.1 km^3 excluding July and August, which suggests that the model also has ice exchange during winter between the Arctic Ocean and the CAA.

The estimate for the full year for the volume flux is 9.0 km^3 through the Queen Elizabeth Islands. This suggest that during July and August, the ice is highly mobile in the model and sea ice is transported out of

Table 2 Comparison of sea-ice volume (in cubic kilometres per year) and sea-ice thickness (in meters) through AG, MS, the gates located in the QEI, LS and JS

	Agnew et al. (2008) excluding July and August			FESOM excluding July and August			FESOM including July and August		
	Volume	Area	Thickness	Volume	Area	Thickness	Volume	Area	Thickness
AG	−42	−42	1.0	−110.3	−49.9	1.5	−84.4	−42.0	1.4
MS	−30	−12	2.5	−108.2	−41.8	2.6	−66.0	−27.7	2.6
QEI	0	0	3.4	−9.1	−7.7	4.1	9.0	−3.0	4.0
LS	102	68	1.5	28.3	8.8	2.5	29.9	10.0	2.5
JS	0	0	1.5	−3.4	−0.8	3.8	−3.1	−0.7	3.7

See Fig. 1a for locations. Note: positive sign indicate northward transport
AG Admunsen Gulf, *MS* McClure Strait, *QEI* Queen Elizabeth Islands, *LS* Lancaster Sound, *JS* Jones Sound

the archipelago. From Table 2, we see that excluding July and August from the analysis makes a difference. Our estimates of flow into the CAA from the Arctic Ocean are smaller when excluding July and August, or even change sign. This shows that the model represents high seasonal variability.

4.3 Long-term mean and variability

In Fig. 8, the series of the monthly mean ice volume fluxes are shown for six gates in the CAA. All these figures show strong seasonal and interannual variability with the magnitude of the variability several times the mean. Note that in this figure the sign convention is northward and eastward is positive. For Admunsen Gulf, we find an average inflow into the CAA of 70.0 ± 175.2 km^3 year^{-1} (Fig. 8a). In 1989, a maximum monthly mean outflow is seen which is preceded by 2 years where only inflow occurred. After 1989, a change in the variability can be seen: The magnitude of monthly mean inflow becomes bigger with lower mean, changing from -73.4 ± 160.4 to -47.5 ± 210.4 km^3 year^{-1}. Both the 1995 and 2005 maxima in the inflow correspond to the (then) record minimum summer ice cover in the Arctic, which made Admunsen Gulf ice free and ice transport through this gate easier. In 1998, no outflow occurred and from this year on the mean increases again while the variability decreases compared to the period 1978–1997: -87 ± 152.2 km^3 year^{-1}.

The volume flux for McClure Straits shows some of the same features (Fig. 8b). The mean inflow through McClure strait is lower, -32.8 ± 270.9 km^3 year^{-1}, but the variability is a bit stronger. Also this gate shows a large outflow in 1989. Though not the maximum outflow seen in this figure, it is biggest since 1979. Like Admunsen Gulf, it is preceded by 2 years with almost no outflow to the Arctic when compared to previous years. In 1991, a maximum monthly mean outflow can be seen, after which each large outflow event is smaller in magnitude. For the period 1979–1988, the volume flux is -62.6 ± 279.4 km^3 year^{-1}. After 1989, the monthly inflow is a bit smaller than during 1978–1988, on average -20.3 ± 323.0 km^3 year^{-1} for 1989–1997. During 1998–2001, the magnitude of variability is much lower than before or after this period. This explains the difference we see in the mean volume transport between the period 1998–2002 (Table 1) and 2002–2007 (Table 2). After 1998, the mean volume flux is -50.9 ± 195.8 km^3 year^{-1}.

The ice volume flux for the QEI region has very strong seasonal variability, with inflow and outflow of the same order (Fig. 8c). For 1979–2007, there

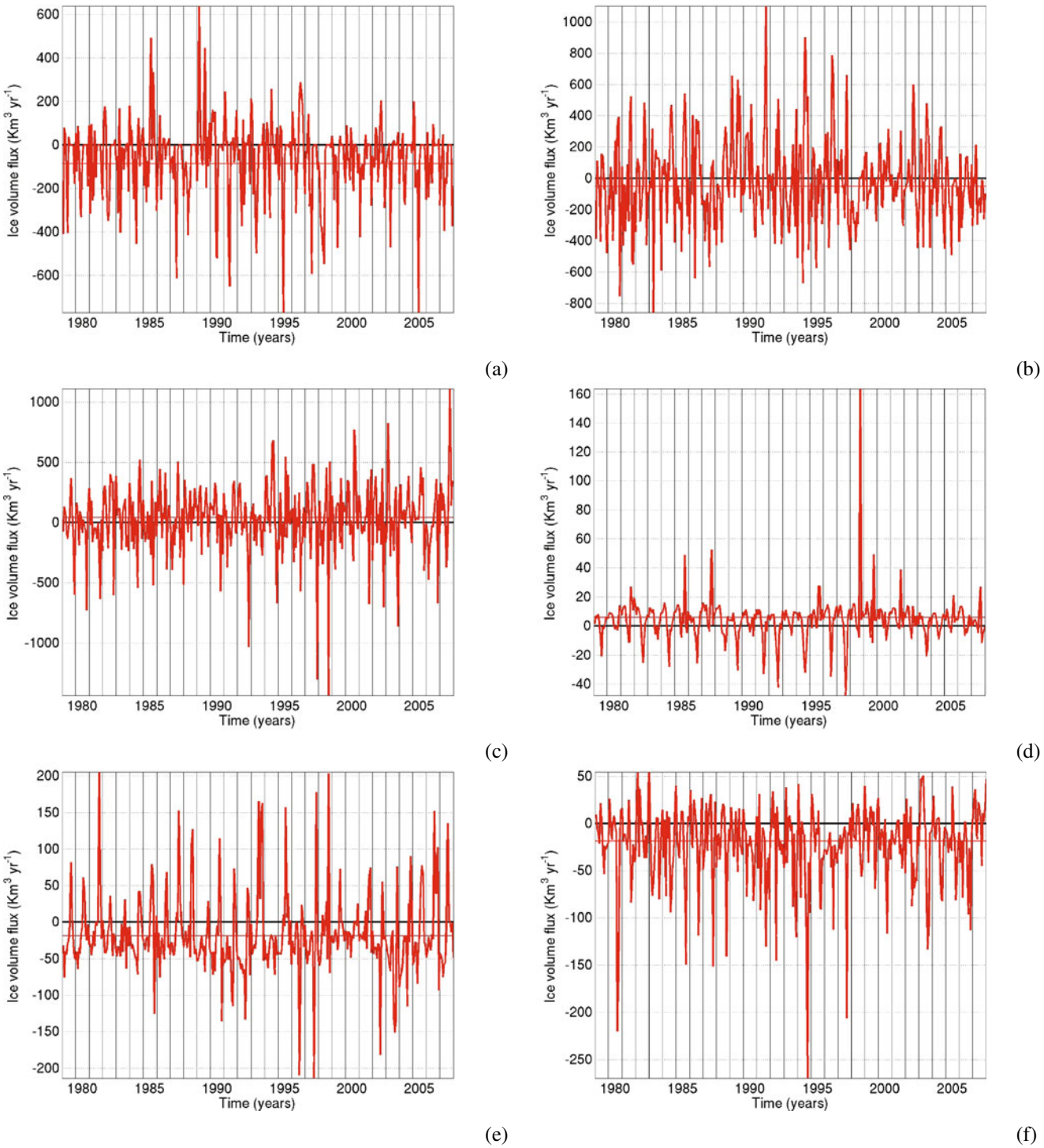

Fig. 8 Time series of ice volume flux: **a** Admunsen Gulf, **b** McClure Straits, **c** the Queen Elizabeth Islands, **d** Jones Sound, **e** Lancaster Sound and **f** Nares Strait. The mean of these time series are indicated by the *horizontal red line*. Note: Positive sign indicates northward transport

is a mean outflow into the Arctic Ocean of 16.6 ± 282.9 km^3 year^{-1}. The magnitude of the variability seems to increase over time, with a maximum monthly volume flux into the QEI region in 1998 and a maximum flow to the Arctic Ocean in 2007.

In Fig. 8d, e, the time series of the monthly mean volume flux are shown for Jones Sound and Lancaster Sound, respectively. Jones Sound has a mean southward transport of 4.0 ± 14.3 km^3 year^{-1}, while the latter has a northward transport of 19.6 ± 53.2 km^3 year^{-1}.

For both, a large outflow to Baffin Bay can be seen in 1998, and in the case of Jones Sound, the magnitude is five to six times larger when compared with other years. The variability seen in the series for Jones Sound is smaller after 1998 (26.3 ± 60.6 km^3 year^{-1}) than the period before 1998 ($18.7 \pm 6{,}053.6$ km^3 year^{-1}). A change can also be seen for Lancaster Sound. During the last 10 years, a trend appears to be present: The peak in monthly mean northward flow decreases, while the flow in the opposite direction becomes more pronounced. This in contrast to the series before 1998, where the mean flow is northward, with some occasional months where there is a big southward transport. Note also that as with the gates previously discussed, after 1989, the variability increases somewhat.

For Nares Strait (Fig. 8d), there is a small flux from the Arctic Ocean to Baffin Bay of -23.6 ± 42.2 km^3 year^{-1}, which is much lower than the estimate given by Kwok (2005). Since local effects of topography are not well presented in the atmospheric forcing fields, some of the features seen in the other figures are not seen here.

To a certain degree and with the exception of Nares Straits, the same changes in the variability can be observed for the gates in Fig. 8: (a) a change around 1988–1989 and (b) a change around 1998. The timing of these changes correspond well with those seen in the anomaly of the total sea-ice area (Fig. 4b). For the former, the changes seen can be for a large part associated with the AO. The timing with the change from neutral to positive AO agrees very well. The AO has influence on the sea-ice thickness (Rigor et al. 2002). For the years 1988–1990, the sea ice north of the CAA, from Nares Strait to Admunsen Gulf, was on average 1–1.5 m thicker. This increases the potential for higher ice volume transports, as is clearly the case for Admunsen Gulf and McClure Strait (Fig. 8a, b). This holds true in general and as such volume transport through the gates will increase on average. As a consequence, more sea ice north of the CAA is more effective at blocking volume transport through the CAA by forming ice bridges. When melt sets in and the bridges no longer obstruct ice flow through the gate, the volume of ice that will be flushed through will be much larger than under neutral or negative AO conditions. This in turn leads to larger variability during the positive AO index years.

It has been observed that the sea-ice concentration was at a record low for the CAA in 1998 due to an anomalous warm summer, and it has been suggested that this conditioned the ice to be more favourable to ice import from the Arctic Ocean (Jeffers et al. 2001). The volume flux for Admunsen Gulf, McClure Strait, the QEI and Nares Strait show this high import into the CAA for 1998, while having lower or absent outflow into the Arctic Ocean. This is especially clear for the QEI and Nares Strait. The large outflow through Lancaster Sound and Jones Sound for 1998 might be related to the same conditions. The changes seen in the variability after 1998 are mainly due to the large-scale changes observed in the Arctic in the past decade.

4.4 Sources of the variability

To investigate the sources driving the variability, the volume flux is separated in different components. The ice thickness H is split in a mean $\bar{H}$ and an anomaly $\tilde{H}$, and likewise the ice velocity $\mathbf{u}$ is split in a mean $\bar{\mathbf{u}}$ and an anomaly $\tilde{\mathbf{u}}$. The volume flux V through a section L can be written as:

$$V = \int_L H\mathbf{u} \times \mathbf{n} dL \tag{5}$$

$$= \int_L (\bar{H} + \tilde{H})(\bar{\mathbf{u}} + \tilde{\mathbf{u}}) \times \mathbf{n} dL \tag{6}$$

$$= \int_L \left[\bar{H}\bar{\mathbf{u}} \times \mathbf{n} + \tilde{H}\bar{\mathbf{u}} \times \mathbf{n} + \bar{H}\tilde{\mathbf{u}} \times \mathbf{n} + \tilde{H}\tilde{\mathbf{u}} \times \mathbf{n}\right] dL \tag{7}$$

where $\mathbf{n}$ is the vector to L. The volume flux has four components: (a) a mean flux (horizontal lines in Fig. 8), (b) flux driven only by the thickness variability (red curves in Fig. 9); (c) flux driven only by the variabilities in the velocity field (green curves in Fig. 9) and (d) flux driven by a combination of thickness variability and variabilities in the velocity field (blue curve in Fig. 9).

As can been seen in Figs. 8 and 9, both the mean and flux driven by thickness variability are the smallest components of the ice volume flux for all the six straits. As such, the variability of the velocity field drives the variability in the ice volume flux. The third term in Eq. 7 is the dominant term for all straits, indicating that for a large part the variability seen in Fig. 8 is due to the advection of ice of a mean thickness due to an anomalous velocity field.

The last term of Eq. 7 is important for some straits. For Admunsen Gulf (Fig. 9a) and McClure Strait (Fig. 9b), the contribution of this term to the ice volume transport is comparable to those driven by the variability in the velocity field. For the other four straits, this term contributes much less. A possible explanation is that both Admunsen Gulf and McClure Strait have a much bigger width, resulting in a much more variable ice cover and small-scale variability such as recirculation of ice. This can be seen to a lesser extent in Lancaster Sound.

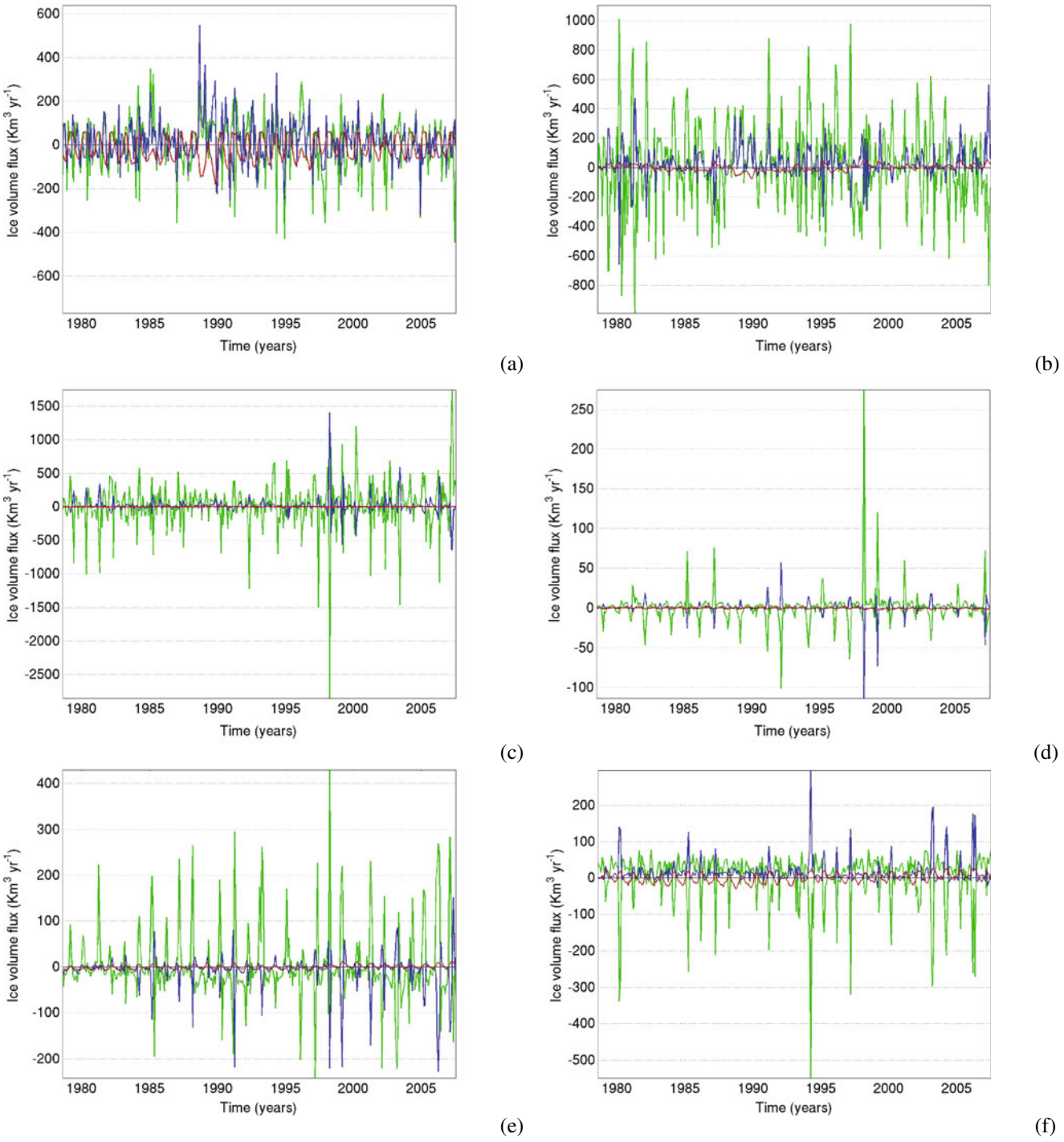

Fig. 9 The time series of the components of the ice volume flux variability: **a** Admunsen Gulf, **b** McClure Straits, **c** the Queen Elizabeth Islands, **d** Jones Sound, **e** Lancaster Sound and **f** Nares Strait. In *red*, $\tilde{H}\tilde{\mathbf{u}} \times \mathbf{n}$; in *green*, $\bar{H}\tilde{\mathbf{u}} \times \mathbf{n}$ and in *blue*, $\tilde{H}\bar{\mathbf{u}} \times \mathbf{n}$. Note: positive sign indicate northward transport

5 Discussion

In this paper, a model for the CAA was presented. This model is based on FESOM (Timmermann et al. 2009), a finite-element coupled sea-ice–ocean model. Instead of a full 3D ocean, the model uses a simple slab ocean, resulting in the same type of model as in Lietaer et al. (2008), except that in this study, both ocean temperature and salinity characterize the slab ocean. The model was forced with NCEP/NCAR reanalysis

data for the period 1973–1978 and with NCEP/DoE reanalysis two data for 1979–2007. A mesh with a coarse resolution for the Arctic Ocean was used, with a highly refined resolution for the CAA. Though this model is relatively simple, it is the first step in a systematic approach to investigate the complex dynamics of the Arctic region, by studying the region using models of increasing complexity.

For winter, the modelled climatological ice thickness and concentration fields were in good agreement with the observations for the central Arctic but extended too far south into the Nordic Seas and Labrador Sea. Since we use a simple slab ocean, this is most likely due to the missing heat advection in the ocean. For summer, however, the model shows good agreement with the observational data. Though the sea-ice concentration is a bit lower than observed, the extent agrees very well apart from Baffin Bay and the Kara Sea, where the ice extended too far to the south. This is a common feature of many sea-ice models (Timmermann et al. 2005; Vancoppenolle et al. 2009). Compared to the study by (Lietaer et al. 2008), the values of the sea-ice concentration were also a bit low and too smooth, due to the different finite-element formulations used in the two models. In addition, our model has some numerical diffusion which smoothens the fields, resulting in less small-scale spatial variability. The main features of the Arctic sea-ice circulation are represented in the ice drift climatology.

The total sea-ice area anomaly agrees well with that observed. Hence, despite slight underestimation of the ice concentration (and thickness) during the summer and overestimation of the sea-ice extent in the winter, our model is able to represent sea-ice variability. Moreover, large-scale climate changes, such as the Arctic oscillation and the recent decline in sea ice are captured by our model. For 1979–1987, the anomaly of the total sea-ice cover was lower and coincides with a neutral AO index, while for 1987–1998, the anomaly is much higher and corresponds well with the positive AO index. Importantly, after 1998, the warming of the Arctic can be clearly seen as a downward trend towards less sea ice.

We compared the fluxes through the CAA with estimates obtained from observational data and other model studies (Sou and Flato 2009; Lietaer et al. 2008) for two 5-year periods. The results found here compare well with these studies. For the Queen Elizabeth Islands and the net inflow into the CAA from the Arctic Ocean, we find mean sea-ice volume fluxes close to estimates from observational data (Kwok 2006). When compared with estimates from observations (Kwok 2005, 2006; Agnew et al. 2008), differences can be explained by differences in methodology. The observational estimates are based on assuming a constant characteristic ice thickness, which does not take the high variability in the sea-ice characteristics into account. Differences were found between the models, but these can be attributed to differences in implementation, model formulation or spatial resolution leading to different distributions of thickness and concentration and different ice drift.

Long-term time series of the monthly mean volume flux show a high interannual and seasonal variability, which is mainly driven by variability in the ice velocity field. In addition, variability in the atmospheric forcing has a strong influence on these fluxes. First of all, the magnitude of the variability of the sea-ice volume flux through most straits increases during positive AO phases. This is mainly due to the dynamical effect the AO has on the winter time ice thickness and the response of the summertime concentration (Rigor et al. 2002). Secondly, the effect of the anomalous warm summer and autumn in 1998 (Jeffers et al. 2001) can be seen as an increase of volume flux into the CAA.

The results presented here indicate that this model is capable of estimating the sea-ice fluxes through the CAA. In addition, the interannual and seasonal variability is well represented, especially in the CAA. Though our model is relatively simple, it is able to capture the sea-ice dynamics in the CAA. As such, it is a good benchmark for further model development, and the model can be used for first-order studies into associated dynamical processes.

Acknowledgements The authors would like to thank the Alfred Wegener Institute, Bremerhaven, Germany—especially Jens Schröter—for making the FESOM code available for this study. The NCEP/NCAR reanalysis data and NCEP/DoE reanalysis 2 data were provided by the NOAA/OAR/ESRL PSD, Boulder, CO, USA. The sea-ice concentration data were provided by the NSIDC, Boulder, CO, USA. This study received funding support from Arcticnet.

References

Aagaard K, Carmack E (1989) The role of sea ice and other fresh water in the Arctic circulation. J Geophys Res 94:14,485–14,498

Adcroft A, Marshall D (1998) How slippery are piecewise-constant coastlines in numerical ocean models? Tellus 50A: 95–108

Agnew T, Lambe A, Long D (2008) Estimating sea ice area flux across the Canadian Arctic Archipelago using enhanced AMSR-E. J Geophys Res 113:C10,011

Balay S, Gropp WD, McInnes LC, Smith BF (1997) Efficient management of parallelism in object oriented numerical software libraries. In: Arge E, Bruaset AM, Langtangen

HP (eds) Modern software tools in scientific computing. Birkhäuser, Boston, pp 163–202
Balay S, Buschelman K, Eijkhout V, Gropp WD, Kaushik D, Knepley MG, McInnes LC, Smith BF, Zhang H (2008) PETSc users manual. Tech. Rep. ANL-95/11—Revision 3.0.0. Argonne National Laboratory, Argonne
Balay S, Buschelman K, Gropp WD, Kaushik D, Knepley MG, McInnes LC, Smith BF, Zhang H (2009) PETSc Web page. http://www.mcs.anl.gov/petsc
Becker E (1976) The finite element method in AIDJEX. AIDJEX Bull 13:144–157
Bourke R, Garrett R (1987) Sea ice thickness distribution in the Arctic Ocean. Cold Regions Sci Technol 13:259–280
Chen C, Beardsley R (2003) An unstructured, finite-volume, three-dimensional, primitive equation ocean model: application to coastal ocean and estuaries. JAtmOceTech 20: 159–186
Chen C, Beardsley R, Cowles G (2006) An unstructured grid, finite-volume coastal ocean model (FVCOM) system. Oceanogr 19:78–89
Chen C, Gao G, Qi J, Proshuntinsky A, Beardsley R, Cowles G, Huang H (2009) A new high-resolution unstructured-grid, finite-volume Arctic Ocean model (AO-FVCOM): an application for tidal simulation. J Geophys Res 114:C08,017
Comiso J (1999) Bootstrap sea ice concentrations from NIMBUS-7 SMMR and DMSP SSM/I. Digital media. National Snow and Ice Data Center, Boulder (updated 2008)
Comiso J (2002) A rapidly declining Arctic perennial ice cover. Geophys Rev Lett 30:1956
Comiso J, Steffen K (2008) Introduction to special section on large-scale characteristics of the sea ice cover from AMSR-E and other satellite sensors. J Geophys Res 113:C02S1
Danilov S, Kivman G, Schröter J (2004) A finite-element ocean model: principles and evaluation. Ocean Mod 6:125–150
Dickson R, Rudels B, Karcher M, Meincke J, Yashayaev I (2007) Current estimates of freshwater flux through Arctic and subarctic seas. Prog Oceanogr 73:210–230
Ebert E, Curry J (1993) An intermediate one-dimensional thermodynamic sea ice model for investigation ice–atmosphere interactions. J Geophys Res 98:10,085–10,109
Fischer H (1995) Vergleichende Untersuchungen eines optimierten dynamischen–thermodynamischen Meereismodells mit Beobachtungen im Wendellmeer. In: Berichte zur Polarforschung, vol 116. Alfred-Wegener Institut (AWI), Bremerhaven, Germany
Flato G, Boer G (2001) Warming asymmetry in climate change simulations. Geophys Res Lett 28:195–198
Fowler C (2003) Polar pathfinder daily 25 km ease-grid sea ice motion vectors. Digital media. National Snow and Ice Data Center, Boulder (updated 2007)
Geuzaine C, Remacle JF (2009) Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities. Int J Numer Methods Eng 79:1309–1331
Goosse H, Fichefet T, Campin JM (1997) The effect of the flow through the Canadian Archipelago in global ice–ocean model. Geophys Res Lett 24:1507–1510
Hibler W (1979) A dynamic thermodynamic sea ice model. J Phys Oceanogr 9:815–846
Hibler W (1980) Sea ice growth, drift and decay. In: Dynamics of snow and ice masses. Academic, New York
Hunke E, Dukovicz J (1997) An elastic–viscous–plastic model for sea ice dynamics. J Phys Oceanogr 27:1849–1867
Jeffers S, Agnew T, De Abreu R, McCourt S (2001) Investigating the anomalous sea ice conditions in the Canadian High Arctic (Queen Elizabeth Islands) during the summer of 1998. Ann Glaciol 33:507–512
Johannessen O (2004) Arctic climate change: observed and modelled temperature and sea-ice variability. Tellus 56A: 328–341
Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Leetma A, Reynolds B, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo K, Ropelewski C, Wang J, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77:437–470
Kanamitsu M, Ebisuzaki W, Woollen J, Yang SK, Hnilo J, Potter G (2002) NCEP-DOE AMIP II reanalysis (R-2). Bull Am Meteorol Soc 83:1631–1642
Kliem N, Greenberg D (2003) Diagnostic simulations of the summer circulation in the Canadian Arctic Archipelago. Atmos Ocean 41:273–289
Kwok R (2005) Variability of Nares Strait ice flux. Geophys Res Lett 32:L24,502
Kwok R (2006) Exchange of sea ice between the Arctic Ocean and the Canadian Arctic Archipelago. Geophys Res Lett 33:L16,501
Kwok R, Cunningham G, Pang S (2004) Fram Strait sea ice outflow. J Geophys Res 109:C01,009
Lambrechts J, Comblen R, Legat V, Geuzaine C, Remacle JF (2008) Multiscale mesh generation on the sphere. Ocean Dyn 58:461–473
Le Provost C (1986) On the use of finite element methods for ocean modelling. In: Advanced physical oceanography numerical modelling. Reidel, Dordrecht, pp 557–580
Lefebvre W, Goosse H (2008) Analysis of the projected regional sea-ice changes in the Southern Ocean during the 21st century. Clim Dynam 30:59–76
Legrand S, Deleersnijder E, Hanert E, Legat V, Wolaski E (2006) High resolution, unstructured meshes for hydrodynamic models of the Great Barrier Reef, Australia. Estuar Coast Shelf Sci 68:36–46
Leppäranta M (1983) Growth model for black ice, snow ice and snow thickness in subarctic basins. Nord Hydrol 14: 59–70
LeRoux D, Lin C, Staniforth A (2000) A semi-implicit finite element shallow-water ocean model. Mon Weather Rev 128:1384–1401
Lietaer O, Fichefet T, Legat V (2008) The effects of resolving the Canadian Arctic Archipelago in a finite element sea ice model. Ocean Model 24:114–152
Long D, Daum D (1999) Spatial resolution enhancement of SSM/I data. EEE Trans Geos Rem Sen 36:407–417
Melling H (2000) Exchanges of freshwater through the shallow straits of the North American Arctic, in the freshwater budget of the Arctic Ocean. Kluwer Academic, Dordrecht
Melling H (2002) Sea ice cover in the northern Canadian Arctic Archipelago. J Geophys Res 107:3181
Millero F (1978) Freezing point of sea water. In: Eight report of the Joint Panel of Oceanographic Tables and Standards, UNESCO Technical Papers in Marine Science, vol 28, pp 29–31
Mukherji B (1973) Crack propagation in sea ice: a finite element approach. AIDJEX Bull 18:69–75
Owens W, Lemke P (1990) Sensitivity studies with a sea ice-mixed layer-pycnocline model in the Weddell Sea. J Geophys Res 95:9527–9538
Parkinson C, Washington W (2007) A large-scale numerical model of sea ice. J Geophys Res 84:311–337
Pietrzak J, Deleersnijder E, Schröter J (2005) The second international workshop on unstructured mesh numerical modelling of coastal, shelf and ocean flows. Ocean Model 10:1–252. Special issue

Rigor I, Wallace J, Colony R, Martin S (2000) Variations in surface air temperature observations in the Arctic, 1979–97. J Climate 13:896–914

Rigor I, Wallace J, Colony R (2002) Response of sea ice to the Arctic oscillation. J Climate 15:2648–2663

Rollenhagen K, Timmermann R, Janjić T, Schröter J, Danilov S (2009) Assimilation of sea ice motion in a finite-element sea ice model. J Geophys Res 114:C05,007

Rothrock D, Yu Y, Maykut G (1999) Thinning of the Arctic sea-ice cover. Geophys Res Lett 26:3469–3472

Schulkes R, Morland L, Staroszczyk R (1998) A finite-element treatment of sea ice dynamics for different ice rheologies. Int Numer Anal Methods Geomesh 22:153–174

Semtner A (1976) A model for the thermodynamic growth of sea ice in numerical investigations of climate. J Phys Oceanogr 6:379–389

Serreze M, Barrett A, Slater A, RA W, Aagaard K, Lammers R, Steele M, Moritz R, Meredith M, Lee C (2006) The large-scale freshwater cycle of the Arctic. J Geophys Res 111:C11,010

Serreze M, Holland M, Stroeve J (2007) Perspectives on the Arctic's shrinking sea-ice cover. Science 315:1533–1537

Sodhi D, Hiber W (1980) Nonsteady ice drift in the Strait of Belle Isle. In: Pritchard RS (ed) Sea ice processes and models. University of Washington Press, Seattle

Sou T, Flato G (2009) Sea ice in the Canadian Arctic Archipelago: modeling the past (1950–2004) and the future (2041–60). J Climate 22:2181–2198

Steele M, Morley R, Ermold W (2001) PHC: a global ocean hydrography with a high quality Arctic Ocean. J Climate 14:2079–2087

Stroeve J, Serreze M, Fetterer F, Arbetter T, Meier W, Maslanik JKK (2005) Tracking the Arctic's shrinking ice cover: another extreme September minimum in 2004. Geophys Rev Let 32:L04,501

Stroeve J, Holland M, Meier W, Scambos T, Serreze M (2007) Arctic sea ice decline: faster than forecast. Geophys Rev Let 34:L09,501

Thomson N, Sykes J, McKenna R (1988) Short-term ice motion modeling with applications to the Beaufort Sea. J Geophys Res 93:6819–6836

Timmermann H R Goosse, Madec G, Fichefet T, Ethe C, Dulière V (2005) On the representation of high latitude processes in the ORCA-LIM global coupled sea ice–ocean model. Ocean Modelling 8:175–201

Timmermann R, Danilow S, Schröter J, Böning D, Sidorenko C, Rollenhagen K (2009) Ocean circulation and sea ice distribution in a finite element global sea ice–ocean model. Ocean Model 27:114–129

Tremblay LB, Mysak L (1997) Modeling sea ice as a granular material, including the dilatancy effect. J Phys Oceanogr 27:2342–2360

Vancoppenolle M, Fichefet T, Goosse H, Bouillon S, Madec G, Morales Maqueda MA (2009) Simulating the mass balance and salinity of Arctic and Antarctic sea ice. 1. Model description and validation. Ocean Model 27:33–53

Wadhams P, Davis N (2000) Future evidence of ice thinning in the Arctic Ocean. Geophys Res Lett 27:3973–3975

Wang L, Ikeda M (2004) A Lagrangian description of sea ice dynamics using the finite element method. Ocean Model 7:21–38

Wessel P, Smith W (1996) A global, self-consistent, hierarchical, high-resolution shoreline database. J Geophys Res 101:8741–8743

White L, Deleersnijder E, Legat V (2008) A three-dimensional unstructured mesh finite element method shallow-water model, with application to the flows around an island and in a wind-driven, elongated basin. Ocean Model 22:26–47

Wilchinsky A, Feltham D (2006) Modelling the rheology of sea ice as a collection of diamond-shaped floes. J Non-Newton Fluid Mech 138:22–32

Yakolev N (2003) Coupled model of ocean general circulation and sea ice evolution in the Arctic Ocean. Izv Atmos Ocean Phys 39:355–368

Zhang J, Walsh J (2006) Towards a seasonally ice-covered Arctic Ocean: scenarios from the IPCC AR4 model simulations. J Climate 19:1730–1747

Ocean Dynamics (2010) 60:1559–1580
DOI 10.1007/s10236-010-0353-8

A study of tides and currents in Cook Strait, New Zealand

Roy A. Walters · Phillip A. Gillibrand · Rob G. Bell · Emily M. Lane

Received: 23 April 2010 / Accepted: 20 October 2010 / Published online: 10 November 2010

Abstract Greater Cook Strait (GCS) lies between the North and the South Islands of New Zealand. Its location at the convergence of the Pacific and Indo-Australian tectonic plates leads to interesting bathymetry with an adjacent shallow shelf and deep ocean trench as well as numerous crossing faults and complex shoreline geometry. Our purpose in this study is to examine tides and currents in GCS and, in particular, identify the major forcing mechanisms for the residual currents. Toward this end, we use an unstructured-grid numerical model to reproduce the tides and currents, verify these results with observations and then use the model to separate the various forcing mechanisms. The physical forcing includes nonlinear generation from tides and tidal currents, differences in sea level between the Pacific Ocean and Tasman Sea boundaries, density variations, wind stress and river discharge into GCS. Each of these mechanisms is important in different areas.

Keywords Tides · Currents · Residual currents · Cook Strait · Shallow water equations · Unstructured grid

Responsible Editor: Vincent Legat

R. A. Walters (✉) · P. A. Gillibrand · E. M. Lane
National Institute for Water and Atmospheric Research,
P.O. Box 8602, Christchurch, New Zealand
e-mail: rawalters@shaw.ca

P. A. Gillibrand
e-mail: p.gillibrand@niwa.co.nz

E. M. Lane
e-mail: e.lane@niwa.co.nz

R. G. Bell
National Institute for Water and Atmospheric Research,
P.O. Box 11 115, Hamilton, New Zealand
e-mail: r.bell@niwa.co.nz

Present Address:
R. A. Walters
Ocean/River Hydrodynamics, 6051 Hunt Rd.,
Victoria BC, Canada

1 Introduction

The two main islands of New Zealand (North and South Islands) present a formidable north–south oriented barrier of approximately 1,400 km to the prevailing west to southwest winds and oceanic currents, with the Greater Cook Strait (GCS) the only gap in the bathymetry and topography. GCS is wider and larger on the western side comprising Golden Bay, Tasman Bay and the South Taranaki Bight, leading through the Narrows (only 24 km wide) and widening again on the eastern side, especially beyond Cape Palliser and Cape Campbell (Fig. 1). In terms of bathymetry, the northwestern section of GCS is a relatively flat shelf with depths down to 100 m. This topography contrasts markedly with the southeastern section where the head of the Cook Strait Canyon falls off rapidly to 1,000 m and joins the head of the 3,000-m-deep Hikurangi Trench. These unique features of GCS have long been a fascination to physical oceanographers (strong convergent tidal flows in the context of a subtropical oceanic convergence zone), marine geologists (convergence of two crustal plates, underwater landsliding and highly

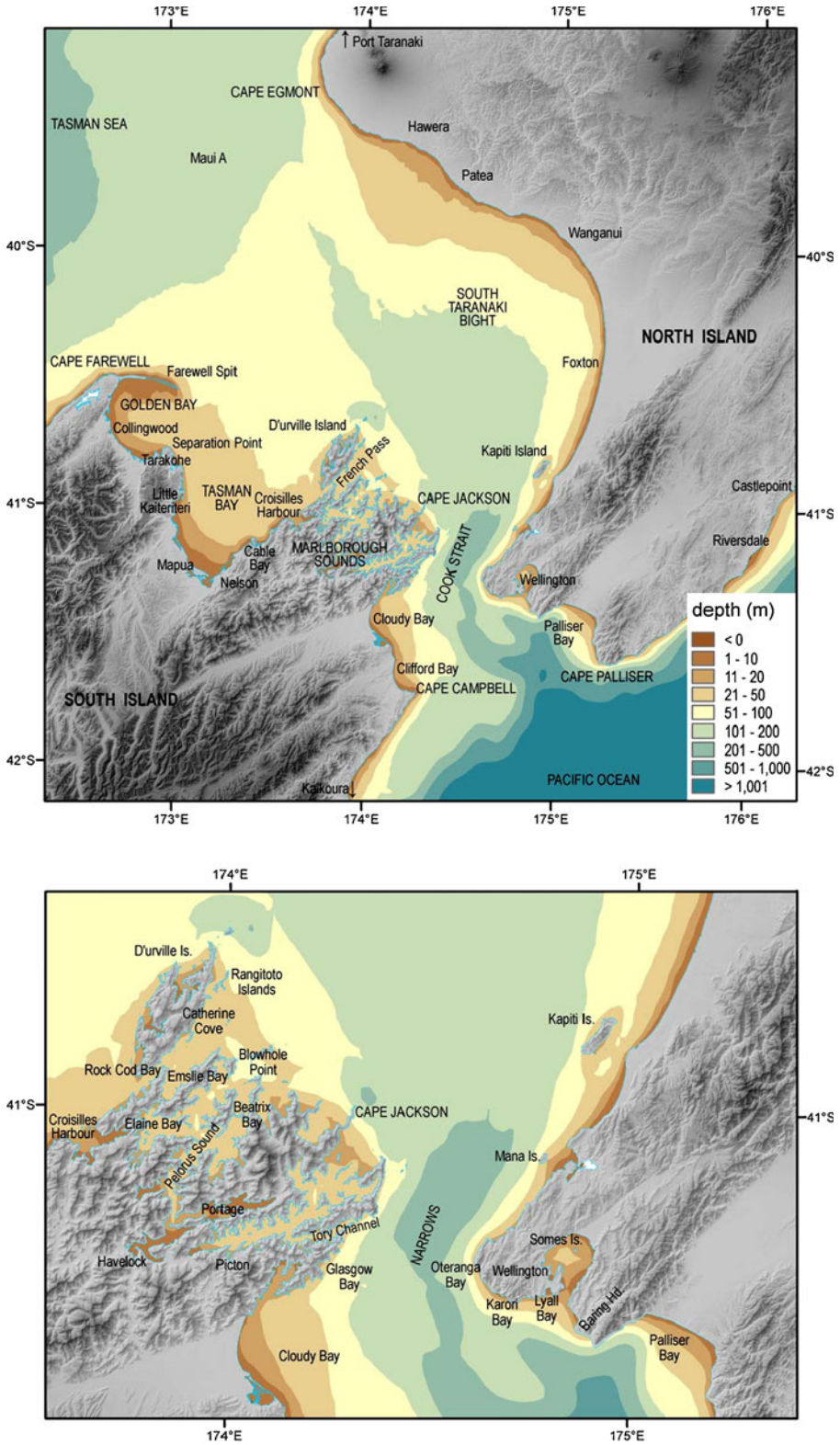

Fig. 1 Index map for Cook Strait with a closer view of the Narrows

erosive currents) and meteorologists (wind funnelling and associated high waves). Extensive reviews of early research studies of GCS are given by Heath (1985) and Harris (1990).

The oceanography of Cook Strait is of particular interest because of the complex interactions between the strong tidal and wind forcing and the regional oceanography. The tides across Cook Strait, for example,

are dominated by a large phase difference between opposite ends of about 145° (or a time lag of 5 h) for the lunar semi-diurnal (M_2) tide, of which 100° phase change occurs over the short 40-km narrowest section. This leads to large currents and interesting tidal dynamics created by the hour-glass planform shape of GCS. The dynamics of the strait are further complicated by the strong wind forcing and the density gradients introduced by the regional oceanography, both of which are described in more detail in the next section. The sub-tidal (residual) circulation resulting from the interactions between these components is the principal focus of this paper.

Numerical modelling of the tidal hydrodynamics of GCS has been previously undertaken by Heath (1974) and Bowman et al. (1980) using depth-averaged (2-D) finite-difference models on regular grid sizes of 7 and 8 km, respectively. Walters et al. (2001) applied a 2-D finite element (spectral in time) model to simulate the tides of the entire New Zealand region on an unstructured grid, which enabled higher resolution to be placed in critical areas like Cook Strait. This particular model is ideal for tides in general but does not provide an accurate estimation of advection and hence tidally generated residuals. Henry and Foreman (2001) further analysed these model results for tidal currents to show that currents associated with a tidal constituent can be represented by the sum of two vectors: a component in phase with the surface oscillation (which accounts for the mean energy flux) and a component in quadrature with the surface oscillation. These components have properties in common with progressive and standing waves, respectively. Studies of the hydrodynamics of sub-domains of the GCS region have also been undertaken by Proctor and Hadfield (1998) for Pelorus Sound (a 3-D baroclinic model on a regular grid) and recently by Tuckey et al. (2006), who applied a 2-D model to Tasman and Golden Bays.

In this paper, we present high-resolution three-dimensional model simulations of features of the tidal hydrodynamics for the entire GCS, including Marlborough Sounds and Tasman/Golden Bays, on an unstructured grid. For a multiply-connected domain with highly variable spatial scales such as GCS, an unstructured grid approach provides an efficient and accurate method to resolve the important spatial details. Our intention is to provide an analysis of the broad-scale features of the circulation and forcing for the residual currents in GCS. The model setup and its verification using numerous tide height and current measurements are described in the next sections and are followed by an analysis of the tidal and residual current characteristics of GCS.

2 Setting

Greater Cook Strait is located adjacent to the boundary between the Indo-Australian and Pacific plates, a setting that defines the bathymetry of the strait. The relatively flat, shallow shelf area to the northwest sits firmly on the Indo-Australian plate; the sea-floor gradually drops to the broad Challenger Plateau which extends northwestwards at a depth of around 1,000 m. In contrast, in the eastern half of the channel, the seabed plunges steeply into the Hikurangi Trough which marks the plate boundary (Fig. 1). A major fault zone passes through the eastern part of GCS resulting in northeast trending reefs and the generally complicated geometry of Marlborough Sounds (between Tasman and Cloudy Bays). Through the central part of the strait then, currents are squeezed through a narrow but deepening (from west to east) channel.

Wind patterns in New Zealand are dominated by the eastward propagation of atmospheric pressure systems, and the prevailing wind is generally from a direction between southwest and northwest. The landmass of New Zealand presents a significant obstacle to westerly air flows, and as a result, considerable funnelling of winds occurs through the Cook Strait channel, with particularly strong accelerations through the Narrows. The other dominant air flow is from the southeast (Harris 1990), which is steered through the Narrows as a southerly before emerging as a low-level jet and veering northwestwards over northern GCS. The overall pattern is one of winds being strongly topographically steered through Cook Strait producing predominantly along-channel wind stress. During summer, sub-tropical cyclones may also pass over or either side of New Zealand, usually travelling southeastwards from the generation area in the tropics to the north of the Tasman Sea and further increasing the wind stresses exerted on Cook Strait.

The spatial distribution of tides around New Zealand reflects the basin-scale pattern of amphidromes set up by the global tide. The major tidal constituents can be placed in three groups: the longer period semi-diurnals M_2 and N_2, the shorter period semi-diurnals S_2 and K_2 and the diurnal constituents K_1, O_1, P_1 and Q_1 (Walters et al. 2001). The M_2 and N_2 constituents have amphidromes northwest and southeast of New Zealand. This pattern creates a Kelvin wave that propagates counterclockwise around the landmass, and from the perspective of Cook Strait, the tidal forcing at the western and eastern entrances is approximately out of phase. On the other hand, the S_2 and K_2 constituents can be described as a plane wave that propagates from the northeast to the southwest. This pattern results in

a Kelvin wave that travels southward along the west coast and has a small amplitude on the east coast that connects with an amphidrome southeast of New Zealand (Walters et al. 2001).

The tides around New Zealand are dominated by the semi-diurnal species and in particular the M_2 constituent which propagates around the NZ coast once during each period. The M_2 tidal wave propagates in from either end of Cook Strait, creating a standing wave component with a node on the northern side of the Narrows that is manifest as a virtual (degenerate) amphidrome at Oteranga Bay (Walters et al. 2001). Tidal heights have now been measured at over 30 sites around GCS by various agencies, considerably extending the database used by Bowman et al. (1980) to verify a tidal model. Previously reported analyses of recording current meter deployments have concentrated on measuring currents in the Narrows section (Heath 1986; Vennell and Collins 1991; Vennell 1994, 1998a). Heath (1986) analysed Aanderaa current meter records from 1982 to 1983 measured 3–5 m above the seabed in three locations across the Narrows, with an additional current meter at 50 m above seabed at two of these sites. Daily-mean currents were dominated by 2–4-week periods associated with through-strait winds, with a residual flow to the south on the western side (South Island) and a northerly residual current on the eastern side (North Island), where the peak tidal currents were also stronger (Heath 1986). The Vennell studies analyzed the vertical structure of the primary tidal currents using acoustic Doppler current profilers (ADCPs) in a shipboard bottom-tracking mode and also subsequently moored on the seabed at three locations (Vennell 1994, 1998a). Results show that the semi-diurnal tidal current speed doubled from 0.7 m s^{-1} on the South Island side to 1.4 m s^{-1} near the North Island coast, with little variation in magnitude over the water column. However, bottom currents led the surface velocity by approximately 10° in phase due to bottom friction. In parallel with these field measurements, Vennell (1998b) developed a diagnostic barotropic model for a short, narrow, channel of variable depth. He found that the phase of the cross-sectional mean velocity is only dependent on the relative amplitude and phase of the surface oscillations at the ends of the narrow channel, explaining the much lower phase change of 12° for M_2 currents through the Narrows as opposed to a 100° phase change for the M_2 sea level.

In addition to the central Narrows in Cook Strait, tidal currents are enhanced through several channels in the Marlborough Sounds, in particular through French Pass which separates d'Urville island from the South Island mainland. Stevens et al. (2008) recently measured spring tide currents of 4 m s^{-1} using moored current meters, drifters and microwave radar. These are the fastest tidal flows that have been measured in New Zealand.

Though tides and winds provide strong forcing of currents in the strait, the circulation is also influenced by the regional oceanography. Cook Strait sits at the confluence of several ocean currents (e.g. Heath 1985). To the west, the Westland and D'Urville Currents transport warm saline surface and sub-surface subtropical water from the Tasman Sea into the strait. During the frequent periods of southerly alongshore winds, cold deep water upwells along the west coast of the South Island (Heath and Gilmour 1987) and is transported into the strait on the D'Urville Current as discrete patches of cold water (Harris 1990). On the eastern fringes of Cook Strait, the southward flowing East Cape Current, transporting subtropical water, meets the cool low-salinity water of the northward flowing Southland Current (Heath 1972). Cool water from the Southland Current penetrates into the strait as far as the Narrows, where it meets the southwestward flowing D'Urville Current (Heath 1971); the convergence is marked by a persistent thermal front (Barnes 1985). The deep canyon is filled with subtropical water from the East Cape Current (Heath 1971). In this paper, we use a modelled climatology of water temperature and salinity (Hadfield et al. 2007) to investigate the mean circulation driven by spatial variations in water density through the strait.

3 Observations

3.1 Tide height and current

Calibration and verification of model predictions against field measurements are a critical component of the modelling process to ensure the results are credible. Because the tidal amplitude and phase change rapidly within the Cook Strait region, particularly in the narrow strait, a primary focus of the tidal data analysis for this study was collating a much larger database of tidal height and currents than has been used in previous modelling studies to better constrain the modelled tide characteristics. Tidal height data were obtained from various sources with records ranging from 2 weeks to several years. Spikes were replaced by linear interpolation, and months with gaps of more than 5 days were excluded. Tidal analysis was undertaken using t-tide (Pawlowicz et al. 2002) to produce amplitudes and phases (UTC) of tidal constituents and their associated 95% confidence intervals (CI). The metadata for each

of the 34 sea-level gauge sites shown in Fig. 2 are listed in Table 1. Quality assurance on multi-year records was assessed by comparing amplitude and phases from single-year analyses and accepting those results that were consistently within a similar 95% CI. Tidal amplitudes and phases derived from a subset of gauges on the open coast of less than 1 year were compared with the SW Pacific regional tidal model of Walters et al. (2001). This comparison revealed a good match with the semi-diurnal lunar (M_2) tide. However, for shorter records of a few months or less, there was often an unsatisfactory match with the semi-diurnal solar (S_2) and lunar elliptical (N_2) tides because neighbouring tides (in frequency space) were not being resolved by the tidal analysis, particularly K_2 and NU_2, respectively, and therefore affecting the results for the primary tides. Consequently, inference of these neighbouring tidal constituents (Foreman 1978) was undertaken in a second phase of tidal analyses, using where possible a longer record (greater than 7 months) from a nearby gauge or alternately sourced from the tidal model of Walters et al. (2001), to compute the ratios of amplitude (inferred tide/primary tide) and phase differences (primary tide–inferred tide) for each site. Fortunately, these amplitude ratios and phase differences tend to be spatially slowly varying, and therefore, the inference procedure is not overly sensitive to the estimates of these parameters except in the case of an amphidrome (which required some care due to presence of a virtual M_2 amphidrome landward of Oteranga Bay). In this way, inferred estimates of amplitude and phase for S_2 and N_2 tide heights on records shorter than 7 months were used to considerably expand the number of sea-level stations that could be used to calibrate or verify the high-resolution model of Greater Cook Strait.

Tidal currents (a vector quantity) were analysed in a similar way using t-tide, with the results presented for each tide species in ellipse form scribed by the velocity vector head. Ellipse features computed from the current vector components were the major and minor semi-axis, inclination of the semi-major axis anticlockwise from east and phase at the most northerly point of the ellipse (which may be peak ebb or flood for different stations) or in the case of localised topographic steering, aligning the phase and inclination with the relevant tidal stream (ebb or flood) in the model. Inference was used for these relatively short current-meter records to estimate K_2 and NU_2 (to better

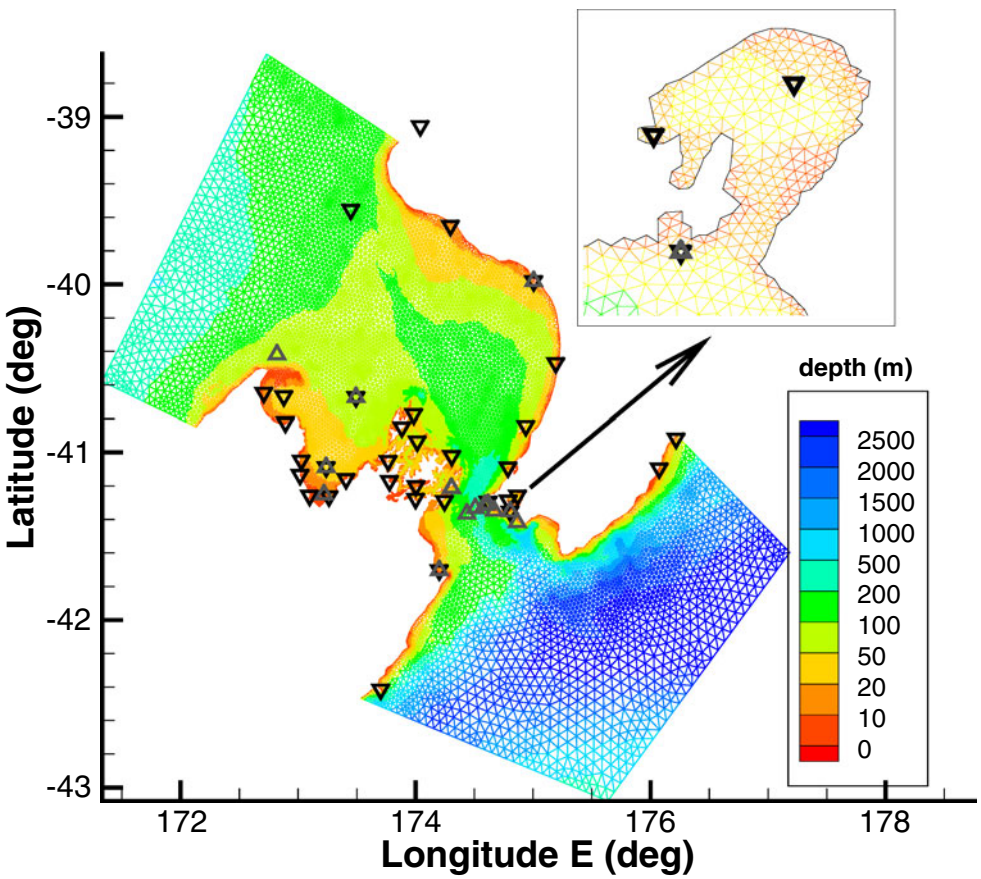

Fig. 2 Locations of the sea level (∇) and current meter (Δ) observation sites in GCS. The grid is colored by depth and a zoomed view of Wellington Harbour is in the *inset*

Table 1 Sea-level gauge field stations used for the tide-height analyses and for comparison with model results

No.	Location	Latitude	Longitude	Days	Data source
1	Kaikoura	−42.4150	173.703	1,460	NIWA
2	Wellington	−41.2860	174.789	1,095	WRC
3	Somes Is.	−41.257	174.868	1,900	WRC
4	Lyall Bay	−41.3519	174.8043	31	NIWA
5	Kapiti Is.	−40.8420	174.938	1,095	NIWA
6	Mana Is.	−41.0892	174.7864	14	NIWA
7	Riversdale	−41.0950	176.074	365	NIWA
8	Castlepoint	−40.9167	176.2167	365	LINZ
9	Little Kaiteriteri	−41.0480	173.027	1,460	TDC
10	Tarakohe Wharf	−40.8224	172.8922	237	TDC
11	Collingwood	−40.647	172.708	70	NIWA
12	Mid Golden Bay (GB)	−40.6675	172.8783	213	NIWA
13	Outer Tasman Bay (CS)	−40.6730	173.4928	137	NIWA
14	South Tasman Bay (TB)	−41.0873	173.2388	108	NIWA
15	Cable Bay	−41.1580	173.4080	33	LINZ
16	Port Nelson	−41.2617	173.2632	315	Port Nelson
17	Port Taranaki	−39.0550	174.040	365	Port Taranaki
18	Hawera	−39.6500	174.2967	64	NIWA
19	Wanganui	−39.9811	175.005	80	NIWA
20	Foxton	−40.4700	175.1933	?	NIWA
21	Oteranga Bay (1)	−41.2937	174.6165	?	LINZ
22	Oteranga Bay (2)	−41.2937	174.6165	30	LINZ
23	Rangitoto Is.	−40.774	173.984	39	LINZ
24	Blowhole Pt.	−40.934	174.014	35	LINZ
25	Cape Jackson	−41.021	174.308	30	LINZ
26	Glasgow Bay	−41.289	174.246	42	LINZ
27	Clifford Bay	−41.7028	174.2017	95	NIWA
28	Picton	−41.2760	174.0000	?	LINZ
29	Havelock	−41.170	173.778	38	LINZ
30	Portage	−41.200	174.003	38	LINZ
31	Maui A	−39.5548	173.448	27	Fugro
32	Elaine Bay	−41.05	173.7667	92	LINZ
33	Catherine Cove	−40.85	173.8833	92	LINZ
34	Rock Cod Bay	−40.930	173.8367	90	LINZ

The sites are shown in Figs. 1 and 2

estimate S_2 and N_2, respectively) in a similar manner to the tide heights analysis but based on ratios and phase differences for the dominant u or v velocity component from the large-domain tidal model of Walters et al. (2001) to circumvent the presence of the virtual amphidrome. This tidal model is ideal for calculating linear tides but does not provide an accurate estimation for advection or residual currents forced by advection. Hence, the model is valuable in calculating the amplitude ratios and phase differences necessary for inference. Table 2 lists the sites where data from recording current-meter deployments (Fig. 2) were used to generate tidal current ellipses including some sites in the Narrows, particularly two sites (C16 and C17) where an upward-looking ADCP was deployed (sites L1 and L2 in Vennell and Collins 1991). These ADCP datasets were particularly useful for evaluating the model results and for providing verification data on features such as the phase advance of the near-bed M_2 velocity relative to the upper-column tidal velocity of around 15 min at the eastern site (C16) and 30 min at the mid-strait site (C17) and the tidal stream changes that occur earlier on the eastern side of the strait.

3.2 Wind

Several wind monitoring stations are located around Greater Cook Strait. However, constructing wind fields by interpolation from these sites at different altitudes is problematic, particularly when there is such strong topographic steering. Instead, we used winds derived from an operational mesoscale meteorological forecast model for the New Zealand region, NZLAM-12 (Lane et al. 2009). The model is a regional implementation of the UK Meteorological Office Unified Model (Davies et al. 2005) with a local spatial resolution of 12 km. Data from a network of observational platforms are assimilated into the weather forecasts using a 3DVAR system (Lorenc et al. 2000). Easterly and northerly components of wind velocity at 10 m height above

Table 2 Current meter locations used for the analyses and for comparison with model results

No	Location	Latitude	Longitude	Depth (m)	Height (m)	Days
C1	Outer Tasman Bay CS1	−40.6730	173.4928	58	48	137
C2	Outer Tasman Bay CS2	−40.6730	173.4928	58	48	138
C3	Sth Tasman Bay TB1	−41.0873	173.2388	32	23	196
C4	Sth Tasman Bay TB2	−41.0873	173.2388	32	23	140
C5	Nelson	−41.2494	173.2206	7	2.8	70
C6	Farewell Spit	−40.4183	172.8200	78	10	128
					55	117
C7	Wanganui	−39.9811	175.0050	10	5.5	80
C8	Karori Bay	−41.348	174.670	19	3	28
C9	Lyall Bay	−41.3519	174.8043	20	14	31
					2	31
C10	Tory Channel Ent N1	−41.2113	174.3043	31	17	33
C11	Tory Channel Ent S1	−41.2129	174.3067	28	13	33
C12	Clifford Bay	−41.7029	174.2015	11	1.4	95
C13	Baring Head	−41.4158	174.8652	20	1	42
C14	Oteranga Bay: CSB1	−41.3067	174.6150	35	1	14
C15	Narrows: CSA2	−41.3650	174.4367	120	85	60
C16	Narrows: ADCP L1	−41.3133	174.5883	178	7	30
					55	30
					99	30
					131	30
C17	Narrows ADCP L2	−41.3317	174.5083	225	10	31
					54	31
					98	31
					130	31
					174	31

Depth is water depth at the site and height is meter placement above the bottom

sea level are provided at hourly intervals to force the hydrodynamic model. The wind vector data are linearly interpolated in space and time to obtain values at the nodal locations of the finite-element ocean grid at each model time step (Lane et al. 2009).

For the present study, the wind-driven circulation for July 2008 was investigated, a period which included two winter storms and may therefore be expected to have experienced relatively strong wind-driven flows through Cook Strait. The hydrodynamic model was spun-up for 4 days using the NZLAM-forecast winds, and the simulation proper ran from 1 to 31 July.

4 Model

The model, RiCOM, is based on the three-dimensional shallow water equations which are derived from the Reynolds-averaged Navier–Stokes equations by using the hydrostatic assumption and the Boussinesq approximation. For incompressible flows, the continuity equation (incompressibility constraint) is

$$\nabla \cdot \mathbf{u} + \frac{\partial w}{\partial z} = 0, \tag{1}$$

and the momentum equation expressed in nonconservative form is

$$\frac{D\mathbf{u}}{Dt} + f\hat{\mathbf{z}} \times \mathbf{u} + g\nabla\eta - \frac{\partial}{\partial z}\left(A_v \frac{\partial \mathbf{u}}{\partial z}\right) - \nabla \cdot (A_h \nabla \mathbf{u}) - F = 0, \tag{2}$$

where the coordinate directions (x, y, z) are aligned in the east, north and vertical directions; $\mathbf{u}(x, y, z, t)$ is the horizontal velocity with components (u, v); $w(x, y, z, t)$ is the vertical velocity; f is the Coriolis parameter; $\hat{\mathbf{z}}$ is the upward unit vector; $\eta(x, y, t)$ is the distance from a reference elevation (mean sea level in this case) to the free surface; g is the gravitational acceleration; A_v and A_h are the kinematic vertical and horizontal viscosities, respectively; ∇ is the horizontal gradient operator $(\partial/\partial x, \partial/\partial y)$ and $F(x, y, z)$ are the body forces that include density forcing, form drag and other processes. The calculation of horizontal density gradients with sigma coordinates follows the method proposed by Stelling and van Kester (2008) where the gradient is interpolated between element centroids.

The free surface equation is derived by vertically integrating the continuity equation and using the kine-

matic free surface and bottom boundary conditions (Pinder and Gray 1977),

$$\frac{\partial \eta}{\partial t} + \nabla \cdot \left(\int_h^\eta \mathbf{u} dz \right) = 0, \tag{3}$$

where $h(x, y)$ is the bottom elevation measured from a reference elevation such that $H(x, y, t)$ is the total water depth given by $H = \eta - h$.

Lateral boundary conditions for Eqs. 1–3 generally fall into two categories: conditions at open (sea) boundaries and conditions at solid (land) boundaries. At open boundaries, sea level η, radiation conditions or a combination of these two are generally set. In addition, discharge may be specified for river or other inflow. At land boundaries, the normal component of velocity vanishes so that $(\mathbf{u} \cdot \hat{\mathbf{n}}) = 0$ where $\hat{\mathbf{n}}$ is the unit normal. If the flow is viscous, then stress conditions also need to be specified.

At the bottom boundary, either a stress condition is specified or $\mathbf{u}|_h = 0$ and the bottom boundary layer needs to be resolved by the vertical grid placement. Since we are not investigating the detailed structure of the bottom boundary layer, we have adopted the first approach because it is more efficient. The bottom stress, $\boldsymbol{\tau}_b$, is given by

$$\frac{\boldsymbol{\tau}_b}{\rho} = C_D |\mathbf{u}| \mathbf{u} \qquad (z = h), \tag{4}$$

where ρ is a reference density and C_D is a bottom drag coefficient.

The surface boundary condition is a wind stress $\boldsymbol{\tau}_s$, given by

$$\frac{\boldsymbol{\tau}_s}{\rho} = C_D^s \left(\frac{\rho_a}{\rho} \right) |\mathbf{W}| \mathbf{W} \qquad (z = \eta), \tag{5}$$

where ρ_a is air density, $\mathbf{W}$ is the 10-m wind velocity and C_D^s is a drag coefficient. Further details about the implementation of the bottom and wind stress are given in Section 4.3.

4.1 Time discretization

The equations are discretized in time using a semi-implicit method such that some of the terms in the equations are treated implicitly and some terms explicitly. The equations are evaluated in the time interval $\Delta t = t^{n+1} - t^n$ where the superscript denotes the time level. The distance through the time interval is given by the weight θ. Using this approximation, the free surface equation becomes

$$\frac{\eta^{n+1} - \eta^n}{\Delta t} + \nabla \cdot \left[\theta \left(\int_h^\eta \mathbf{u}^{n+1} dz \right) + (1 - \theta) \left(\int_h^\eta \mathbf{u}^n dz \right) \right] = 0 \tag{6}$$

The material derivative in the momentum equation is approximated using Eulerian–Lagrangian methods (ELM) or semi-Lagrangian methods (SLM) which take advantage of the simplicity of Eulerian methods and the enhanced stability and accuracy of Lagrangian methods (Staniforth and Côté 1991). With ELM, only the term at time level n in the material derivative ($\mathbf{u}^*$) is evaluated at the foot of the trajectory. With SLM, all terms evaluated at time level n are evaluated at the foot of the trajectory so the entire equation is treated in a Lagrangian sense.

Using an ELM approach, the momentum equation becomes

$$\frac{\mathbf{u}^{n+1} - \mathbf{u}^*}{\Delta t} - \frac{\partial}{\partial z} \left(A_v \frac{\partial \mathbf{u}^{n+1}}{\partial z} \right) + f \hat{\mathbf{z}} \times \mathbf{u}^n = -g \nabla \left[\theta \eta^{n+1} + (1 - \theta) \eta^n \right] + \nabla \cdot (A_h \nabla \mathbf{u}^n) + F \tag{7}$$

where the superscripts $n + 1$ and n denote variables evaluated at the fixed nodes in the Eulerian grid at times t^{n+1} and t^n, respectively. At each time step, the velocity is integrated backwards with respect to time to determine where a particle starts at time t^n in order to arrive at a grid node at time t^{n+1}. This point is referred to as the foot of the trajectory (Staniforth and Côté 1991). The superscript $*$ denotes a variable evaluated at the foot of the trajectory. The material derivative, the first term, thus has a very simple form.

The proper choice of tracking and interpolation methods is critical for maintaining accuracy with SLM and ELM approximations (Staniforth and Côté 1991). Here, a power-series approach in time is used to calculate the trajectory, and the global-quadratic interpolation method is used to evaluate the velocity at the foot of the trajectory (Walters et al. 2008).

Note that the vertical viscosity term is treated implicitly to remove stability constraints whereas the horizontal viscosity term is treated explicitly to avoid the solution of a large matrix equation. This introduces a stability constraint that is generally not important (Casulli and Cattani 1994). In this study, the horizontal stress term is neglected.

The Coriolis term is integrated using a third-order Adams–Bashforth scheme that is both efficient and avoids any major stability constraints (Durran 1991; Walters et al. 2009b).

4.2 Space discretization

The governing equations, Eqs. 6 and 7, are approximated in space using finite element methods. The interpolation functions on each element are a piecewise constant approximation for sea level, η, and a linear approximation for velocity such that the normal velocity is constant on each edge. This element is known as the RT_0 element (Raviart and Thomas 1977), and the discrete three-dimensional shallow water equations are developed in Miglio et al. (1999) and Walters et al. (2009a).

Both the continuity equation and free surface equation reduce to a finite volume approximation using the RT_0 element. With the hydrostatic approximation, the continuity equation is a diagnostic equation for vertical velocity and is not considered further. The free surface equation can be written in a finite volume form that conserves mass both locally and globally,

$$A_e \frac{{\eta_e}^{n+1} - {\eta_e}^{n}}{\Delta t} = \int_{\Gamma_e} \left[\theta\left(H^{n+1}{u_n}^{n+1}\right) + (1-\theta)\left(H^n {u_n}^n\right)\right] d\Gamma_e, \tag{8}$$

where subscript e denotes the value for a specific element, subscript n denotes a normal component, A is the area in the x–y plane and Γ_e is the bounding surface along the edges of the element. The last term has been converted from a volume divergence to a surface integral using the Gauss divergence theorem.

Likewise, the momentum equation becomes (Miglio et al. 1999; Walters et al. 2009a)

$$\mathbf{A}\mathbf{u}_n^{n+1} = \mathbf{G} - g\theta\,\Delta t \mathbf{\Delta}\eta^{n+1}\mathbf{Z}, \tag{9}$$

where $\mathbf{A}$ is a block tridiagonal matrix that contains the mass matrix and the implicit terms arising from the vertical component of stress, $\mathbf{u}_n$ is the vector of normal velocity on each edge, $\mathbf{G}$ contains all the explicit terms and the known component of the implicit terms and the last term is the implicit part of the surface pressure gradient term such that $\mathbf{\Delta}\eta$ is a vector of the difference in element values across an edge and $\mathbf{Z}$ is a vector of depth increments. The mass matrix is lumped using the technique for finite-element node point integration described in Walters et al. (2009a). This is equivalent to discretizing the pressure gradient term using the distance between element centroids as measured orthogonal to an element edge.

The tridiagonal matrix at each velocity node can be inverted and u_n^{n+1} substituted into the free surface equation (Eq. 8) to derive a discrete form of a wave equation that has only η at the $n+1$ time level. Typically, η^{n+1} is solved for first, followed by a calculation of $\mathbf{u}_n^{n+1}$ from Eq. 9. This method provides an efficient means to solve the equations. The specific forms of the matrices are shown in detail in Walters et al. (2009a).

Note that in the free surface equation and the discrete wave equation, water depth H is a factor in all the side flux terms. When $H = 0$ (i.e. the side is dry), there is automatically no water flux through that side. When all sides of an element are dry, the water level is stationary in time. Hence, wetting and drying are implemented without any special treatment but are limited by the Courant number.

4.3 Model setup

The grid for Greater Cook Strait is composed of triangular elements in the horizontal, and at each vertex, the depth is discretized using σ-coordinates with 15 vertical levels. The three-dimensional elements are then pie-shaped with five faces. The horizontal grid was generated using the software described in Henry and Walters (1992) using a combination of clusters and a frontal marching method. First the boundaries were discretized, and then the interior elements were generated with a smooth transition in size from the detailed land boundary to the coarse open boundary. This procedure allows a detailed representation of the land geometry while retaining model accuracy and efficiency. Grid resolution is based on a range of requirements including depth and gradient of the sea floor, complexity of the local geometry and the necessity of resolving certain bays and passages.

This particular grid has about twice the resolution of the regional grid used for sea level forecasting (Lane et al. 2009) but still does not resolve some of the detailed coastal geometry. The intention was to create a background grid that describes the broad features in GCS and can be refined in specific areas for local studies. This grid contains 24,100 nodes and 44,423 elements (two dimensions) or 361,500 nodes and 621,922 elements (three dimensions). Edge lengths range from approximately 100 m (in some sections of Marlborough sounds) to 10 km (in the open ocean).

Bathymetric data are derived from several sources. Underlying bathymetry from the EEZ grid includes the NIWA seabed database for offshore, the shelf and Cook Strait (swath bathymetry in some areas) supplemented by digitizing contours and soundings in the nearshore waters and the Marlborough Sounds from the Land Information NZ Hydrographic Charts (NZ 463, 4633, 614–615, 6142, 6151–6154, 6212).

The open boundaries approximately follow M_2 tide amplitude and phase isopleths from the tidal model of

Walters et al. (2001). This procedure seemed to give a more accurate representation of the tidal forcing. Boundary conditions include the amplitude and phase of the six largest tidal components, M_2, S_2, N_2, O_1, K_1, Q_1 and mean sea level, Z_0. For this study, K_2 and P_1 were neglected so that tidal analysis could be performed with 30 days of simulated time rather than 180 days. These two constituents are relatively small. The amplitudes and phases of the tidal constituents were derived from a regional scale tidal model (Walters et al. 2001). The amplitude for Z_0 was obtained from another model for the regional ocean climatology (Hadfield et al. 2007).

In addition, temperature and salinity that characterise the mean state of GCS were obtained from another model (Hadfield et al. 2007). From these variables, density was calculated and used in diagnostic calculations to evaluate the influence of density forcing on the residual currents. With diagnostic calculations, the density field is variable in space but held constant in time. The objective is to not let the density field evolve in time since this necessitates application of boundary conditions that we do not possess. Our approach meets our objective of assessing the relative influence of density variations.

The value for vertical viscosity coefficient, A_v, can be determined in a number of ways including empirical formulations and turbulence closure schemes. GCS is relatively shallow and well mixed so we have chosen a mixing length formulation given by

$$A_v = \kappa u_*(z-h)(1-(z-h)/H) \quad (h+z_0 \le z \le \eta). \tag{10}$$

where $\kappa = 0.4$ is von Karman's constant, u_* is shear velocity and z_0 is roughness height.

Using results for flows over rough topography, the bottom drag is specified in terms of C_D and a roughness height, z_0. The governing equations (Eqs. 1–3) are spatially averaged over unresolved geometric features such as bottom roughness elements or structures imbedded in the flow (Nikora et al. 2007). Many experiments have shown that the spatially averaged velocity profile below roughness crests is approximately linear (Nikora et al. 2002). Near the roughness crests, the flow transitions into a logarithmic layer and the matching condition on stress give a value of $C_D \approx 0.02$. We have adopted this value so that the free parameter in bottom stress now becomes the roughness height, z_0 (Walters and Plew 2008). Sensitivity tests with C_D (0.01–0.02) and z_0 (0.1–1 m) show that the results for sea level are relatively insensitive to these parameters because of negative feedback between them and the bottom velocity. Velocity profiles have greater shear in the lower water column for larger values of C_D and z_0 as is expected. Overall, $C_D = 0.02$ and z_0 between 0.1 and 1 m gave the most reasonable fits between observed and modelled velocity profiles. The results presented here are for $z_0 = 0.1$ m. Other than inspection of the velocity profiles, no formal tuning of the model parameters was done.

Finally, the surface drag coefficient is given by Wu (1982)

$$C_D^s = .001(0.8 + 0.065|W|), \tag{11}$$

where $|W|$ is the 10-m wind speed.

5 Results

The results fall within two general categories: comparison of model results with observations of tidal heights and currents and evaluation of the different forcing mechanisms for the residual current. The comparisons are presented in the next two subsections, and the analysis of the residual currents is contained in the third subsection.

Evaluation of the residual currents can be determined by several approaches. One is to calculate tides and currents with all forcing mechanisms and use time-series analysis methods to separate the tidal, mean and low frequency components. We have found that a more straightforward method is to calculate the tides first, evaluate the tidal residuals, then consider the aperiodic wind forcing and finally examine the influence of density variations. The dominant energy input is through the tides so that the residual components interact weakly and hence the procedure is tractable.

In using the latter approach, the results in Sections 5.1, 5.2, 5.3.1 and 5.3.2 are from calculations with only tides; Section 5.3.3 considers the difference between a calculation with tides only and with both tide and wind effects and finally, Section 5.3.4 compares tide alone with tide and diagnostic density effects.

5.1 Tides

As noted earlier, the basin-scale forcing for the M_2 and N_2 constituents is very different than for the S_2 and K_2 constituents. For the former pair, the tidal elevations (boundary conditions) at the east and west ends of GCS are approximately out of phase so that there is a node point of the oscillation in between. For the latter pair, there is an oscillatory boundary on the west but approximately null amplitude on the east.

The M_2 constituent has a spatial pattern with an amphidrome in the narrow section west of Oteranga Bay (Fig. 3), an area of intense tidal currents ranging up to approximately 2.5 m s^{-1}. The tidal wave characteristics (standing or progressive wave) are determined by the phase difference between sea level and velocity with a phase difference of 90° indicating a standing wave and no phase difference indicating a progressive wave. Examining a transect from the western entrance to the eastern entrance of GCS, the tidal wave west of the Narrows is found to be a standing wave whereas in the Narrows it is a progressive wave. East of the Narrows it is a mixture of a progressive and standing waves. This pattern would be expected from a physical situation with two large oscillating basins connected by a narrow constriction. The tidal wave is amplified in Tasman and Golden Bays which have some of the largest tidal ranges in New Zealand. The spatial pattern for the N_2 constituent is similar to the M_2 but with reduced amplitude.

On the other hand, the spatial pattern for the S_2 constituent is very different (Fig. 4). The tidal wave west of the Narrows is a standing wave forced by the Kelvin wave travelling southward along the west coast with a secondary north–south mode in South Taranaki Bight. There is no amphidrome such as with the M_2 constituent. The S_2 amplitude is small east of the Narrows and hence the phase tends to be inaccurate. The spatial pattern for the K_2 constituent is similar to the S_2 but has much smaller amplitude (see Walters et al. 2001).

In general, the diurnal species have rather small amplitude with the land mass of New Zealand at or near an amphidrome (Walters et al. 2001). However, in some locations, the currents can be more significant presumably from the occurrence of shelf waves. We do not consider the diurnal constituents in GCS because of their small amplitude.

Many of the sea level observation sites are located sightly outside the grid because of small differences between the modelled and actual shoreline. The amplitude and phase at these sites are interpolated using an inverse distance weighting of the three closest points. The comparison between observations and model results for the M_2, S_2 and N_2 constituent at the sites in Table 1 is shown in Table 3. Note that the different sites have a wide range of data quality and grid resolution. Table 1 gives a general indication of data quality with the long-term open-ocean sites considered to be more accurate than the short-record shallow-bay sites.

Typical differences in amplitude and phase are a few centimetres and a few degrees, with larger differences

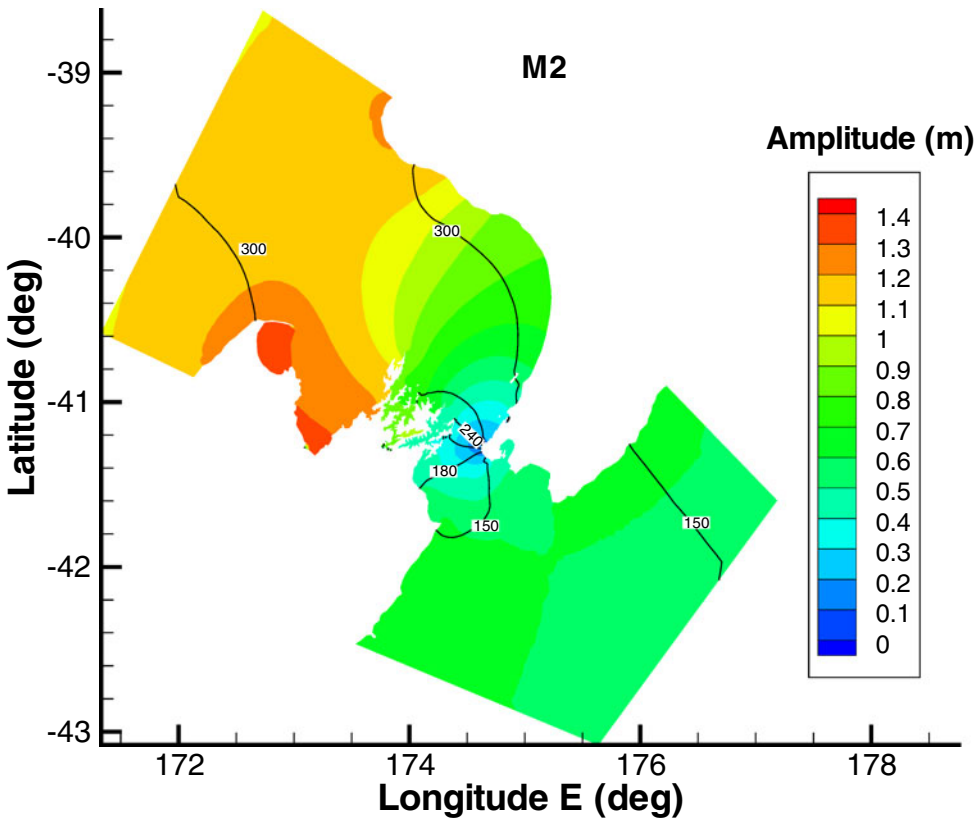

Fig. 3 Amplitude (*filled contours*) and phase (*lines*) for the M_2 tidal constituent

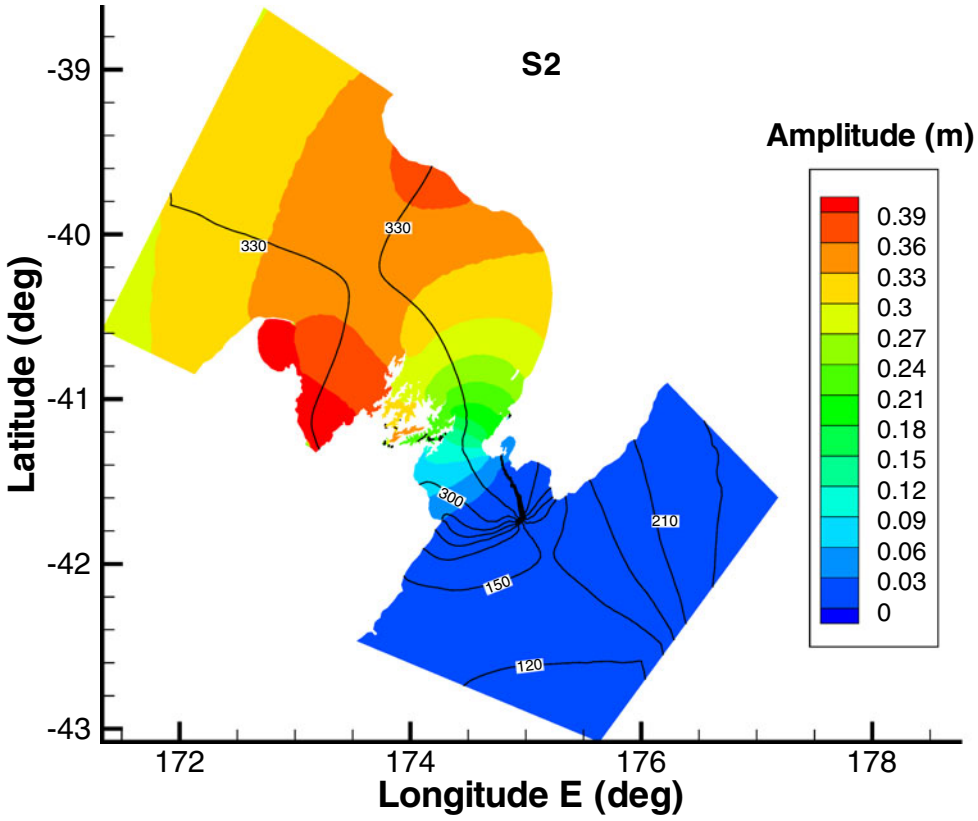

Fig. 4 Amplitude (*filled contours*) and phase (*lines*) for the S_2 tidal constituent

in certain areas and at sites with poorer spatial resolution. In the eastern part of GCS, the phases tend to be too small which mostly seems to be related to boundary conditions. The S_2 amplitudes are also quite small there, and the phases are more uncertain. The amplitude and phase for the M_2 and N_2 near the amphidrome have larger errors because of the large gradients in this area and the sensitivity of the location. Havelock (29) undergoes wetting and drying in the model whereas it is connected to a channel that stays wet at the field site. The Maui A site (31) is an open water site that shows excellent results.

5.2 Currents

The amplitude and phase of tide elevation depends on the broad-scale features in geometry and bathymetry. On the other hand, tidal currents respond strongly to local features. Hence, the accuracy of the currents is, in general, not as good as the tidal height. The accuracy can be improved by refining the grid around the current meter locations in order to resolve the local topography. One advantage of the unstructured grid is that this can be done locally without requiring a refinement of the entire grid. However, we have not done so in this study as our intention is to represent the broad-scale features.

The spatial pattern in depth-averaged current speed is shown in Fig. 5. The speed is calculated as the root mean square (rms) velocity averaged over the 30-day simulation period which contains two spring-neap tidal cycles. Wind stress and baroclinic forcing are not included in this simulation but are considered later. The large current speeds in the Narrows near the M_2 amphidrome are readily apparent. In general, the spatial pattern in speed reflects the influence of bottom topography and shoreline geometry. The speed is relatively large along the centerline of GCS where the depth is the largest. Large speeds also occur where the flow is accelerated around Farewell Spit and across the shoal area in the South Taranaki Bight. The relatively large speeds in the eastern corner of the grid are a part of the southward flowing East Cape Current and are forced by the residual (Z_0) boundary conditions.

The comparison between observed currents and modelled currents for the M_2, S_2 and N_2 constituent at the sites in Table 2 is shown in Table 4. Most of the current meter sites relate to specific local investigations near the shoreline, except for the Tasman Bay and Narrows deployments. Depending on the grid resolution, there is a wide range in model accuracy at these sites. Poor results occur at C8 and C14 which are close to shore in indented rock platforms that are expressions of two major tectonic faults (perpendicular to the main

Table 3 Results for the tide height analyses and model results

Site	M2						N2						S2					
	A_o	G_o	A_m	G_m	ΔA	ΔG	A_o	G_o	A_m	G_m	ΔA	ΔG	A_o	G_o	A_m	G_m	ΔA	ΔG
1	0.674	144	0.645	140	−0.029	−4	0.150	115	0.149	111	−0.001	−4	0.024	174	0.016	130	−0.008	−44
2	0.488	148	0.485	142	−0.003	−6	0.126	117	0.119	110	−0.007	−7	0.024	349	0.034	5	0.010	16
3	0.491	150	0.485	142	−0.006	−8	0.125	120	0.118	110	−0.007	−10	0.028	353	0.034	5	0.006	12
4	0.473	149	0.460	141	−0.013	−8	0.121	114	0.113	109	−0.008	−5	0.027	327	0.033	1	0.006	34
5	0.554	292	0.538	299	−0.016	7	0.088	295	0.090	306	0.002	11	0.263	337	0.265	340	0.002	3
6	0.373	282	0.339	292	−0.034	10	0.059	283	0.049	313	−0.010	30	0.216	333	0.216	341	0.000	8
7	0.566	160	0.622	152	0.056	−8	0.137	130	0.143	122	0.006	−8	0.026	220	0.012	225	−0.014	5
8	0.633	160	0.617	153	−0.016	−7	0.144	124	0.141	124	−0.003	0	0.027	228	0.012	233	−0.015	5
9	1.320	288	1.307	288	−0.013	0	0.240	277	0.230	277	−0.010	0	0.408	329	0.400	330	−0.008	1
10	1.353	288	1.335	290	−0.018	2	0.251	277	0.237	279	−0.014	2	0.410	329	0.400	331	−0.010	2
11	1.376	291	1.357	291	−0.019	0	0.254	278	0.241	280	−0.013	2	0.455	331	0.406	332	−0.049	1
12	1.343	290	1.330	291	−0.013	1	0.250	279	0.236	280	−0.014	1	0.419	332	0.398	332	−0.021	0
13	1.188	287	1.168	288	−0.020	1	0.225	277	0.208	278	−0.017	1	0.392	328	0.363	328	−0.029	0
14	1.329	286	1.293	287	−0.036	1	0.247	275	0.227	277	−0.020	2	0.431	330	0.397	328	−0.034	−2
15	1.288	289	1.282	286	−0.006	−3	0.239	273	0.225	276	−0.014	3	0.399	326	0.394	327	−0.005	1
16	1.365	284	1.324	287	−0.041	3	0.249	275	0.233	276	−0.016	1	0.423	326	0.407	328	−0.016	2
18	1.195	290	1.204	291	0.009	1	0.233	276	0.235	280	0.002	4	0.338	319	0.336	320	−0.002	1
19	1.097	287	1.090	302	−0.007	15	0.203	279	0.213	296	0.010	17	0.380	312	0.373	331	−0.007	19
20	0.810	300	0.831	303	0.021	3	0.149	291	0.152	301	0.003	10	0.337	336	0.335	337	−0.002	1
21	0.763	301	0.721	302	−0.042	1			0.128	303	0.128				0.309	339	0.309	
22	0.207	211	0.161	187	−0.046	−24	0.040	180	0.038	106	−0.002	−74	0.113	345	0.117	340	0.004	−5
23	0.204	216	0.161	187	−0.043	−29	0.021	153	0.038	106	0.017	−47	0.098	345	0.117	340	0.019	−5
24	0.871	274	0.828	274	−0.043	0	0.147	262	0.129	265	−0.018	3	0.288	323	0.292	324	0.004	1
25	0.739	270	0.764	271	0.025	1	0.129	258	0.114	262	−0.015	4	0.262	325	0.280	324	0.018	−1
26	0.545	258	0.507	258	−0.038	0	0.078	241	0.060	246	−0.018	5	0.220	324	0.214	324	−0.006	0
27	0.435	198	0.404	193	−0.031	−5	0.079	156	0.071	148	−0.008	−8	0.112	301	0.103	312	−0.009	11
28	0.610	167	0.595	163	−0.015	−4	0.121	129	0.127	130	0.006	1	0.049	262	0.042	275	−0.007	13
29	0.490	257	0.438	257	−0.052	0	0.057	233	0.045	249	−0.012	16	0.226	328	0.212	329	−0.014	1
30	0.961	293	0.871	274	−0.090	−19	0.133	286	0.129	265	−0.004	−21	0.333	331	0.322	328	−0.011	−3
31	0.961	286	0.927	275	−0.034	−11	0.145	275	0.136	267	−0.009	−8	0.345	322	0.344	331	−0.001	9
32	1.158	294	1.169	295	0.011	1	0.214	284	0.226	285	0.012	1	0.325	323	0.342	325	0.017	2
33	0.825	282	0.818	272	−0.007	−10	0.120	263	0.121	263	0.001	0	0.327	321	0.301	326	−0.026	5
34	0.817	283	0.825	272	0.008	−11	0.124	267	0.126	263	0.002	−4	0.308	321	0.290	324	−0.018	3
35	1.220	286	1.318	288	0.098	2	0.215	267	0.231	277	0.016	10	0.410	328	0.404	329	−0.006	1

A_o observed amplitude (metres), G_o observed phase (degree relative to UTC), A_m model amplitude, G_m model phase, ΔA $A_m - A_o$, ΔG $G_m - G_o$

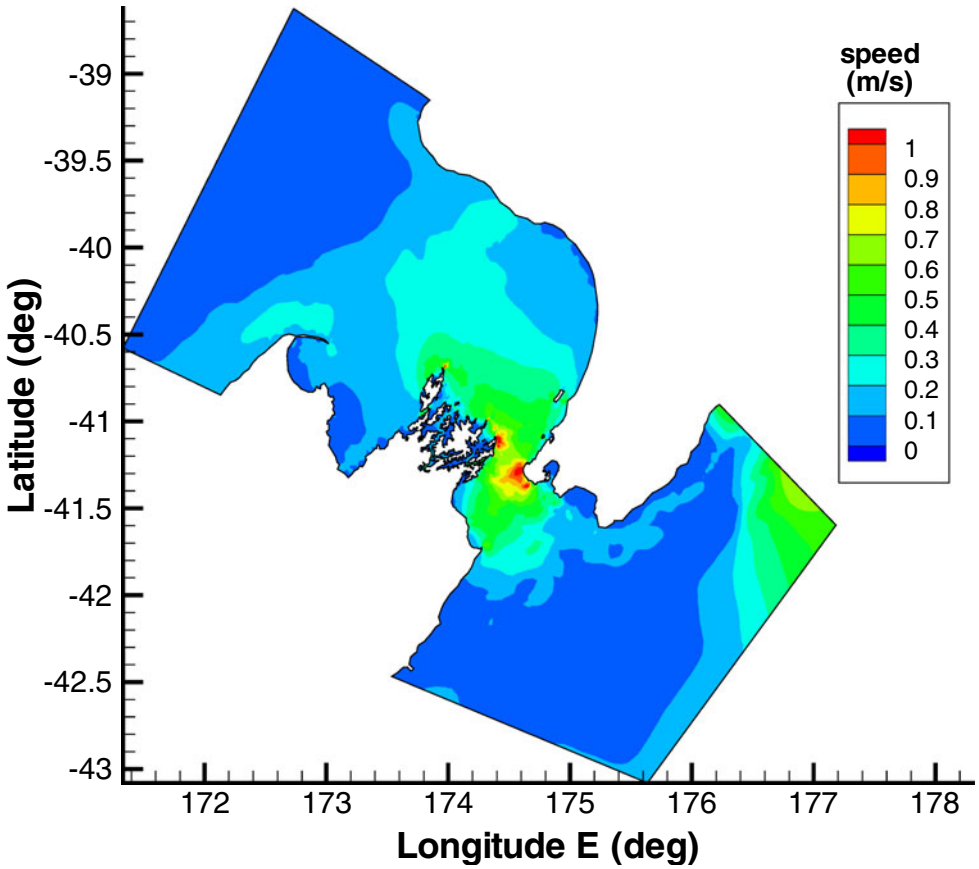

Fig. 5 Average (rms) current speed calculated from the depth-averaged velocity

flow) and near where a flow separation is located. The model does not resolve these indented submarine rock platforms very well so that the observations are weaker than the model results. There are also poor results at Lyall Bay, near the entrance to Wellington Harbour. The flood tide into Wellington Harbour starts early passing by Lyall Bay while the currents are still ebbing at Baring Head. This is a case of inadequate grid resolution to treat the complexity of the flow in this area. Other unpublished tests with the model have shown that with adequate resolution, the model can represent the flow separation accurately. In future work, the area from Wellington Harbour Entrance to the Narrows will be resolved in greater detail.

For the deeper sites such as C16 and C17 in the Narrows, both the magnitude and vertical profile of velocity are reasonably accurate. In particular, the increase in current speed from south to north and the phase lead at the north site is in accordance with the results from Vennell (1994). The phase lead for the model at the bottom measurement height is 6° and 8° at C16 and C17, as opposed to 7° and 16° that were observed. The phase lead depends on the form of the vertical viscosity formulation and is dependent on z_0. A larger value for z_0 at C17 would provide better agreement.

5.3 Residual currents

One of the long-term goals in examining GCS hydrodynamics is to evaluate the solute and particulate transport mechanisms and their magnitudes. One of these mechanisms is the transport by the mean (residual) current; the other is the dispersive transport that involves correlations between currents and concentrations. As a starting point, we examine the components of the residual currents.

A natural method to define residual currents is through a time average of the flux terms in the free-surface equation (Eq. 3). Denoting a time average with angle brackets $<>$

$$< H\mathbf{U} >= H_0\mathbf{U}_0 + < \eta'\mathbf{U}' > \qquad (12)$$

where $H = H_0 + \eta'$ and $\mathbf{U} = \mathbf{U}_0 + \mathbf{U}'$, the subscript 0 denotes the residual part and the prime denotes the perturbation in time. Hence, the currents considered here are Eulerian residual currents. Accordingly, the mean of the depth averaged currents are plotted in

Table 4 Comparison of observed and modelled current ellipses

Site	M2 observations			M2 model output			M2 errors		N2 observations			N2 model output			N2 errors		S2 observations			S2 model output			S2 errors	
	AM	Inc	*G*	AM	Inc	*G*	ΔA	ΔG	AM	Inc	*G*	AM	Inc	*G*	ΔA	ΔG	AM	Inc	*G*	AM	Inc	*G*	ΔA	ΔG
C1	0.224	16	41	0.255	23	36	0.031	−5	0.054	13	24	0.053	19	16	−0.001	−8	0.051	44	71	0.060	53	66	0.009	−5
C2	0.215	17	43	0.255	23	36	0.040	−7	0.054	10	22	0.053	19	16	−0.001	−6	0.045	43	66	0.060	53	66	0.015	0
C3	0.094	47	30	0.133	49	26	0.039	−4	0.018	42	22	0.024	47	14	0.006	−8	0.031	56	71	0.044	57	68	0.013	−3
C4	0.093	48	32	0.133	49	26	0.040	−6	0.019	41	24	0.024	47	14	0.005	−10	0.029	59	63	0.044	57	68	0.015	5
C5	0.111	70	15	0.147	60	21	0.036	6	0.024	58	12	0.026	60	9	0.002	−3	0.031	52	67	0.047	62	64	0.016	−3
C6	0.173	6	122	0.189	1	107	0.016	−15	0.032	6	78	0.036	2	66	0.004	−12	0.034	168	53	0.035	169	52	0.001	−1
	0.191	177	317	0.215	179	294	0.024	−23	0.031	1	106	0.041	1	71	0.010	−35	0.057	173	44	0.041	169	58	−0.016	14
C7	0.102	133	203	0.153	129	203	0.051	0	0.024	144	180	0.033	131	183	0.009	3	0.028	131	228	0.029	126	196	0.001	−32
C8	0.307	188	170	1.178	146	173	0.871	3	0.055	205	163	0.263	146	154	0.009	3	0.068	192	192	0.189	145	191	0.121	−1
C9	0.119	156	114	0.191	177	176	0.072	62	0.022	125	43	0.036	179	149	0.014	106	0.019	145	94	0.030	173	184	0.011	90
	0.091	162	116	0.143	174	170	0.052	54	0.013	134	42	0.024	176	143	0.011	101	0.013	152	79	0.022	171	179	0.009	100
C10	1.550	16	351	1.057	24	337	−0.493	−14	0.287	15	332	0.187	24	320	−0.100	−12	0.324	15	14	0.219	24	6	−0.105	−8
C11	1.355	18	338	1.047	24	337	−0.308	−1	0.284	13	294	0.185	24	320	−0.099	26	0.327	11	353	0.215	25	6	−0.112	13
C12	0.141	142	124	0.154	149	128	0.013	4	0.022	139	84	0.028	150	101	0.006	17	0.033	139	143	0.024	149	145	−0.009	2
C13	0.490	124	207	0.411	131	207	−0.079	0	0.066	124	211	0.078	131	183	0.012	−28	0.088	124	218	0.080	130	218	−0.008	0
C14	0.250	139	130	0.860	114	168	0.610	38				0.159	114	147						0.147	114	182		
C15	0.678	102	194	1.104	93	209	0.426	15	0.144	106	187	0.226	93	192	0.082	5	0.169	112	246	0.252	94	233	0.083	−13
C16	0.979	102	183	0.952	104	182	−0.027	−1	0.190	111	160	0.183	104	163	−0.007	3	0.269	102	211	0.179	104	201	−0.090	−10
	1.344	103	187	1.339	104	186	−0.005	−1	0.241	108	164	0.260	103	167	0.019	3	0.348	102	212	0.267	103	205	−0.081	−7
	1.450	102	189	1.447	104	188	−0.003	−1	0.266	98	164	0.283	103	168	0.017	4	0.373	100	210	0.295	103	206	−0.078	−4
	1.460	100	190	1.470	104	188	0.010	−2	0.288	90	163	0.287	103	168	−0.001	5	0.371	99	209	0.301	103	206	−0.070	−3
C17	0.847	101	194	0.975	100	202	0.128	8	0.160	106	173	0.198	99	187	0.038	14	0.229	101	230	0.210	100	230	−0.019	0
	1.053	103	199	1.241	100	206	0.188	7	0.194	107	178	0.255	100	191	0.061	13	0.275	101	233	0.280	100	233	0.005	0
	1.203	99	204	1.315	100	208	0.112	4	0.214	104	183	0.271	100	192	0.057	9	0.294	98	233	0.302	100	235	0.008	2
	1.278	97	207	1.349	100	209	0.071	2	0.238	99	187	0.271	100	192	0.041	6	0.310	97	235	0.314	100	236	0.004	1
	1.291	96	210	1.370	99	210	0.079	0	0.278	95	191	0.284	100	194	0.006	3	0.311	99	236	0.321	100	237	0.010	1

The multiple values for C6, C9, C16 and C17 are at the depths given in Table 2

AM is the major axis (metres per second), *Inc* is the inclination (° anticlockwise from east)), *G* is the phase (degree relative to UTC), $\Delta A = \mathrm{AM(model)} - \mathrm{AM(observation)}$, $\Delta G = G(\mathrm{model}) - G(\mathrm{observation})$

Figs. 6 and 7 as is appropriate for barotropic tidal currents. For the current profiles, the velocity at each level is processed with harmonic analysis, and the Z_0 component is extracted as the residual.

Residual currents are generated by a variety of tidal and nontidal mechanisms. Tidal generation mechanisms arise from the time-averaged nonlinear terms in the governing equations (Eqs. 1–3; Walters and Werner 1991). One of these is the tidal stress which is the time-averaged advection term, and another is the wave drift which is the time-averaged nonlinear term in the free surface equation. Nontidal mechanisms include river discharge into GCS, throughflow due to surface elevation differences at the east and west open boundaries of GCS, wind stress and baroclinic currents due to density variations. In the following subsections, we attempt to evaluate the relative importance of the tidal and nontidal mechanisms.

River discharge into GCS can have a strong local effect on currents and stratification (Sutton and Hadfield 1997; Proctor and Hadfield 1998). For instance, in a study of Beatrix Bay off Pelorus Sound, Sutton and Hadfield (1997) noted a two-layer stratified water structure. The conceptual picture is that freshwater discharge peaks spread rapidly along Pelorus Sound and then are mixed vertically. However, these are local effects that are mostly confined to an area of freshwater influence. For the most part, these local areas are below the spatial scales we are addressing here. Hence, they are not dealt with in this broader study.

5.3.1 Tidal mechanisms

The process denoted as wave drift results from a net mass flux during the passage of a tidal wave and a return residual flow to balance this flux (Walters and Werner 1991). The forcing mechanism is the time-averaged nonlinear term in the free surface equation given by $\nabla \cdot < \eta \bar{\mathbf{u}} >$ where the angle brackets denote a time average and the overbar denotes a depth average. The forcing is the strongest when η and $\mathbf{u}$ are in phase (a progressive wave) and have a large magnitude. These conditions are best satisfied in Tasman and Golden Bays and have a broad spatial pattern. The largest currents generated by this mechanism are approximately 0.05 m s^{-1} in French Pass and are created by a sea level setup in Tasman Bay. However, this forcing is not significant anywhere else as compared to other mechanisms in GCS and can be neglected.

Tidal stress is important around headlands and other features where there is flow separation and back-flowing eddies (Geyer and Signell 1991). In these locations, the residual flow is directed along the shoreline

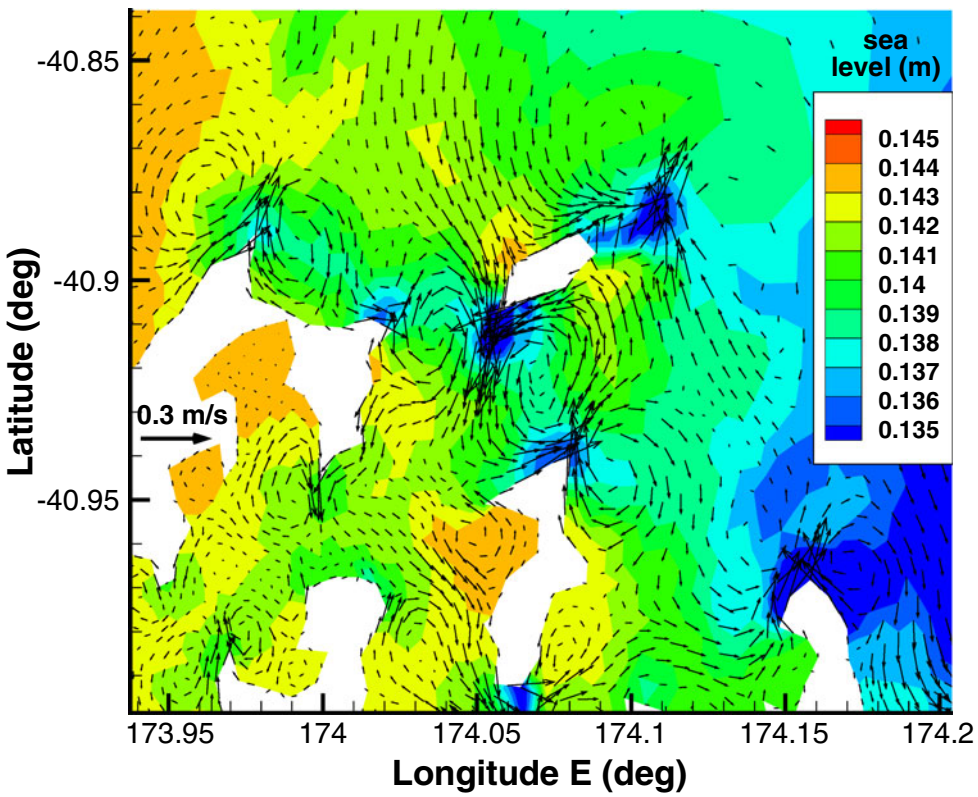

Fig. 6 Residual currents in outer Pelorus Sound. Tidal stress is the dominant forcing mechanism without winds. Grid resolution is 200–500 m around in this area

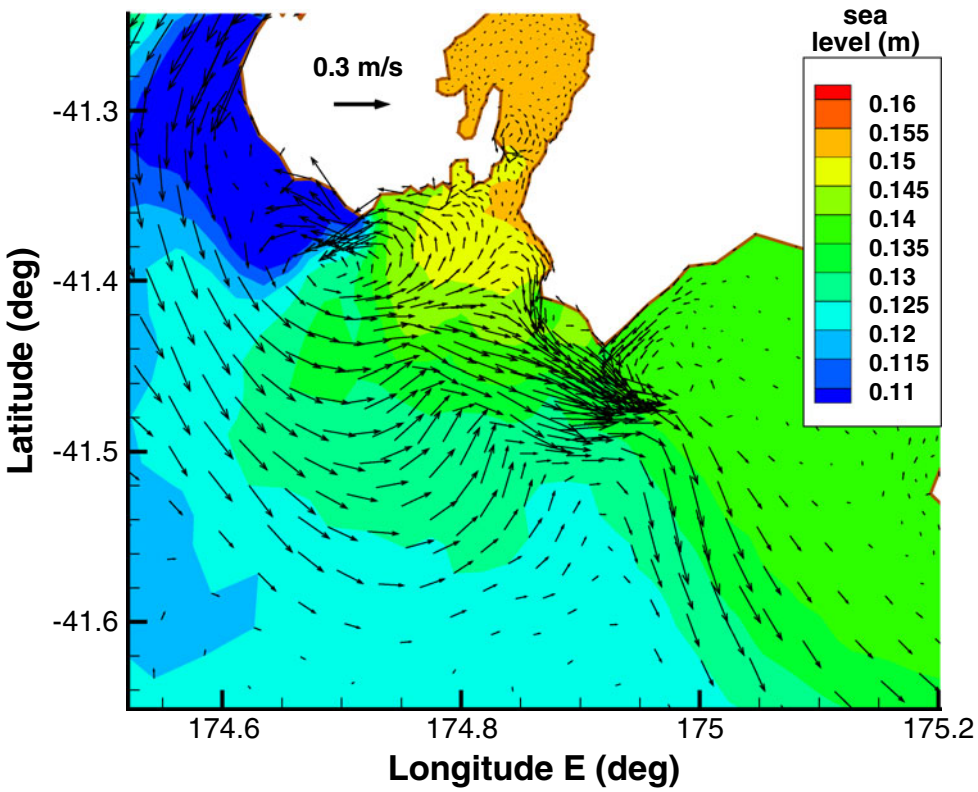

Fig. 7 Residual currents along the northern side of Cook Strait. Tidal stress is significant near the headlands, but throughflow is dominant otherwise (without winds)

toward the headland then out to sea. The flow has relatively small spatial scales so the unstructured grid enables us to better resolve these features. Marlborough Sounds is one area where tidal stress is a major forcing mechanism. This area lies between D'Urville Island and Tory Channel (Fig. 1) and is characterised by a complicated geometry with a number of bays, islands and sounds. Grid resolution in Marlborough Sounds ranges from 100 to 500 m depending on what is required. A typical tidal stress residual flow for outer Pelorus Sound is shown in Fig. 6. Residual current speeds for this mechanism range up to approximately 0.4 m s^{-1}.

5.3.2 Throughflow

A residual flow in GCS from west to east arises because of sea level differences between the Tasman Sea to the west and the Pacific Ocean to the east. At the western entrance, the flow entering or exiting GCS depends on the wind conditions. For the case of the predominant southwesterly winds, the Westland Current flows up the west coast of the South Island and becomes the D'Urville Current when it enters GCS (Harris 1990). For other wind conditions, the flow may enter as a current flowing down the west coast of the North Island or there may be a flow to the west in GCS (Heath 1986). The results presented here show a dominance of the Westland Current flowing past Cape Farewell into GCS and are a consequence of the mean sea surface height that is used for the boundary conditions. After passing Cape Farewell, the flow crosses over to South Taranaki Bight, then strengthens along the north side of GCS, flows along the coast south of Wellington and then turns north at Cape Palliser to follow the east coast of the North Island. The northward flow is examined by Chiswell (2000) where the name Wairarapa Coastal Current is proposed. The calculated flow through GCS is similar to the accepted behaviour of the D'Urville Current. The calculated volumetric transport for the mean flow is 0.61 Sv, where 1 Sv = 10^6 m^3 s^{-1}. This is approximately the accepted value of about 1 Sv for the D'Urville Current, although this value is highly variable in time.

When the inflow at the western entrance passes Farewell Spit, it breaks into a meandering current such as described in Harris (1990). The outflow at the eastern entrance is a combination of the Southland Current passing northward along the east coast of the South Island, the D'Urville Current exiting GCS and meanders from the East Cape Current (Heath 1971). The residual flow on the northern side of Cook Strait is shown in Fig. 7. Around the headlands, tidal stress can be a significant forcing mechanism. Overall, however, the throughflow dominates the residual currents

without winds. Note the northwesterly flow off Wellington and the southeasterly flow otherwise. This pattern is consistent with the current meter results from Heath (1986) for the Narrows and Bell (1991) for the south Wellington coastline and is a consequence of the tidal stress near shore and the throughflow at a distance from shore. Bell (1991) derived residual currents for 13 current meters generally along the 20-m-depth contour along the coast near the Wellington Harbour entrance. These values are in good agreement with the values shown in Fig. 7.

5.3.3 Wind stress

Local winds (as opposed to regional winds that affect throughflow) are highly variable in GCS and hence have a wide range of relative effects. For areas where other forcing mechanisms are weak such as Tasman Bay and Golden Bay, the wind component can be dominant. For areas such as Marlborough Sounds where tidal stress is generally dominant, wind stress can be locally significant in open water away from land boundaries. In the open waters of GCS between Cape Farewell and Cape Palliser, wind stress can be significant at times and on average can generate depth-averaged currents that are comparable to the throughflow.

The effects of wind stress are somewhat difficult to quantify because they are aperiodic and depend on the passage of fronts and weather systems. The greatest effects occur intermittently, when steep fronts and cyclones pass through the area. In calm weather, there may be no effects. A simple average of the wind induced residual currents underestimates their significance by reducing the importance of large, intermittent events. Moreover, wind generated currents are concentrated in the surface layer, unlike tidally generated residual currents and throughflow which extend over the entire water column. Hence, if transport is considered, the importance of wind generated currents is reduced further.

Our approach is to simulate a typical winter period (July 2008) and then approximate the contribution from the wind generated currents. The model was run with and without winds for a 30-day period. Then, the two output files were used to calculate differences in sea level and velocity in hourly steps. Finally, the rms differences were calculated and are taken as an estimate of the wind forced part, i.e. the time-averaged wind-generated currents. For velocity, both the value at the surface and for the depth-average were derived as a rough measure of the importance of the surface wind-driven current. Typically, the surface value is 50% larger than the depth average.

At the west and east entrances to GCS, the time-averaged wind-driven current can be larger than the tidal current. North of Farewell Spit and across the South Taranaki Bight, wind-generated currents (Fig. 8) are the major component of the residual current, followed by throughflow in the form of the D'Urville Current. The wind-generated currents are also dominant north of Marlborough Sounds and south of Wellington (Fig. 8). At the west end of Palliser Bay, the D'Urville Current is slightly larger than the wind-driven component.

In Tasman and Golden Bays, wind-generated currents are the dominant residual component and can be comparable to the tidal currents at times. In general, however, the rms tidal currents are about twice as large as the rms wind-generated currents.

5.3.4 Baroclinic currents

In general, GCS is well mixed in the vertical with respect to density. As noted by Harris (1990), cold core eddies can be advected in through the western entrance with the Westland Current. These eddies would modify the meandering flow derived in our simulations. In addition, Harris (1990) points out a possible entrainment of deep water by the eastern outflow of the D'Urville Current. In order to examine this issue in greater detail, a diagnostic calculation was made using a density field derived from climatology. The density field was derived from model results of Hadfield et al. (2007) which describe the mean state of GCS (Fig. 9).

Two sets of simulations were compared. In the first set, the model simulations included only tides. In the second set, density forcing was added as a diagnostic calculation where density does not change in time. The comparison of the results included calculating the difference between the two data sets, comparing the flow patterns and comparing vertical profiles of velocity at several sites spanning GCS.

The overall results indicate that the throughflow through GCS is dependent on a delicate balance between sea level difference and density difference across GCS. Heath (1986) suggests that this throughflow is variable and can in fact be directed in either a westerly or easterly direction. For the boundary conditions used here, the flow is reduced considerably compared to the case with no density variations discussed previously.

In addition, the results indicate that density forcing has a minimal effect on velocity profiles west of the Narrows. In Tasman and Golden Bays, there is a density current component but the residual speed is small

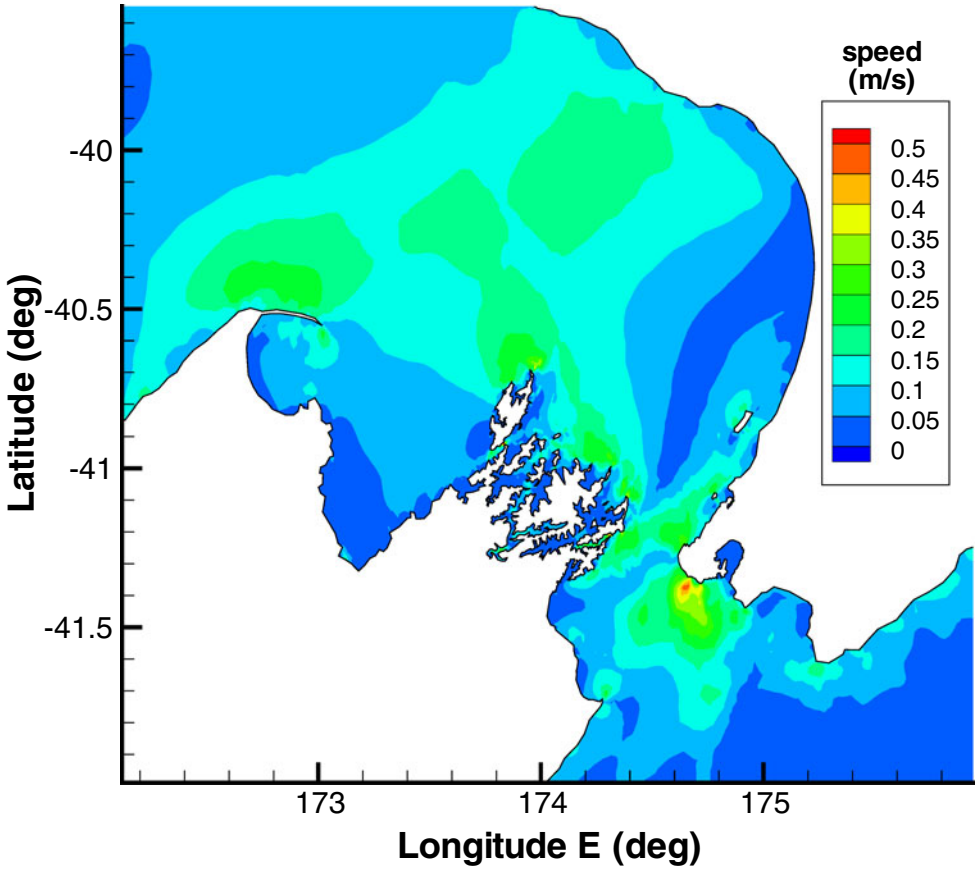

Fig. 8 Wind-generated rms currents (depth average)

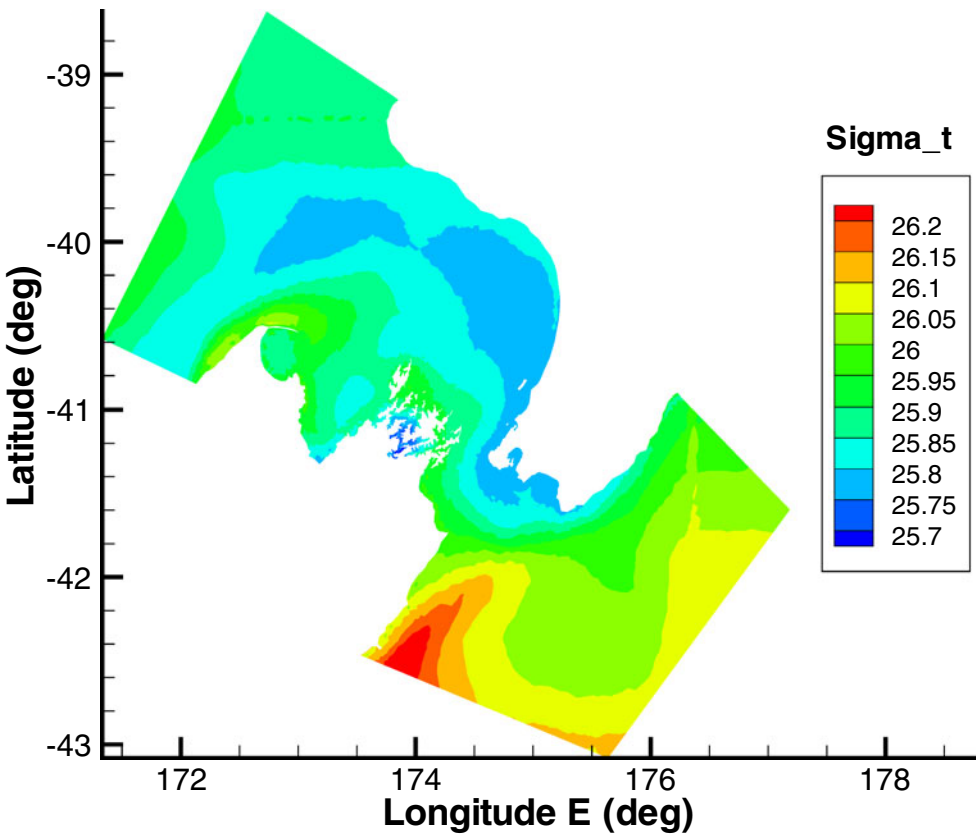

Fig. 9 Surface density distribution used in the diagnostic calculations

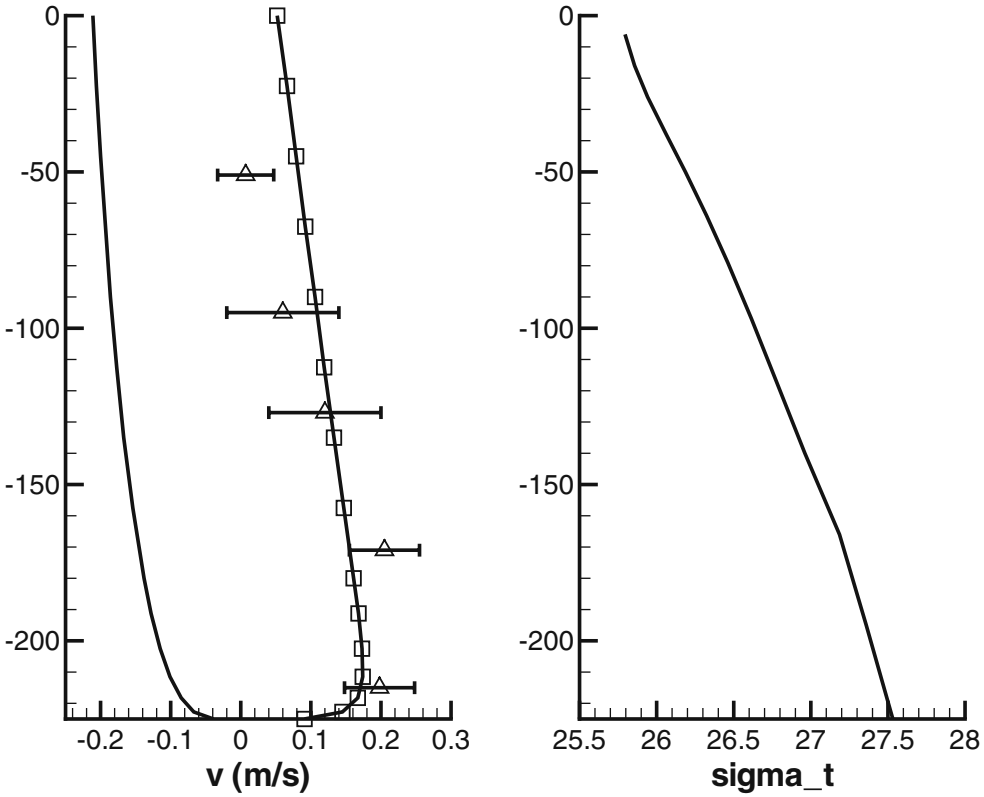

Fig. 10 Profile of northward velocity at current meter site C17 (left). *Line* no density forcing, *line with square* density forcing, *triangle* current meter data, *bars* the range of daily means. Sigma_t is the density with respect to depth at the same point (right). The *vertical axis* is distance from mean sea level in metres

of the order 0.01 m/s. East of GCS, the longshore flow patterns are modified by the density gradients such that the geostrophic currents are dependent on both the boundary conditions and the density gradients. In and to the south of the Narrows, a density current intrudes up the Cook Strait Canyon and displaces the eastward throughflow to the south.

Current meter site C17 is at a saddle point at the head of Cook Strait Canyon, west of Oteranga Bay. Model results at this site with no density forcing display a southward outflow with a typical barotropic velocity profile (Fig. 10). Model results with density forcing show a density current intruding northward with a velocity profile typical of estuarine circulation. The Z_0 component of the C17 data is shown in triangles with bars that indicate the range of daily mean values reported in the observations. The current meter data suggest that the velocity profile retains the same shape but is translated with changes in local winds and setup across GCS.

Note that the ambient conditions for the data and model are different. The data represent conditions for density, winds and setup that occurred during the deployment (1 month). On the other hand, the model uses climatology for the density and setup from two different data sets and does not include winds. Nonetheless, the model and observations are in reasonable agreement which may indicate that the model boundary conditions are within the normal range of observations, or density intrusion indicated by the shape of the vertical velocity profile is relatively invariant at this site.

A useful way to view these results is to examine the surface density distribution (Fig. 9). The Westland Current brings a low density water mass around Farewell Spit into GCS where the D'Urville Current carries this through GCS, along the Wellington Coast, and turns northward up the east coast. This conceptually represents a passive tracer carried by the dominant large-scale residual flow component. In the Narrows, a density current intrudes up Cook Strait Canyon and modifies the barotropic flow pattern. At the east entrance to GCS, there is an interaction between the northward flowing Southland Current, the D'Urville Current and a part of the East Cape Current.

6 Conclusions

Greater Cook Strait has a unique set of features in relation to geometry, location and physical forcing. Its location in the convergence zone between the Pacific and Indo-Australian tectonic plates has a strong influence on topography with a plateau and ocean trench at the entrances to GCS and complex bathymetry due to the

numerous crossing faults. In addition, New Zealand lies across the oceanic sub-tropical convergence and presents a barrier to the prevailing westerly winds. Finally, the tidal forcing shows a large difference between the M_2 and longer period species and the S_2 and shorter period species. All these features affect the flow patterns that are observed in this study.

The geometric complexity of GCS can be well represented using unstructured grid methods. The land boundaries are discretized with more detail than is necessary for the open boundaries. A smooth transition in element size between these boundaries provides an accurate discretization, while the reduction in computational elements compared to a fine-structured grid provides increased efficiency.

Our purpose has been to examine tides and currents in order to assess the importance of the various physical forcing mechanisms. To accomplish this, we have analysed a large number of tide and current observations and used these to assess a numerical model of GCS. An analysis of the model results then provides a means to separate the forcing mechanisms. We believe that this is the first detailed analysis of the residual circulation in GCS and extends the qualitative description by Harris (1990) and earlier authors.

The modelled tides compare well with observations with a typical error of 0.02 m in M_2 amplitude and 8° in phase at the well-resolved sites. The currents depend more strongly on local bathymetry so the agreement with observations is not as good. Even so, the model quantitatively reproduces the spatial pattern and vertical profile of velocity at the well-resolved sites. The model correctly reproduced the large difference in the M_2 and S_2 tide species.

The residual currents show a distinct spatial pattern in the dominant generation mechanisms. Tidal stress is the most significant forcing mechanism around headlands and in the complex geometry of Marlborough Sounds. Typical residual currents for this mechanism range up to 0.5 m s^{-1}. Wave transport (forcing from the nonlinear term in the free surface equation) was negligible everywhere. Throughflow in the form of the D'Urville Current accounted for a significant part of the residual current along the axis of GCS with a transport of 0.61 Sv. This throughflow is dependent on a balance between the sea level gradient and density gradient across GCS. Density forcing had a small effect on the vertical structure of the currents to the west of the Narrows. However, the density gradients had a significant effect on the flows in the east entrance to GCS.

The influence of winds is the greatest in a band from Farewell Spit across South Taranaki Bight where wind-generated currents are the dominant component of the residual current. Wind effects were also important north of Marlborough Sounds, south of Wellington and in Tasman and Golden Bays. Wind-generated currents can be locally important at certain times in places where other mechanisms are weak. Such is the case in Tasman and Golden Bays where tidal currents are relatively weak and the wind driven and tidal currents can be of comparable magnitude at times.

The use of the unstructured grid in this modelling work allowed better resolution in areas of complex geometry such as Marlborough Sounds without the need for nesting grids or for resolving the entire grid to the same level. Although some stations were not as well resolved in this broad brush survey, this shows that, had they been the primary focus of the study, those stations could also be well resolved within the same grid.

Acknowledgements The authors acknowledge Fraser Callaghan for his expertise in setting up the grids, Mark Hadfield for providing the salinity and temperature from which density was calculated, Geoffroy Lamarche for supplying gridded swath bathymetry of Cook Strait and Sanjay Wadhwa for generating Fig. 1 in ArcGIS. The project was partially funded by the New Zealand Foundation for Research, Science and Technology (Contracts C01X0401, C01X0812 and C01X0507).

References

Barnes EJ (1985) Eastern Cook Strait region circulation inferred from satellite-derived, sea-surface, temperature data. N Z J Mar Freshw Res 19:405–411

Bell RG (1991) Oceanography and microbial modelling studies for Wellington and Lower Hutt outfalls. In: Bell RG, Hume TM, Healy TR (eds) Coastal engineering—climate for change. Proceedings of 10th Australasian conference on coastal and ocean engineering, Auckland. DSIR Water Quality Centre Publication no. 21, pp 251–256

Bowman MJ, Kibblewhite AC, Ash DE (1980) M2 tidal effects in greater Cook Strait, New Zealand. J Geophys Res 85(C5): 2728–2742

Casulli V, Cattani E (1994) Stability, accuracy, and efficiency of a semi-implicit method for three-dimensional shallow water flow. Comput Math Appl 27(4):99–112

Chiswell SM (2000) The Wairarapa coastal current. N Z J Mar Freshw Res 34:303–315

Davies T, Cullen MJP, Malcolm AJ, Mawson MH, Staniforth A, White AA, Wood N (2005) A new dynamical core for the Met Office's global and regional modelling of the atmosphere. Q J R Meteorol Soc 131:1759–1782

Durran DR (1991) The third-order Adams–Bashforth method: an attractive alternative to leapfrog time differencing. Mon Weather Rev 119:702–720

Foreman MGG (1978) Manual for tidal currents analysis and prediction. Pacific marine science report 78-6. Institute of Ocean Sciences, Victoria, 70 pp

Geyer WR, Signell RP (1991) Measurements and modelling of the spatial structure of nonlinear tidal flow around a headland. In: Parker BB (ed) Tidal hydrodynamics. Wiley, New York, pp 403–418

Hadfield MG, Rickard GJ, Uddstrom MJ (2007) A hydrodynamic model of Chatham Rise, New Zealand. N Z J Mar Freshw Res 41:239–264

Harris TFW (1990) Greater Cook Strait form and flow. DSIR Marine and Freshwater, Wellington, 212 pp

Heath RA (1971) Hydrology and circulation in central and southern Cook Strait, New Zealand. N Z J Mar Freshw Res 5:178–199

Heath RA (1972) The Southland current. N Z J Mar Freshw Res 6:497–533

Heath RA (1974) The lunar semi-diurnal tide in Cook Strait, New Zealand. Dtsch Hydrogr Z 27:214–224

Heath RA (1985) A review of the physical oceanography of the seas around New Zealand—1982. N Z J Mar Freshw Res 19:79–124

Heath RA (1986) In which direction is the mean flow through Cook Strait New Zealand—evidence for 1 to 4 week variability? N Z J Mar Freshw Res 20:119–137

Heath RA, Gilmour AE (1987) Flow and hydrological variability in the Kahurangi Plume off north-west South Island, New Zealand. N Z J Mar Freshw Res 21:125–140

Henry RF, Foreman MGG (2001) A representation of tidal currents based on energy flux. Mar Geod 24:139–152

Henry RF, Walters RA (1992) A geometrically-based, automatic generator for irregular triangular networks. Commun Appl Numer Methods 9:555–566

Lane EM, Walters RA, Gillibrand P, Uddstrom M (2009) Operational forecasting of sea level height using an unstructured grid ocean model. Ocean Model 28:88–96

Lorenc AC, Ballard SP, Bell RS, Ingleby NB, Andrews PLF, Barker DM, Bray JR, Clayton AM, Dalby T, Li D, Payne TJ, Saunders FW (2000) The Met. Office global three-dimensional variational data assimilation scheme. Q J R Meteorol Soc 126:2991–3012

Miglio E, Quateroni A, Saleri F (1999) Finite element approximation of quasi-3D shallow water equations. Comput Methods Appl Mech Eng 174:355–369

Nikora V, Goring D, Biggs B (2002) Some observations of the effects of micro-organisms growing on the bed of an open channel on the turbulence properties. J Fluid Mech 450: 317–341

Nikora V, McLean S, Coleman S, Pokrajac D, Walters RA (2007) Double-averaging concept for rough-bed open-channel and overland flows: theoretical background. J Hydraul Eng 133(8):873–883

Pawlowicz R, Beardsley R, Lentz S (2002) Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE. Comput Geosci 28:929–937

Pinder GF, Gray WG (1977) Finite elements in surface and subsurface hydrology. Academic, New York

Proctor R, Hadfield M (1998) Numerical investigation into the effect of freshwater inputs on the circulation in Pelorus Sound, New Zealand. N Z J Mar Freshw Res 32:467–482

Raviart PA, Thomas JM (1977) A mixed finite element method for 2nd order elliptic problems. In: Mathematical aspects of the finite element method. Lecture notes in math. Springer, Berlin

Staniforth A, Côté J (1991) Semi-Lagrangian integration schemes for atmospheric models—a review. Mon Weather Rev 119: 2206–2223

Stelling GS, van Kester JATM (2008) On the approximation of horizontal gradients in sigma co-ordinates for bathymetry with steep bottom slopes. Int J Numer Methods Fluids 18:915–935

Stevens C, Sutton P, Smith M, Dickson R (2008) Tidal flows in Te Aumiti (French Pass), South Island, New Zealand. N Z J Mar Freshw Res 42:451–464

Sutton PJH, Hadfield MG (1997) Aspects of the hydrodynamics of Beatrix Bay and Pelorus Sound. N Z J Mar Freshw Res 31:271–279

Tuckey BJ, Gibbs MT, Knight BR, Gillespie PA (2006) Tidal circulation in Tasman and Golden Bays: implications for river plume behaviour. N Z J Mar Freshw Res 40: 305–324

Vennell R (1994) Acoustic Doppler current profiler measurements of tidal phase and amplitude in Cook Strait, New Zealand. Cont Shelf Res 14(4):353–364

Vennell R (1998a) Observations of the phase of tidal currents along a strait. J Phys Oceanogr 28(8):1570–1577

Vennell R (1998b) Oscillating barotropic currents along short channels. J Phys Oceanogr 28(8):1561–1569

Vennell R, Collins N (1991) Acoustic Doppler current profiler measurements of tides in Cook Strait, New Zealand. In: Bell RG, Hume TM, Healy TR (eds) Coastal engineering—climate for change. Proceedings of 10th Australasian conference on coastal and ocean engineering, Auckland. DSIR Water Quality Centre Publication no. 21, pp 529–534

Walters RA, Plew DR (2008) Numerical modelling of environmental flows using DAM: some preliminary results. Acta Geophys 56(3):918–934

Walters RA, Werner FE (1991) Nonlinear generation of overtides, compound tides, and residuals. In: Parker BB (ed) Tidal hydrodynamics. Wiley, New York, pp 297–320

Walters RA, Goring DG, Bell RG (2001) Ocean tides around New Zealand. N Z J Mar Freshw Res 35:567–579

Walters RA, Lane EM, Henry RF (2007) Semi-Lagrangian methods for a finite element coastal ocean model. Ocean Model 19:112–124

Walters RA, Hanert E, Pietrzak JD, Le Roux DY (2009a) Comparison of unstructured, staggered grid methods for the shallow water equations. Ocean Model 28:106–117

Walters RA, Lane EM, Hanert E (2009b) Useful time-stepping methods for the Coriolis term in a shallow water model. Ocean Model 28:66–74

Wu J (1982) Wind-stress coefficients over sea surface from breeze to hurricane. J Geophys Res 87:9704–9706

Ocean Dynamics (2010) 60:1581–1596
DOI 10.1007/s10236-010-0346-7

Seasonal circulation and influence factors of the Bohai Sea: a numerical study based on Lagrangian particle tracking method

Jinhua Wang · Yongming Shen · Yakun Guo

Received: 19 December 2009 / Accepted: 27 September 2010 / Published online: 9 October 2010

Abstract Seasonal circulation of the Bohai Sea (BS) in 1992 was investigated using Lagrangian particle tracking method. The hydrography of the BS was simulated based on an unstructured grid, finite-volume, three-dimensional primitive equation ocean model. With the use of the unstructured triangular grid, the model can easily fit the irregular coastal boundary of the BS. The simulated tides, tidal current, and thermohaline field agreed well with the observations. The transport of particles has three-dimensional structure in the BS. Compared with central Bohai and Bohai Strait, the differences of particles' transportation between surface and bottom layer in three bays are small. The circulation in the summer is stronger than that in the winter, with the average residual velocity in the surface layer being about 3.7 cm/s during the summer while only 1.8 cm/s during the winter. Using the same model, several well-designed numerical experiments were performed to investigate the effect of oceanic tide, river discharge, wind stress, and thermal stratification on the circulation. It is shown that winds play an important role in the circulation of the BS during both the winter and the summer. Density circulation is important during the summer; however, it is negligible during the winter. River runoff only affects the area around the river mouth. Compared with wind and thermohaline effect, the contribution of tides is small during the summer, and the circulation under only M_2 tidal constituent could not reflect the actual circulation of the BS.

Keywords Bohai Sea · Lagrangian · Unstructured grid · Particle tracking method · Three-dimensional hydrodynamic model

Responsible Editor: Vincent Legat

J. Wang · Y. Shen (✉)
State Key Laboratory of Coastal and Offshore Engineering,
Dalian University of Technology,
Dalian 116023, People's Republic of China
e-mail: ymshen@dlut.edu.cn

Y. Guo
School of Engineering, University of Aberdeen,
Aberdeen AB24 3UE, UK

1 Introduction

The Bohai Sea (BS) is a shallow and nearly closed sea with the average depth being only 18 m. Over the past several decades, our understanding and knowledge of the structure of the BS has been continually improved. Based on the observation, some researches about the circulation of the BS have been conducted. Kuang et al. (1991) analyzed the observed data and found that the seasonal variation of the circulation of the BS is not strong. Based on the analysis of the measured datasets of current, sea temperature, and salinity, Guan (1962) suggested that the mean circulation in the BS is that the current flows in from the north and flows out from the south in the Bohai Strait. The Yellow Sea Warm Current extension enters the BS through the deep trench of the Bohai Strait, when arriving the west bank of the BS it divides into two branches with one branch moving towards the north along the west bank of Liaodong Bay and moving out along the east bank forming a clockwise gyre; while another branch moving into the Bohai Bay and passing through the old Yellow River mouth into the Laizhou Bay, and flowing out from the south of the Bohai Strait.

Although physical circulation processes on the shelf are well described in this area (see, for example, Fang et al. 2000; Hainbucher et al. 2004; Huang et al. 1998, 1999; Li et al. 2005b; Xu et al. 2006), few researches have been reported on the transport of particulate and dissolved substances. Therefore, it is essential to develop some numerical models

in order to fill this research gap. One possible model is the Lagrangian particle tracking approach. The use of Lagrangian method has a long history, and it has primarily been used by oceanographers for estimating the mean flows in the oceans and marginal seas (Malhadas et al. 2010; Nicolle et al. 2009). Understanding the transport pathways of materials can provide considerable information about the upper-ocean circulation over the large area of the BS, with potential implications for the physical environment and for the marine life that depend on it. These results could refine important aspects of the regional circulation that would be difficult to verify from observed and Eulerian model data alone.

Meanwhile, the dynamics of the shelf circulation are governed by the combined interactive forces of many factors such as stratification, winds, river discharge, and oceanic tide. As such, the simulation results of the BS's circulation and the explanation of driving mechanism vary among researchers. It is not clear which factor is the control mechanism driving the circulation of the BS. To elucidate this, extensive investigation is required.

The present study is an attempt to fill the particular gaps addressed above. An unstructured grid, finite-volume, three-dimensional, primitive equations ocean model is used in order to present the seasonal circulation in the region. In particular, the dynamical features obtained in this realistic simulation are compared with the available field observations. The model is used to hypothesize what the circulation may look like by means of a Lagrangian numerical integration method. It is expected that the passive transport problem may also be useful for interpreting the transport of quasi-conserve tracers (e.g. nutrients, primary biological productivity) in the oceans. This paper fully considers the effects of winds, heat flux, oceanic tide, and river discharge on the behavior of the BS's circulation during both the winter and the summer.

The arrangement of the paper is as following. Section 2 describes the configurations of the model. The simulated results including tides, tidal currents, temperature, and salinity filed, as well as the surface circulation of the BS are presented in Section 3, with comparisons to available field observations. Section 4 presents the results of a series of numerical experiments designed to examine the influences of tidal currents, the river discharge, wind stress, and thermal stratification on the behavior of the circulation of the BS, following a summary in Section 5.

2 Numerical model and setup

2.1 Model description

Three-dimensional ocean models have been widely used in coastal seas and estuaries and can be generally divided into three types: (1) finite-difference model, e.g., POM (Blumberg and Mellor, 1987), ROMS (Shchepetkin and McWilliams, 2005); (2) finite-element model, e.g., QUODDY (Lynch et al. 1996), ADCIRC (Luettich et al. 1992); (3) finite-volume model, e.g., ELCIRC (Zhang et al. 2004), FVCOM (Chen et al. 2003). The numerical experiments made in this study were conducted using FVCOM: a prognostic, unstructured grid, finite-volume, free-surface, three-dimensional primitive equation coastal ocean, and estuarine model developed by Chen et al. (2003). The model has been used to simulate the typhoon-induced storm surge in the Hangzhou Bay (Guo et al. 2009). The FVCOM utilizes a modified Mellor and Yamada level 2.5 (Galperin et al. 1988) and Smagorinsky (1963) turbulent closure schemes for the default setup of vertical and horizontal mixing, respectively. Unlike other coastal finite-difference and finite-element models, FVCOM solves the hydrostatic primitive equations by calculating fluxes resulted from a discretization of the integral form of these equations on an unstructured triangular mesh. This approach not only takes advantages of finite-element methods for grid flexibility and finite-difference methods for numerical efficiency but also provides a good numerical representation of momentum, mass, salt, and heat conservation.

Fluid particles were tracked by solving the x, y, and z velocity equations

$$\frac{dx}{dt} = u, \frac{dy}{dt} = v, \frac{d\sigma}{dt} = \frac{\omega}{H+\zeta} \tag{1}$$

where u, v, and ω are the x, y, and σ velocity components; H, ζ are the bottom depth (relative to z=0) and the height of the free surface (relative to z=0). The relation between ω and w is defined as

$$\omega = w - (2+\sigma)\frac{d\zeta}{dt} - \sigma\frac{dH}{dt} \tag{2}$$

where w is the vertical velocity in the z coordinate direction. Equation (1) was integrated by means of the fourth-order Runge–Kutta scheme. Particle velocities used in this calculation were obtained using a bilinear interpolation from eight nearest grid points.

2.2 Model configuration

The simulated area covers 37°07′ N–41° N and 117°35′ E–122.5° E. In order to better fit the irregular coastline, the horizontal resolution is about 3 km around the coast and about 8–9.5 km in the interior and near the open boundary. The computation has 3,200 nodes and 5,887 triangular elements as shown in Fig. 1c. In the vertical, it comprises 10 uniformly distributed σ layers, which result in a vertical resolution of about 0.1–1 m in the coastal region which is shallower than 10 m, and about 6 at 60 m isobaths. The bathymetry used in this model is provided by the field observation carried out by the Navigation Guarantee

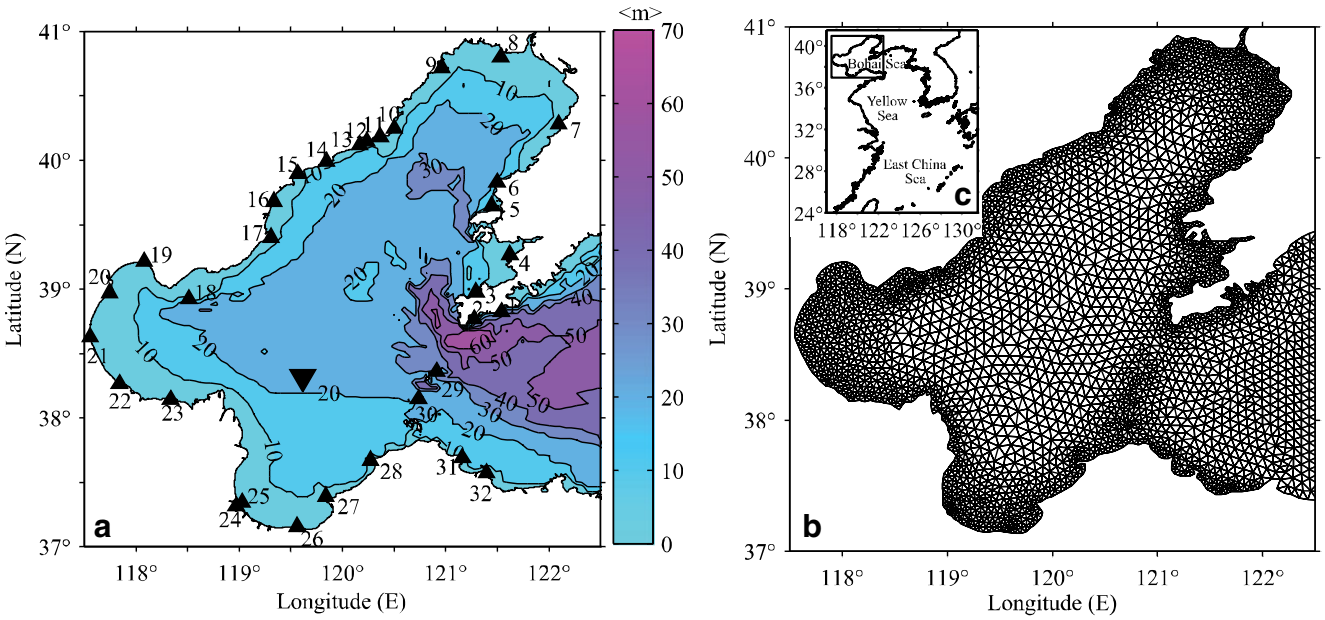

Fig. 1 **a** Model bathymetry of the BS. *Numbers* next to *filled triangles* denote the tidal stations where tidal harmonic constants are compared between simulations and observations (see Fig. 4). *Filled inverted triangle* denotes the tidal current observation station; the comparison between simulations and observations is shown in Table 1. **b** The numerical computation unstructured grid of the BS area and **c** model domain

Department of the Chinese Navy Headquarters as shown in Fig. 1a and interpolated to the mesh grid by distance weighted interpolation method. Based on the CFL condition, the external time step is 24 s and the internal mode is 15 times of the external mode. The simulation started on January 1, 1991, and ended on December 31, 1992 and the results of year 1992 are analyzed and presented in this paper.

The harmonic constants of M_2, N_2, S_2, K_2, K_1, O_1, and P_1 obtained from the coastal gauges at the northern and southern coasts were interpolated onto the open boundary. Along the open boundary, monthly temperature and salinity derived from Levitus dataset (Levitus and Boyer 1994; Levitus et al. 1994) were linearly interpolated in time and the vertical layers. On the sidewalls and bottom, the normal gradients of temperature and salinity were set to zero. The model area includes seven rivers whose positions of mouths are shown in Fig. 2. The discharge rates of these rivers were derived according to the method of Zhang et al. (2007). The bottom roughness z_0 was chosen equal to 0.001 m. The meteorological parameters (wind components at 10 m above the sea level, air temperature, pressure, relative humidity, cloudiness, and precipitation rate) were obtained from the analysis of the National Center for Environmental Prediction (NCEP), with a bilinear interpolation in space and linear interpolation in time. Using these parameters, the heat forcing at the air–sea interface can be calculated according to the formulas presented by Shen et al. (2004). The initial temperature and salinity distribution were extracted from monthly averaged Levitus dataset (Levitus and Boyer 1994; Levitus et al. 1994), and the initial sea level and current were set to zero.

3 Results

3.1 Tides and tidal currents

The model-predicted time series of surface elevation and currents at each grid point was fitted by a least squares

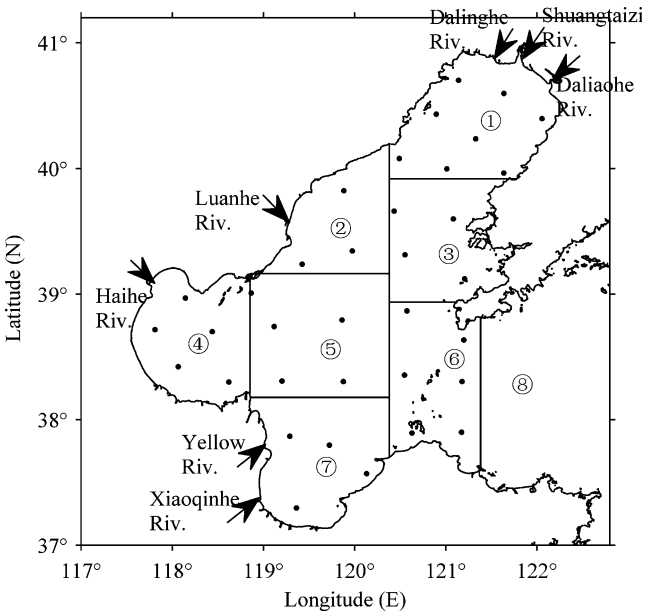

Fig. 2 Locations of rivers, initial particles, and partitions of the BS. ①–⑧ represents the Liaodong Bay, Qinhuangdao area, Changxingdao area, Bohai Bay, Bohai central, Bohai Strait inside, Laizhou Bay, Bohai Strait outside

harmonic analysis method. Since the major diurnal and semi-diurnal tidal constituent are M_2 and K_1, only these two tides are analyzed. The resulting co-amplitude, co-phase of each tidal constituent is shown in Fig. 3. The model predicted two amphidromic points of M_2, one is located offshore of Qinhuangdao on the northwest coast and another is near the old Yellow River mouth. In contrast, only one amphidromic point of K_1 appears at the southern part of the Bohai Strait. There are two areas where M_2 current is strong. One is the Laotieshan Passage, another one is close to Changxingdao. The amplitudes of M_2 resultant current in these areas are about 0.8 m/s. Because the Laotieshan Passage is also an area having the strongest diurnal current about 0.3 m/s, the total tidal current thus is very strong there.

A comparison of tidal harmonic constants at 32 tidal stations between model results and observations is shown in Fig. 4; moreover, the simulated current amplitudes and phases at a current observation station as shown in Fig. 1a is compared with the observation (Qiao et al. 2008). Both the comparisons (see Fig. 4 and Table 1) indicate a good reproduction of the four major tidal constituents in the BS; furthermore, the simulated distributions have been found to be in overall qualitative agreement with the observed results (Editorial Board for Marine Atlas 1992).

3.2 Temperature and salinity field

Comparison between the observed and the simulated results of temperature and salinity at four oceanographic stations are shown in Fig. 5. The simulated SST agrees well with the observed data, with the maximum errors being less than 4°C for the whole simulation period. Due to the insufficient precision of the precipitation data, the model cannot catch the sudden change of the salinity field. However, the model, in general, accurately reproduces both the temperature and salinity variability observed at these sites.

Temperature and salinity distributions of the surface layer in both the winter and summer are shown in Fig. 6. During summer, temperature in three Bays is higher due to the small heat content of the shallow depth. There exists a high temperature area adjacent to the Qinhuangdao area because of weak current (also shown in Fig. 3). This is resulted from particularly weak mixing in the surface layer

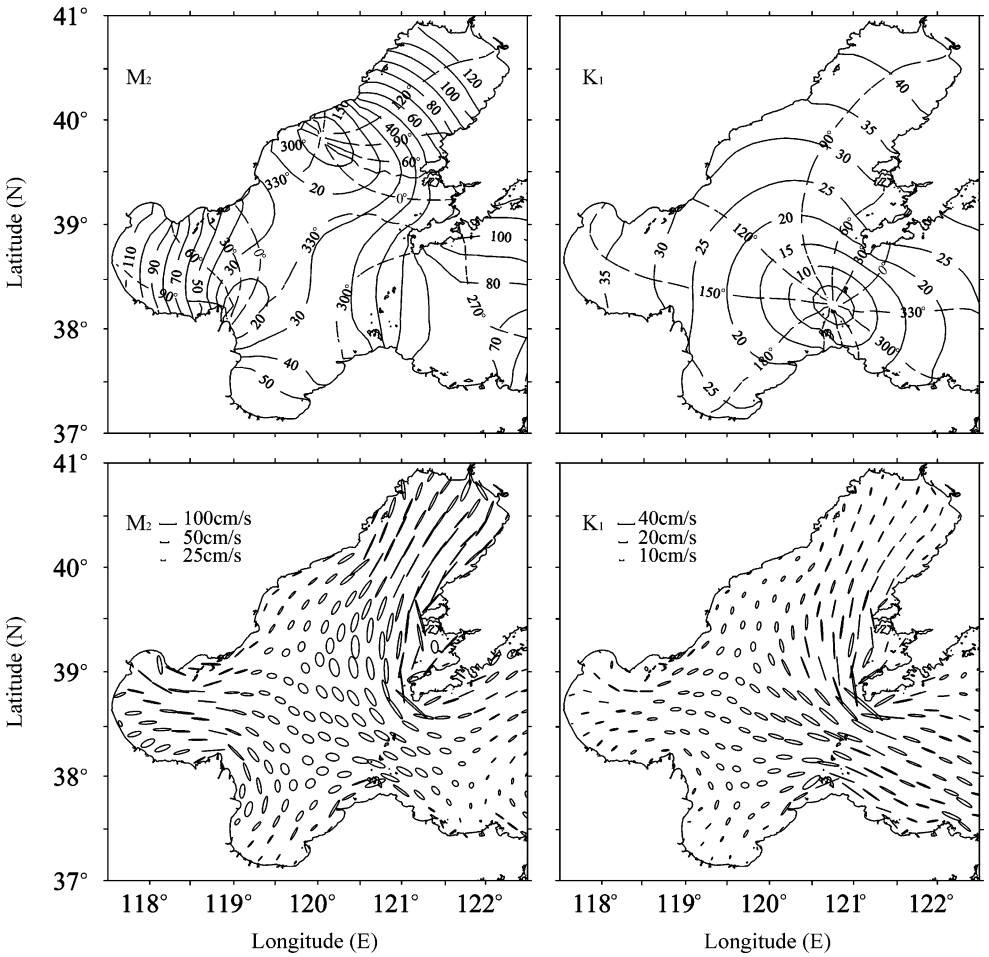

Fig. 3 Co-amplitude *line* (*solid*, in cm) and co-phase line (*dashed*, in degree) of M_2 and K_1 (*upper panels*) and their tidal current ellipses (*lower panels*)

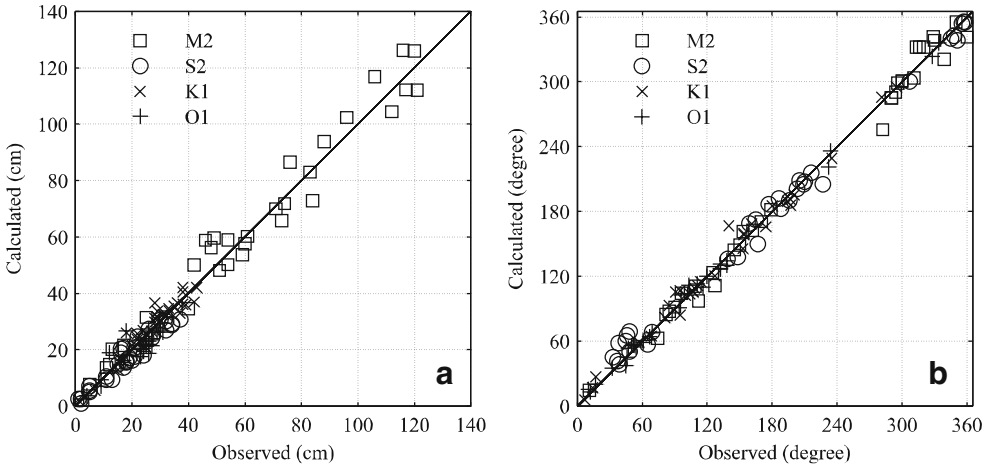

Fig. 4 Comparison of harmonic constants derived from simulations (ordinate) and observations (abscissa) at 32 tidal stations as shown in Fig. 1a for **a** amplitude and **b** phase. *Different symbols* denote different tidal constituents

since winds and tidal currents are very weak and there is not enough energy to mix the water column against the surface heat-induced stratification. In the winter, the temperature in three Bays is low, even below zero at the headland of Liaodong Bay. The salinity is high in the central of the BS, and low around the Liaodong Bay and the Yellow River mouth. Both the Haihe River and Luanhe River have little effect on the salinity distribution for lower river discharge in year 1992. From the salinity isobars of 30 psu, the low salinity front has the trend to move to the south of the Bohai Bay, and extends to the Bohai Strait as a result of large river discharge in the summer.

3.3 Surface circulation of the BS during both the winter and the summer

Lagrangian track of water parcels generally cannot appropriately sample all relevant space and time-scales. Additionally, sensitivity to initial location and time of release is known to affect the particle trajectories. While we acknowledge that releasing more particles would improve our simulations, we implicitly assume that the number of particles available to us in this study adequately (and at least qualitatively) capture the main features of the transport on the shelf during the various releases. In this study, the circulation is estimated from the Lagrangian trajectories of the released particles. Particles were initialized at each cell center and were tracked from February 1 to April 1 and July 1 to September 1 1992, representing both winter and summer, respectively.

The surface circulation on the inner- and mid-shelf varies seasonally with the winds. Being controlled by the Asian Monsoon, the winds over the BS show an apparent seasonal variation: the northerly winds prevail in the winter and the southeasterly winds in the summer (Fig. 7). During the winter months (February–March), relatively strong (0.4 dyn/cm^2) northerly and northwesterly winds (2-months averaged NCEP wind stress) prevail over the BS. The winds stress curl in the northeastern areas is about 3×10^{-6} and about -7×10^{-6} N/m^3 in the southwestern areas. During the

Table 1 Comparison between model-calculated and observed tidal current harmonic constants at the current observation station (38°19′39.426″ N, 119°37′1.546″ E)

		H_U (cm/s)		G_U (°)		H_V (cm/s)		G_V (°)	
		Observed	Calculated	Observed	Calculated	Observed	Calculated	Observed	Calculated
Surface	M_2	20.11	22.18	45.51	43.80	31.13	30.88	176.50	169.98
	S_2	4.79	4.94	122.34	131.98	7.38	7.39	236.41	250.65
	K_1	7.03	4.57	49.87	38.60	14.22	12.50	224.70	236.07
	O_1	6.32	4.15	13.94	358.41	8.73	7.48	189.43	171.97
Bottom	M_2	16.85	15.06	15.46	35.72	25.88	22.78	156.52	165.66
	S_2	4.29	3.14	89.10	115.90	7.23	5.32	214.42	237.30
	K_1	3.22	3.39	52.74	66.15	16.52	10.24	238.84	231.27
	O_1	4.16	2.74	53.37	78.53	13.37	5.62	204.91	169.47

H_U, G_U are amplitude and phase lag of northern current component; H_V, G_V are amplitude and phase lag of eastern current component

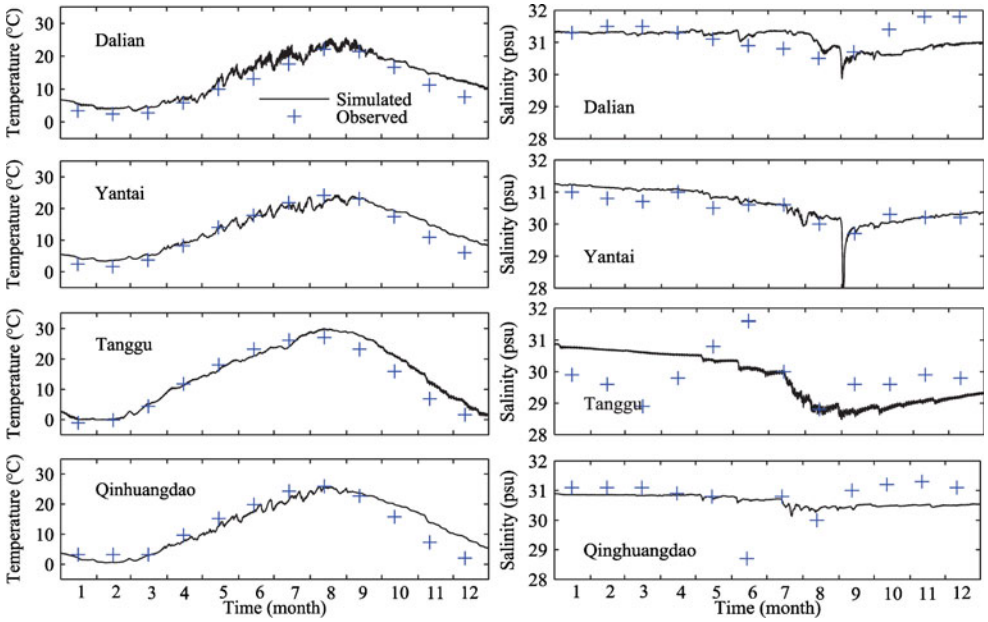

Fig. 5 Comparison of surface temperature and salinity between simulated and observed at four oceanographic stations

summer months (June–July), relatively weak (0.2 dyn/cm^2) southeasterly winds prevail over the BS. As such, the dynamics in the BS is strongly influenced by wind conditions (Fang et al. 2000; Hainbucher et al. 2004; Li et al. 2005a, b). Therefore, a strong dependence of surface circulation on seasonal wind is expected.

The main characteristics of the large-scale surface circulation in the BS can be identified by the displacement

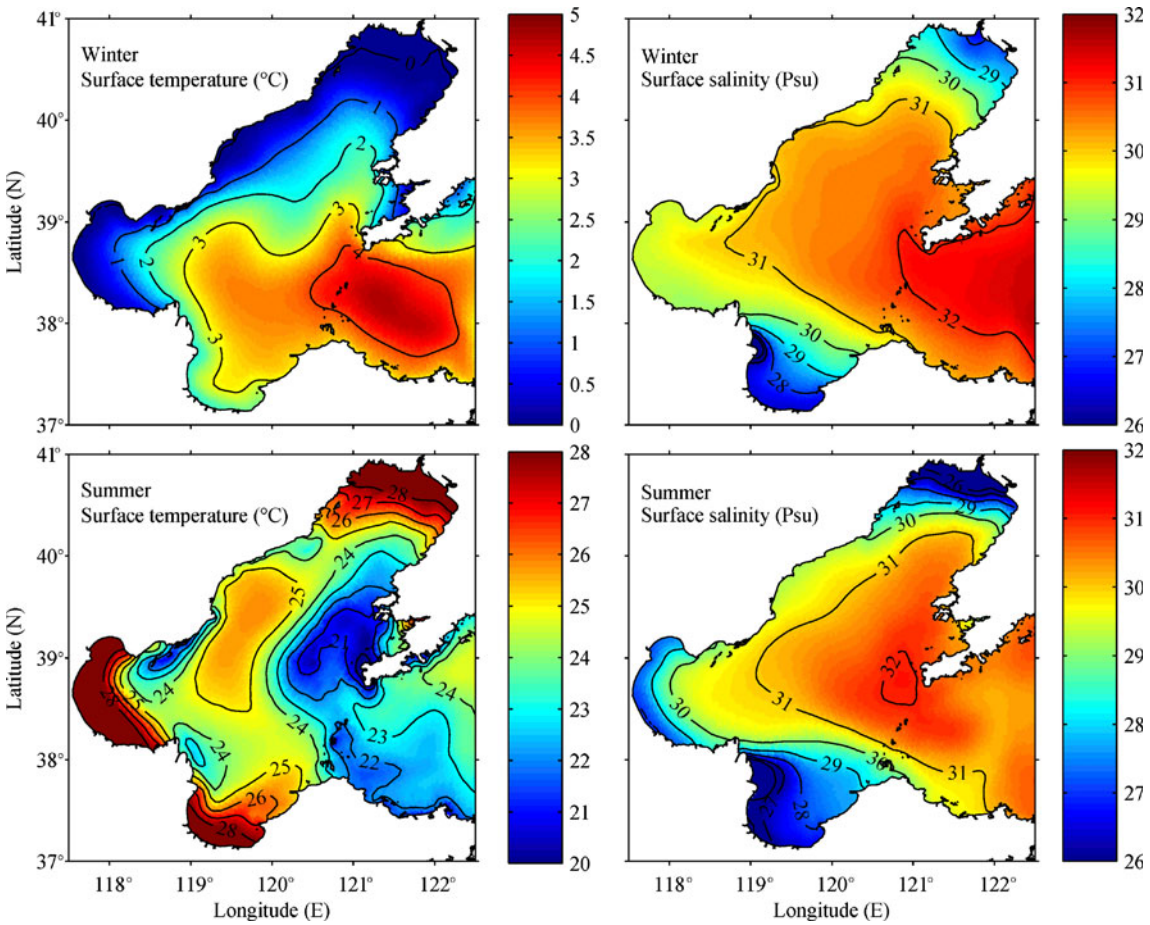

Fig. 6 Distribution of temperature and salinity field of the surface layer in the mid of February and August

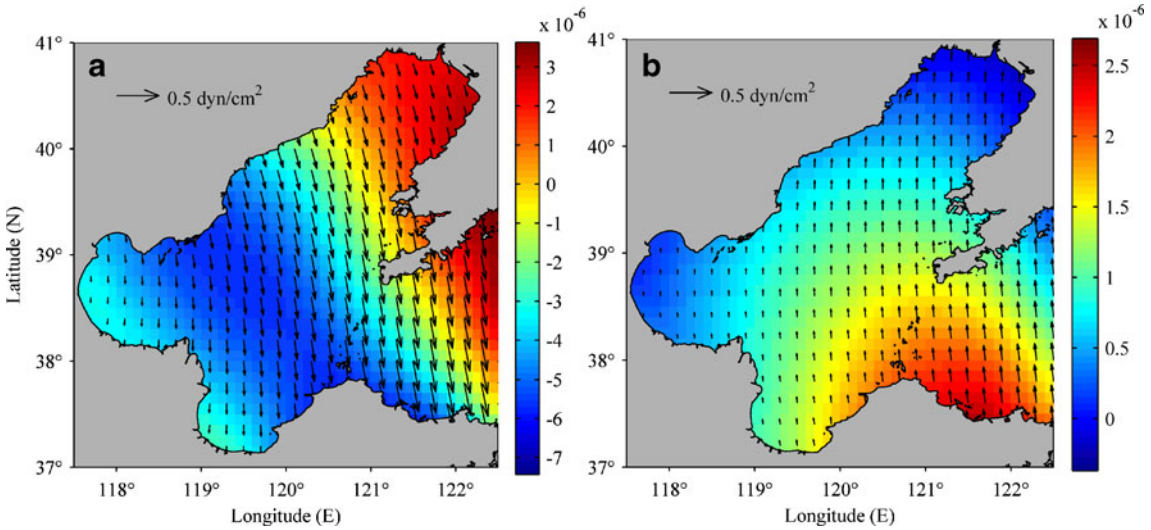

Fig. 7 NCEP wind stress and wind stress curl (N/m^3) over the BS. This is an average for the periods **a** February to March 1992; **b** July to August 1992

vectors shown in Fig. 8. During the winter months (Fig. 8a), the surface particles move out from the Bohai Strait, and a clockwise gyre is found in Liaodong Bay, which is consistent with the observation of surface circulation by Guan (1962). The particles located outside of Qinhuangdao move to the southeast, converge with the clockwise gyre at the northern mouth of the Bohai Bay, pointing to the old Yellow river mouth. An anti-clockwise gyre is found in the south Bohai Bay, flowing along the coastline of Laizhou Bay, and finally flowing out from the south Bohai Strait. In Laizhou Bay, the particles' transport distances near the Yellow River mouth and central of Laizhou Bay are slightly longer. The particles released at the Yellow River mouth move south along the west bank of Laizhou Bay, which can also be seen from the low salinity front shown in Fig. 6. The low salinity plume moves to the headland of Laizhou Bay and has the tendency to turn to the southeast as a result of the influence of the Coriolis force (Li et al. 2005a). Particles located at the central of the BS move south, then turn to the southeast and move out from the south of Bohai Strait finally. Particle's moving distances are long in the Qinhuangdao area and within the Bohai Strait.

During the summer months (Fig. 8b), surface particles in the Bohai Bay move east and flow out from the Bohai Strait. Particles in the Bohai central divide into two branch:

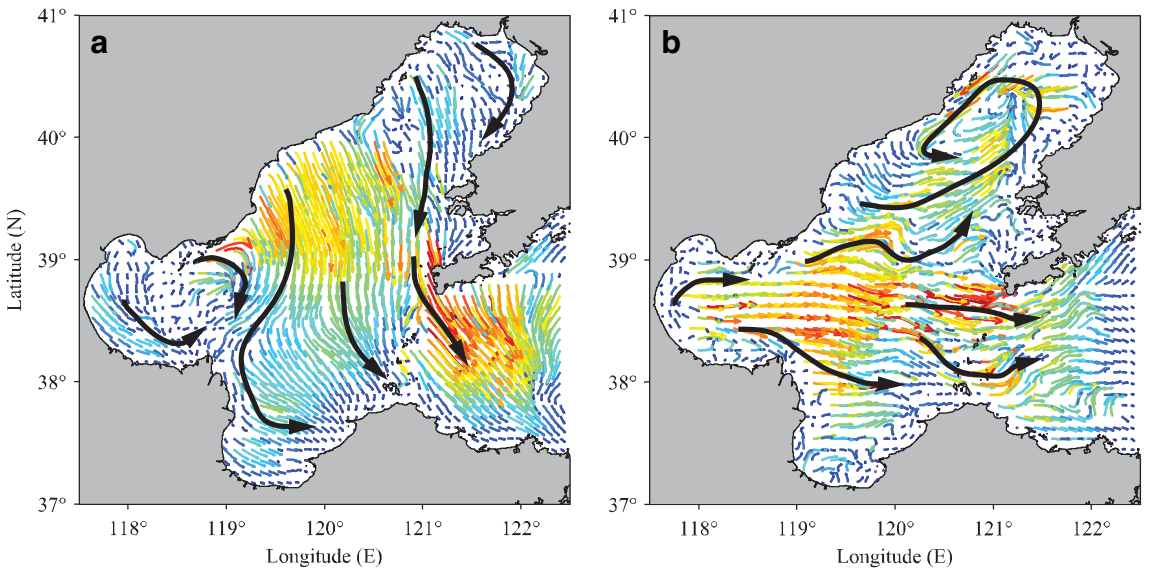

Fig. 8 Displacement vectors of all particles from the release locations where *red color* indicates longer displacement **a** February to March 1992; **b** July to August 1992. *Thick arrows* show the schematic diagrams of the upper circulation of the BS deduced from the displacement vectors; these results should be viewed as suggestive rather than quantitative

one converge with the particles in the Bohai Bay moving out form the Bohai Strait; another branch moves to northeast together with the particles in the Liaodong Bay form an anti-clockwise gyre. Particle moving distances are long in the Bohai Bay, Bohai central, and inside the Bohai Strait. Particles around the Yellow River mouth and at the central of the Laizhou Bay move east, which is in accordance with the low salinity front.

In general, the results obtained from surface particle trajectories not only confirmed the circulation structures discussed in the literature (Hainbucher et al. 2004) but also provided a more quantitative and higher resolution description of the BS's surface circulation features due to the high density of particles released in this area.

4 Discussion

4.1 Effects of different factors on the seasonal circulation of the BS

To investigate the effects of different factors on the seasonal circulation in the BS, particles were released near the surface layer (5 m below the sea level) and bottom layer at the beginning of February and August and were tracked for a month. Thirty-five particles were placed in the BS in order to not only reflect the Lagrangian water transport but also be convenient to describe the particle trajectories. The initial locations of particles are shown in Fig. 2.

Six numerical experiments (Table 2) were conducted using the calibrated model, to identify the prevailing forcing functions and the BS characteristics: (1) the benchmark experiment with full forcing (EXP1); (2) a baroclinic experiment initialized as in EXP1 but without winds forcing (EXP2); (3) a baroclinic experiment initialized as in EXP1 but without surface heat flux forcing (EXP3); (4) a baroclinic experiment initialized as in EXP1 but without freshwater forcing (EXP4); (5) a barotropic experiment forced with river discharge (EXP5), and (6) a barotropic experiment forced by M_2 tidal constituent (EXP6). The difference between the control run and the experiment runs EXP2, EXP3, EXP4 is used to investigate the effects of the winds, heat fluxes, river discharge on the Lagrangian transport of the BS. Comparison between EXP5 and EXP2 shows the baroclinic effect, while the comparison between EXP 5 and EXP 6 show the tides effect on the Lagrangian transport of the BS.

4.1.1 Particle trajectories in the winter

The particle trajectories in surface and bottom layer for 6 experiments in the winter are shown in Fig. 9. Figure 9 shows that for control run (EXP1), the longest net transport distance of the surface particles is about 30 km near coastline, and increases to 146 km at the central south of Bohai Strait. In the whole BS, the difference of net transport distance between bottom particles is small with the longest net transport distance being about 32 km and taking place at the Bohai Strait.

Wind forcing is likely to play an important role on the surface circulation. A comparison between EXP1 and EXP2 reveals the winds effect. In the surface layer, the particles located at three bays show similar trajectories to those in EXP1. However, there is noticeable difference at some areas. One area is at the central south of the Liaodong Bay where particle moves to the central, while it moves to the east coast of the Liaodong Bay in EXP1. Another area is outside of Qinhuangdao where particles' transport distances are short comparing to those in EXP1. The third area is at the northern Bohai Strait where particle moves in and has short transport distance. In other areas, the move directions of particles are similar to EXP1, but the net transport distances are shorter. There is no obvious difference between EXP1 and EXP2 for the bottom particle trajectories. According to the above analysis, it can be concluded that winds play a leading role in the BS's circulation during the winter.

When the surface heat forcing is eliminated (EXP3), a slight change occurs in the particle trajectories in comparison to EXP1, indicating that the density change caused by temperature has little effect on the particles' moving due to small horizontal temperature gradient in the winter.

Table 2 Experiments designed in this article

Experiment	Conditions used in the experiment
1	Tides (seven major tidal constituents are considered) + Wind + Heat flux + River discharge
2	Tides (seven major tidal constituents are considered) + Heat flux + River discharge
3	Tides (seven major tidal constituents are considered) + Wind + River discharge
4	Tides (seven major tidal constituents are considered) + Wind + Heat flux
5	Tides (seven major tidal constituents are considered) + River discharge
6	Tides (only M_2 tidal constituent is considered)

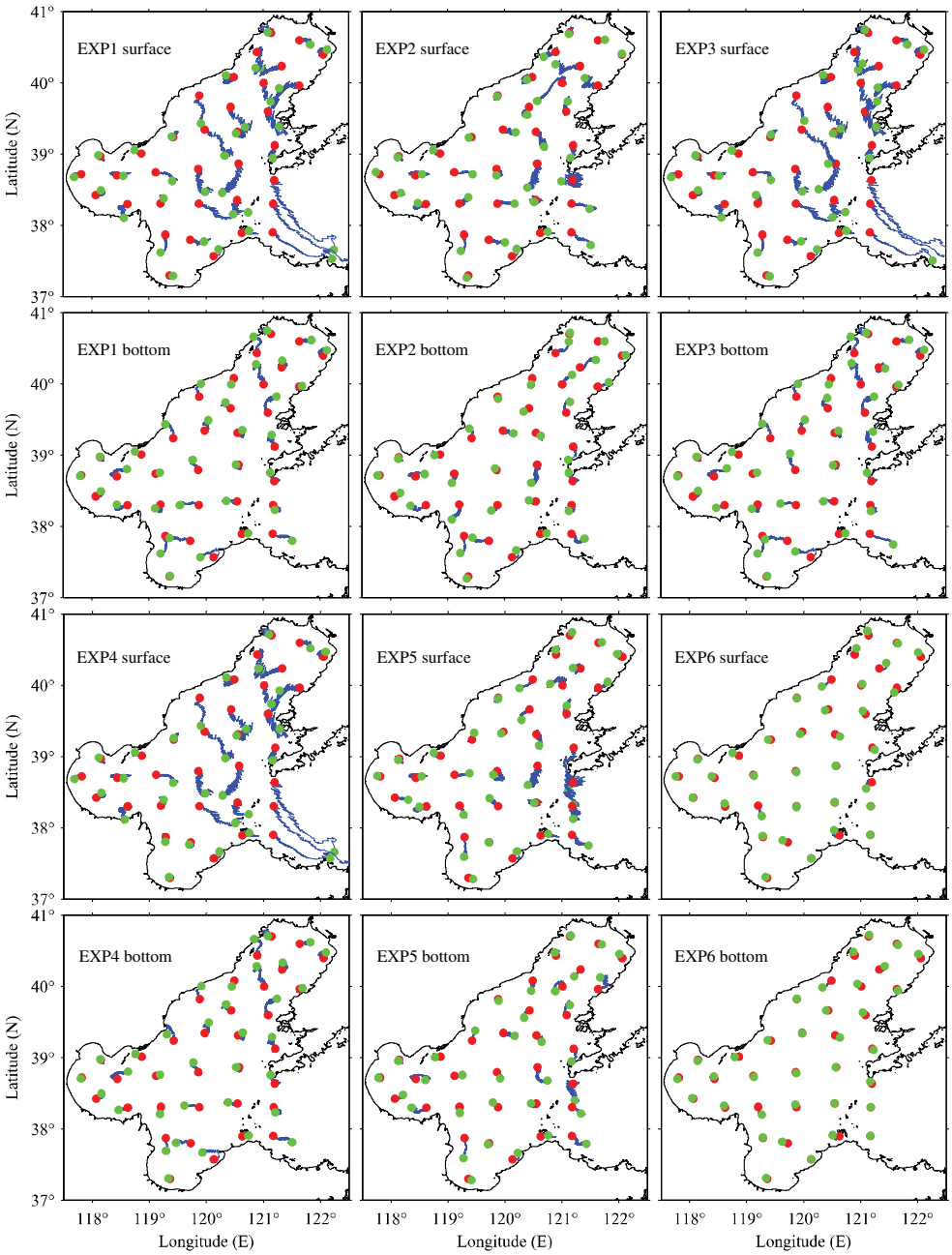

Fig. 9 Trajectories of particles in a month during winter, *red points* represent initial locations and *blue* represent final locations

Without considering the effect of river discharge (EXP4), only the particles near the Yellow River mouth are affected. The net transport distances decrease at the location of the Yellow River mouth and the central of Laizhou Bay. The other rivers' discharges are small comparing to the Yellow River in the winter; thus, they have little effect on the Lagrangian water transport.

In EXP5, the model was driven by composite tide and river discharge with temperature and salinity fixed. Except the area around the Liaodong Bay and the Yellow River mouth, the particle trajectories are similar than those in EXP2. It shows that the density current caused by the thermohaline has little contribution to the particles' movement; the Lagrangian water transport in the winter is mainly controlled by winds and tides.

When only M_2 tidal constituent was considered (EXP6), the net transport distances of most particles in a month are short. Whereas the net transport distances of particles at the location of the Yellow River mouth, the west coast of the Liaodong Bay and two sides of the Bohai Strait are slightly longer due to the nonlinear effect caused by mutation of the terrain and topography of the coastline. The moving distances are shorter than those in EXP5. The longest moving distance is 18 km which results in Lagrangian residual velocity being less than 1 cm/s. So it can be concluded

that though M_2 is the major tidal constituent it does not reflect the actual circulation while analyzing the Lagrangian water transport.

4.1.2 Particle trajectories in the summer

The particle trajectories in surface and bottom layer for 6 experiments in the summer are shown in Fig. 10. For control run (EXP1), the longest net transport distance of the surface particles is about 80 km near coastline, and increases to 175 km at the Bohai central. In the whole BS, the difference of net transport distance between bottom particles is small, and the longest net transport distance is about 52 km, occurring at the Bohai Strait.

In EXP2, the average net displacement of surface particles is of the magnitude of EXP1. The main difference is the moving direction at the north of the BS where particles move towards the south in EXP2 and the southeast in EXP1 as a result of wind-induced Ekman transport. Particle trajectories in the bottom layer are similar than those in EXP1, indicating that winds effect on the bottom water transport in the summer is weak.

Surface particles' moving distances decrease significantly in EXP3, but the moving directions are similar to that in

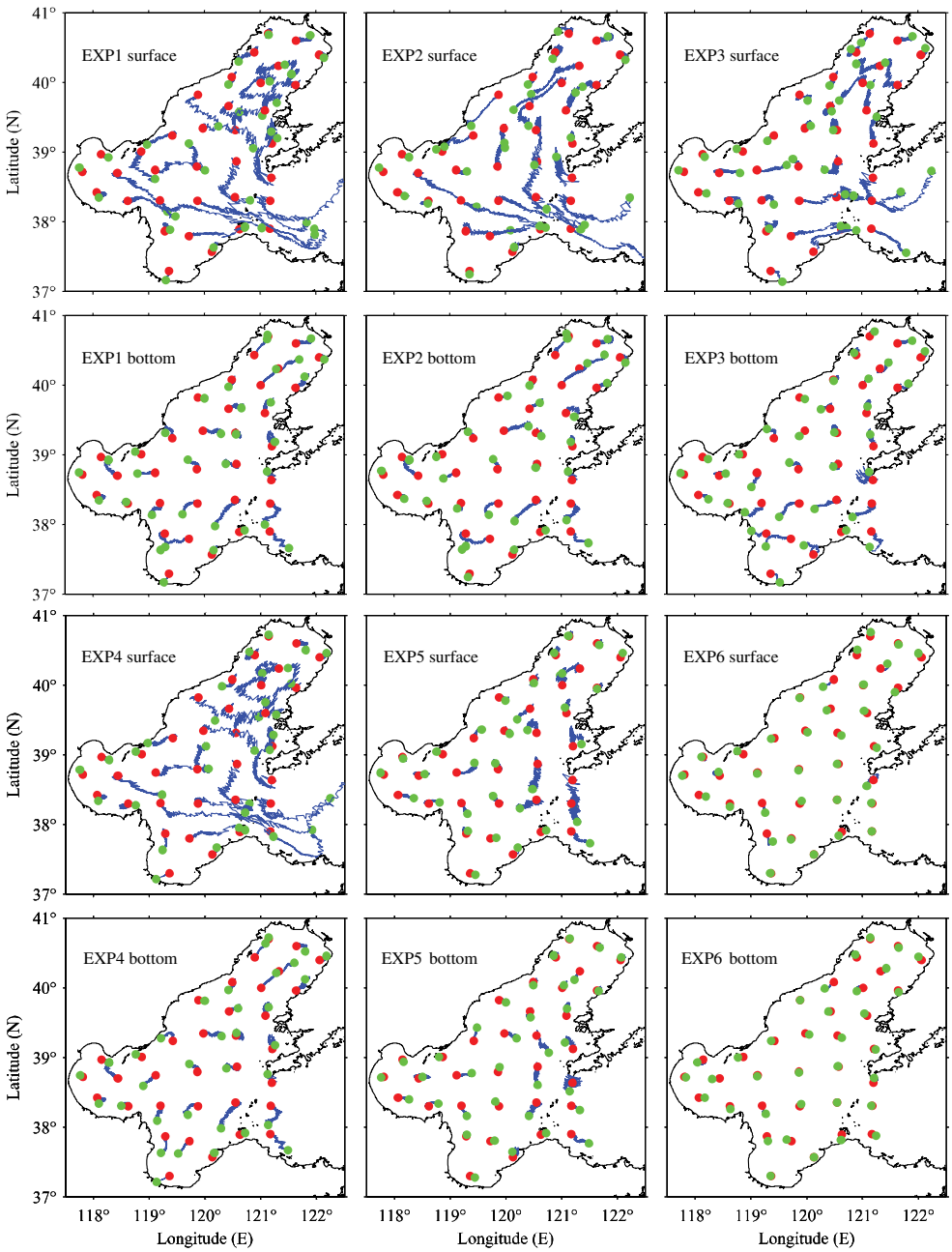

Fig. 10 Trajectories of particles in a month during the summer, *red points* represent initial locations and *blue* represent final locations

EXP1, showing that in the summer the heat flux enhances the particles' moving which differs from that in the winter. In the bottom layer, there are some changes of moving direction for particles located at the mouth of the Bohai Bay and outside of Qinhuangdao, as a result of weaken circulation in the surface layer in these areas.

Without considering the effect of river discharge (EXP4), except some changes around the Yellow River mouth and the Liaodong Bay the particle trajectories in other places are similar to those in EXP1 in the surface layer. Although rivers' discharges are larger in the summer, they only change the velocities near the river mouth and have little influence on the summer motion throughout the BS. A detailed discussion about the river run-off influence is shown in Section 4.2.

Though the particle trajectories are similar between EXP2 and EXP5 during the winter, the particle trajectories in EXP5 show great differences to those in EXP2 in the summer. The net transport distance of the particles located at the Bohai central, the Bohai Strait, three mouths of the Liaodong Bay, the Bohai Bay and Laizhou Bay is smaller, indicating that the density circulation is the dominant factor in the summer. In EXP6, the particle trajectories at the surface and bottom layer are similar as those in EXP6 of the winter since only M_2 tidal constituent was concerned.

4.1.3 Statistical comparison between the winter and the summer

The area-averaged net displacement of the surface particles after 1 month during the winter and the summer for EXPs 1–6 is shown in Fig. 11. The area-averaged net displacement of the surface particles during the summer is 96.04 km resulting in a residual velocity of 3.7 cm/s during the summer, which is about twice of that in the winter. In EXP2, the sharp of the histograms is similar to EXP1 except the displacement being about 20 km smaller. Without considering the effect of surface heat flux (EXP3), particle average displacements during the winter and the summer are almost equal. The area-averaged net transport distance during the summer in EXP5 is about 20 km and only 25% of that in EXP2, indicating that the density circulation dominates the summer circulation. The surface area-averaged net transport distance in EXP6 is about 1/3~1/2 of that in EXP5, proved that more tidal constituents are needed to describe the circulation of the BS.

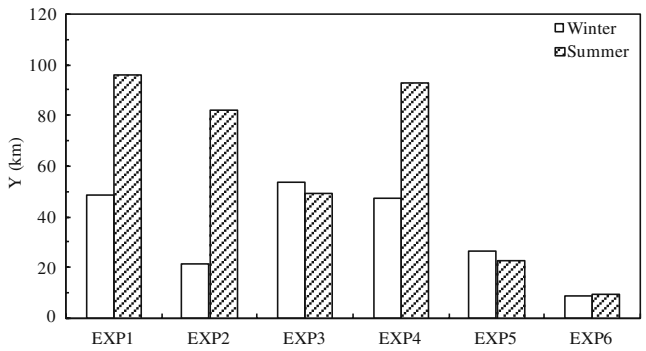

Fig. 11 Area-averaged net displacement of the surface particles as shown in Fig. 2 for EXPs 1–6 during the winter and summer

4.2 River run-off influence to the circulation of the BS

In order to describe the response of the general circulation to river runoff effect during the winter and the summer, the surface salinity and residual patterns were presented in Fig. 12. The residual values were calculated by tidally averaging the transient solution computed by the model for 60 days of integration during the winter and the summer.

Results reveal noticeable residual currents in the Liaodong Bay and the Bohai Bay. In winter, the prevailing northerly winds produce three regions of surface water flows in the Liaodong Bay: one flow to south along the western coast, one flow to south along the northern and eastern coast, and in the central body water flow to southeast. As the river discharge in the summer is larger than that in winter, a clear lower salinity plume is formed during the summer in Liaodong Bay (Fig. 12b). In comparison with the residual currents in winter (Fig. 12a), a strong anticlockwise gyre is formed in the summer (Fig. 12b) with the maximum residual velocity equal to 5 cm/s. Except that the residual velocity is large around the river mouths, the residual patterns during the winter and the summer in EXP1 (Fig. 12a and b) are similar to those in EXP4 (Fig. 12e and f). In the Bohai Bay, an anticlockwise gyre develops near the head of the bay in the winter while clockwise gyre is developed and flow out in the northeast direction in the summer, clearly related to the strong seasonal wind forcing. Particularly, the southerly winds in the summer drive the low-salinity water into the central BS; combined with the large river discharge and summertime thermal stratification the river plume shows a strong northeastward extension. The presence of the Yellow river discharge strengthens the residual currents around the river mouth by comparison between EXP1 and EPX4.

To provide additional evaluation of how well the circulation responds to the river runoff, the EXP4 without considering the run off described above was contrasted with the EXP1 focusing on the area adjacent to the river mouth. A total of 4,000 particles are initialized in two box areas locate at Laizhou bay and Liaodong bay as shown in Fig. 13, typical of the river mouth region. Particles are distributed on the middle sigma layer in vertical. The particles are released at February 1 and July 1, followed for

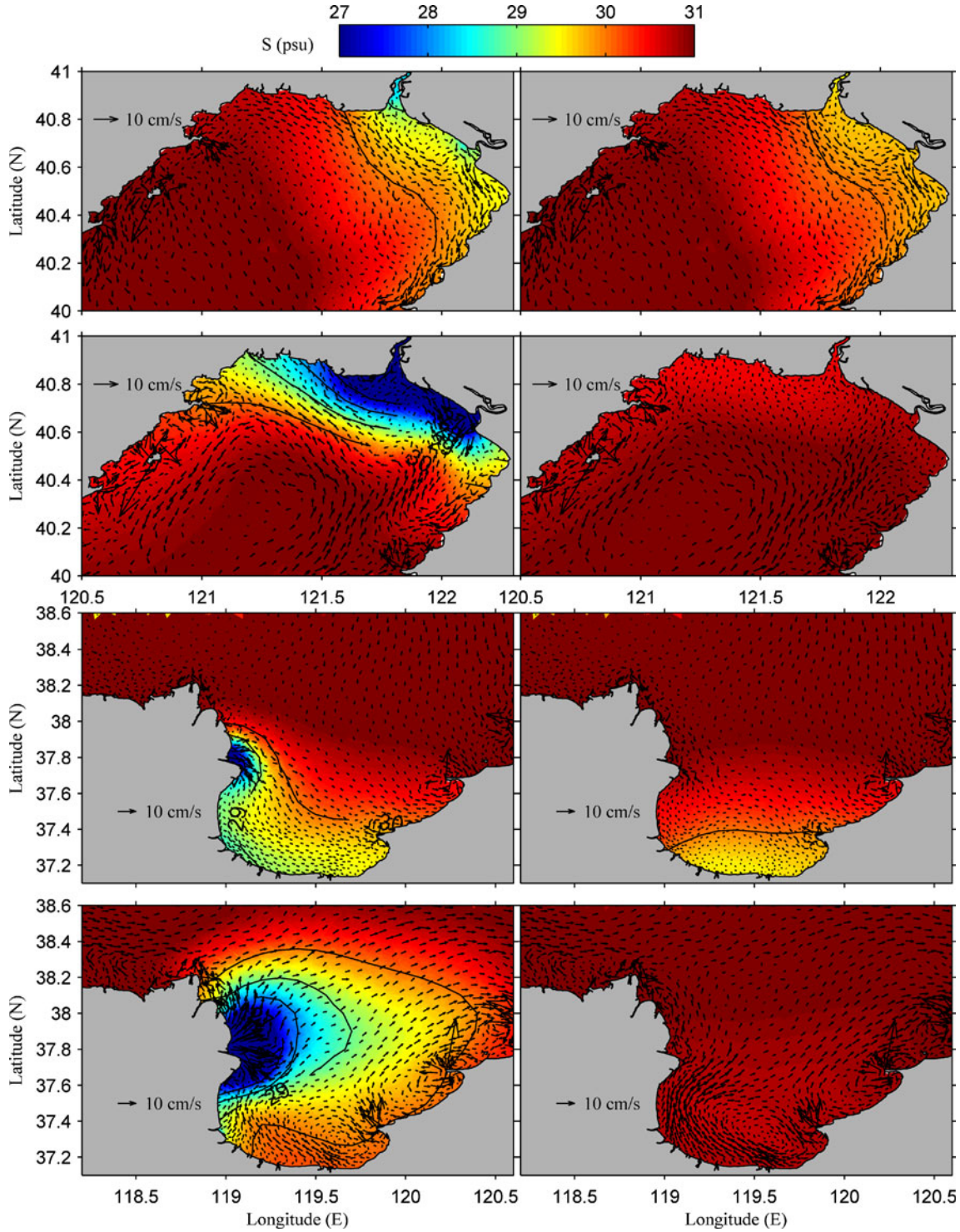

Fig. 12 Sea surface salinity and surface residual currents calculated for EXP1 (**a**–**d**) and EXP4 (**e**–**h**); (**a**, **c**, **e**, **g**) and (**b**, **d**, **f**, **h**) represent the surface salinity and subtidal currents calculated during the winter and summer, respectively

a period of 2 months to give descriptive of the water parcel motion during the winter and summer. Comparison of the snapshots of the particle locations in mid-February, early March, mid-March late March between EXP1 and EXP4 are shown in Figs. 14 and 15, color of the trajectories represents water depth.

During the winter, the particle displacements in the Liaodong Bay are small from mid-February to late

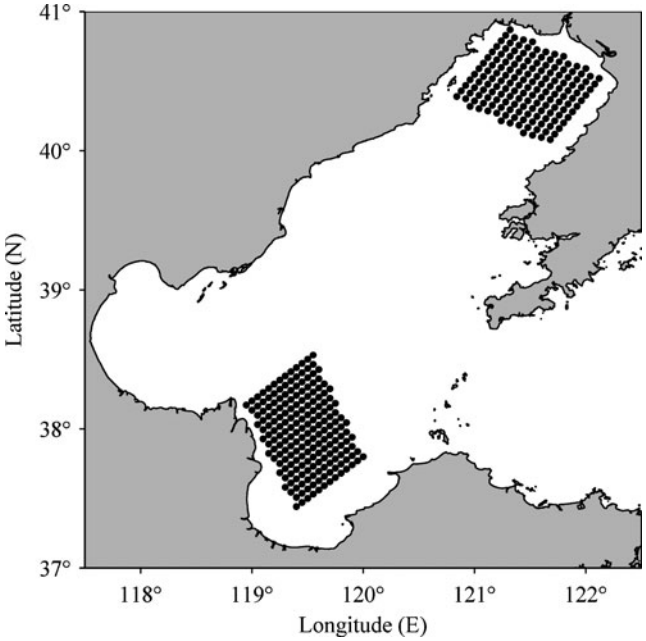

Fig. 13 Dots in the figure indicate the positions of the model particles released in the Yellow River mouth area and around the Daliaohe River and Shuangtaizi River area. In the vertical, particles are released in each middle sigma layer of the vertical grid

March, until the late March most of the particles still stay up the 40° line (Fig. 14). In the Laizhou Bay, the surface particles move to southeast, while the bottom particles firstly move north (Fig. 14b and f) then turn northeast form a clockwise movement and the head of the bottom particles reach the sea off Qinhuangdao area until the late March. The possible reason for this phenomenon is the compensation of the surface current for volume conservation. The model results thus confirm the observed distribution of the coastal deposits (Editorial Board for Marine Atlas 1992), and also some particles are found to flow into the Bohai Bay which is in confirmation of the existence of sand intrusion from the Yellow River mouth. The discrepancy between EXP1 and EXP4 is negligible as the river discharge is small during winter. Nevertheless, Fig. 14d and h shows that most of the surface particles move to east of the Laizhou bay in Exp 1 while they stay in the western part of the Laizhou Bay in EXP 4 till the late March. This can be explained by the surface subtidal currents as shown in Fig. 12c and g, considering the effect of the Yellow river runoff, the subtidal currents along the coastline of the Laizhou bay is stronger than that without.

During the summer, the particles locate in the Liaodong Bay move long distance after a half month moving, in clear contrast to the particles as shown in Fig. 14a and e. Compared with the particles in the Liaodong Bay, the particle distribution in the Laizhou Bay still shows nearly the initial box sharp. After 1 month, the particles in the Liaodong Bay continue move in anti-clockwise direction along the west coast of the bay with the front particles moving beyond the 40° N line. In the Laizhou Bay, particles spread over the entire bay, and the surface particles in EXP1 move long distance in comparison with that in EXP4 as shown in Fig. 15b and f. The particles remain in the motion tendency, the front of particles released in the Liaodong Bay move around the 39° N line while the particles released around the Yellow River mouth still locate in the Laizhou Bay after 2 months. The vaguely similar distribution of the particles is found between Fig. 15d and h; however, a few differences are found around the river mouths, where the most surface particles in the Laizhou Bay move northeast in EXP1 while around the mouth in EXP4, meanwhile the particles in the Liaodong Bay move to west coast in EXP1 while spread the whole bay in EXP4. Probably, this distribution is related to the strong density current caused by the large river discharge in the summer.

5 Conclusions remarks

A high resolution model of the BS was established based on an unstructured grid, finite-volume, three-dimensional primitive equation ocean model. The model is easy to fit the irregular coastal boundary of the area of the BS using the unstructured triangular grid. The model considers the effects of tides, winds, heat flux, precipitation, evaporation and river runoff. The simulated tides, tidal currents, temperature, and salinity were well verified, successfully reproduced the hydrodynamic characteristics and variation of temperature and salinity during a year.

Seasonal circulations of the BS in 1992 were studied by means of Lagrangian trajectory. The results indicate that circulation in the summer is about twice stronger than that in winter, with the average residual velocity in the surface layer is near 3.7 cm/s during the summer while this is only 1.8 cm/s during winter. There exists a clockwise gyre during the winter; on the contrary, a counterclockwise gyre takes place during the summer at Liaodong Bay. A double gyre in the Bohai Bay is found during the winter, with the counterclockwise one in the south and a clockwise one in the north. Seawater flows out along the central of the Bohai Bay. Compared with central BS and the Bohai Strait, where a strong three dimensional structure takes place, the transportation in three Bays and the differences between surface and bottom layer are small.

The effects of various physical factors on the circulation of the BS were analyzed via a series of numerical experiments. The model results indicate that winds play an important role in the circulation both in the winter

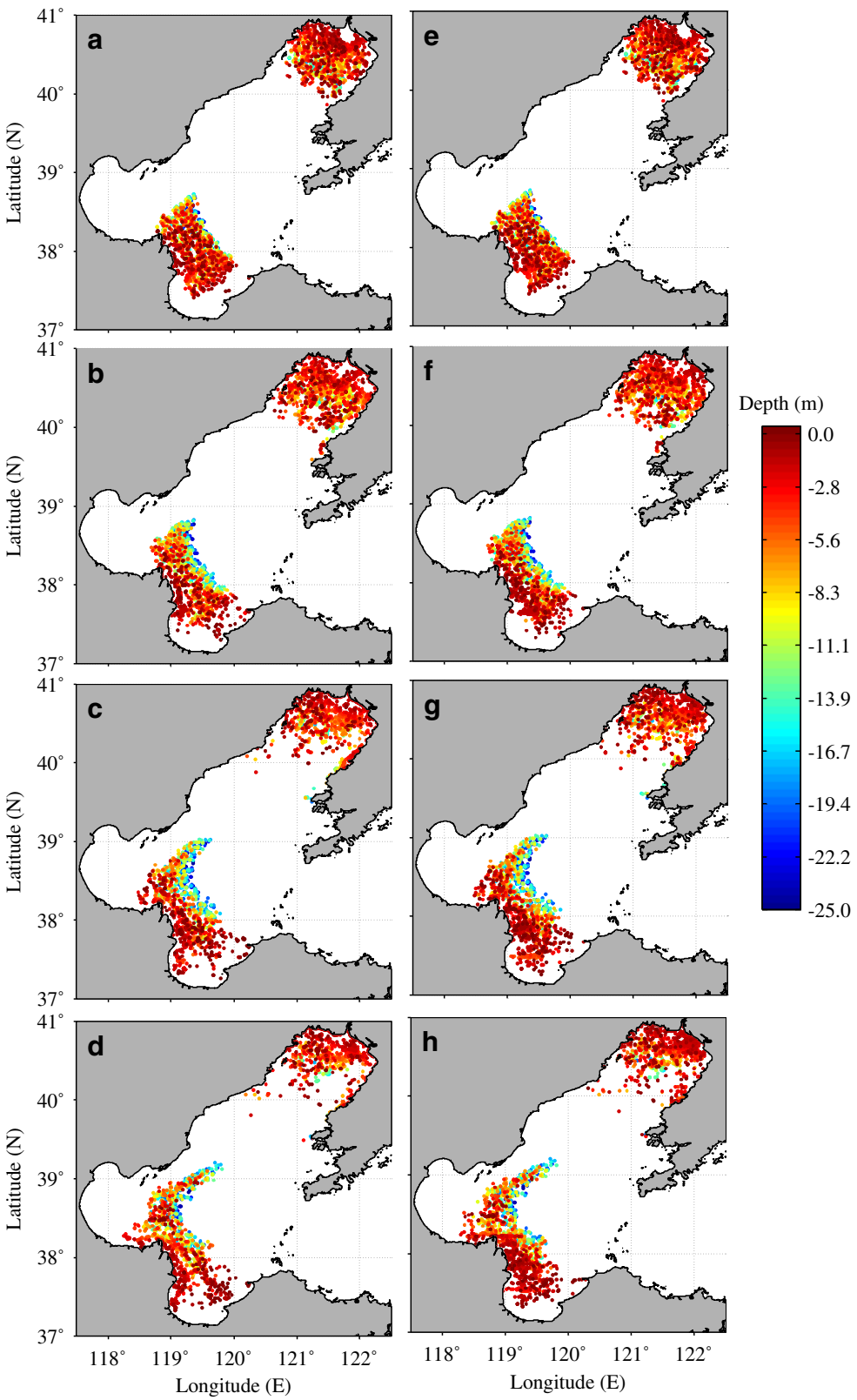

Fig. 14 Displacement of all particles from the release location during winter; color of the trajectories represents water depth. **a**–**d** are snapshots in mid-February, early March, mid-March late March in the control run (EXP 1); **e**–**h** are snapshots in mid-February, early March, mid-March late March without the river discharge (EXP4)

and summer, and density circulation is negligible during the winter while is important during the summer with the area-averaged net transport distance quadrupled without considering the density circulation. Compared with surface heat flux and wind stress, the contribution of tidal effect is small during summer, and circulation under only M_2 tidal constituent could not reflect the actual circulation of the BS. In addition, the circulation responds

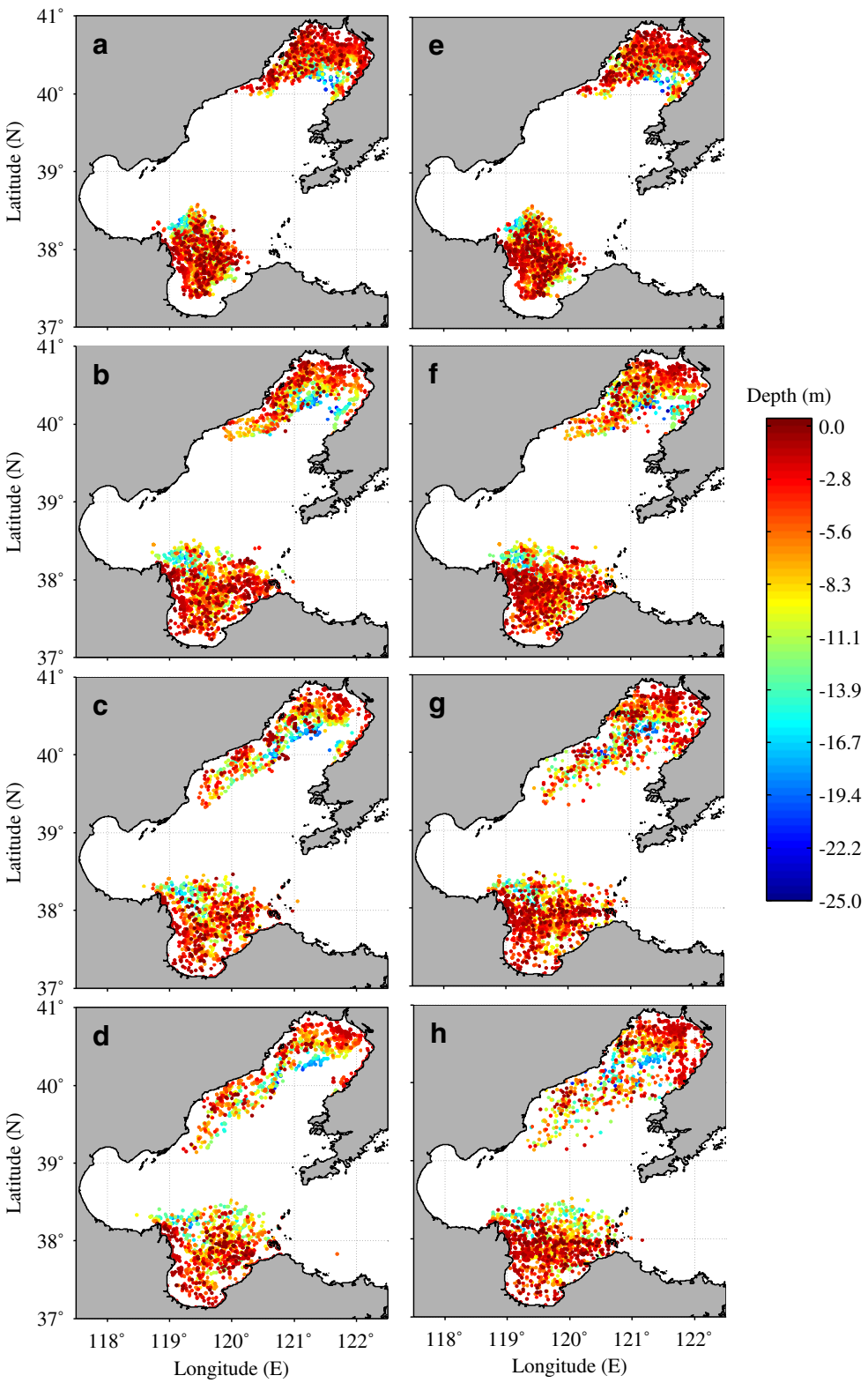

Fig. 15 Displacement of all particles from the release location during the summer; color of the trajectories represents water depth. **a**–**d** are snapshots in mid-February, early March, mid-March late March in the control run (EXP1); **e**–**h** are snapshots in mid-February, early March, mid-March late March without the river discharge (EXP4)

to the river runoff is evaluated focusing on the area adjacent to the river mouth. Remarkable similarities of the particle trajectory indicated that the general circulation is not changed with only the area around river mouth affected.

Although the model Lagrangian results are qualitative, the simulated trajectories provide a systematic overview of the general circulation and give a more visually result about the main driving forces of the seasonal circulation of the BS.

Acknowledgements This research is sponsored by the National Natural Science Foundation of China under Grant No.50839001, by the National Basic Research Program of China under Grant No. 2006CB403302 and by the Scientific Research Foundation of the Higher Education Institutions of Liaoning Province under Grant No.2006T018. The authors are also grateful to three anonymous reviewers for providing constructive comments on earlier versions of this manuscript.

References

Blumberg AF, Mellor GL (1987) A description of a three-dimensional coastal ocean circulation Model. In: Heaps NS (ed) Three-dimensional coastal ocean models. American Geophysical Union, Washington, pp 1–16

Chen CS, Liu HD, Beardsley RC (2003) An unstructured grid, finite-volume, three-dimensional, primitive equations ocean model: application to coastal ocean and estuaries. J Atmos Ocean Technol 20:159–186

Editorial Board for Marine Atlas (1992) Marine atlas of the Bohai Sea, Yellow Sea and East China Sea: Hydrology. China Ocean, Beijing

Fang Y, Fang GH, Zhang QH (2000) Numerical simulation and dynamic study of the wintertime circulation of the Bohai Sea. Chin J Oceanol Limnol 18:1–9

Galperin B, Kantha LH, Hassid S et al (1988) A quasi-equilibrium turbulent energy model for geophysical flows. J Atmos Sci 45:55–62

Guan BX (1962) Some problems concerning the study of the current structure of the near-shore area of China Sea (in Chinese). Oceanol Limnol Sin 4:121–141

Guo YK, Zhang JS, Zhang LX et al (2009) Computational investigation of typhoon-induced storm surge in the Hangzhou Bay, China. Estuar Coast Shelf S 85:530–537

Hainbucher D, Hao W, Pohlmann T et al (2004) Variability of the Bohai Sea circulation based on model calculations. J Mar Syst 44:153–174

Huang DJ, Su JL, Zhang LR (1998) Numerical study of the winter and summer circulation in the Bohai Sea (in Chinese). Acta Aerod Sin 16:115–121

Huang DJ, Su JL, Backhaus JO (1999) Modelling the seasonal thermal stratification and baroclinic circulation in the Bohai Sea. Cont Shelf Res 19:1485–1505

Kuang GR, Zhang Q, Dai YZ (1991) Observation of long-term currents and analysis of residual currents in central Bohai Sea (in Chinese). T Oceanog Limino 2:1–11

Levitus S, Boyer TP (1994) World ocean atlas 1994, volume 4: temperature. NOAA Atlas NESDIS 4, U.S. Department of Commerce, Washington

Levitus S, Burgett R, Boyer TP (1994) World Ocean Atlas 1994, Volume 3: Salinity. NOAA Atlas NESDIS 3, U.S. Department of Commerce, Washington

Li GS, Wang HL, Dong C (2005a) Numerical simulations on transportation and deposition of SPM introduced from the Yellow River to the Bohai Sea (in Chinese). Acta Geogr Sin 60:707–716

Li GS, Wang HL, Li BL (2005b) A model study on seasonal spatial-temporal variability of the Lagrangian residual circulations in the Bohai Sea. J Geogr Sci 15:273–285

Luettich JR, Westerink JJ, Scheffner NW (1992) ADCIRC: An advanced three-dimensional circulation model for shelves, coasts, and estuaries. Report 1: Theory and methodology of ADCIRC-2DDI and ADCIRC-3DL. Technical Report DRP-92-6,Department of the Army

Lynch DR, Ip JT, Naimie CE et al (1996) Comprehensive coastal circulation model with application to the Gulf of Maine. Cont Shelf Res 16:875–906

Malhadas MS, Neves RJ, Leitao PC et al (2010) Influence of tide and waves on water renewal in Óbidos Lagoon, Portugal. Ocean Dyn 60:41–55. doi:10.1007/s10236-009-0240-3

Nicolle A, Garreau P, Liorzou B (2009) Modelling for anchovy recruitment studies in the Gulf of Lions (Western Mediterranean Sea). Ocean Dyn 59:953–968. doi:10.1007/s10236-009-0221-6

Qiao LL, Bao XW, Wu DX (2008) The observed currents in summer in the Bohai Sea. Chin J Limnol Oceanogr 26:130–136

Shchepetkin AF, McWilliams JC (2005) The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. Ocean Model 9:347–404. doi:10.1016/j.ocemod.2004.08.002

Shen YM, Zheng YH, Wu XG (2004) A three-dimensional numerical simulation of hydrodynamics and water quality in coastal areas (in Chinese). Prog Nat Sci 14:694–699

Smagorinsky J (1963) General circulation experiments with the primitive equations. I. The basic experiment. Mon Weather Rev 91:99–164

Xu RY, Zhao BR, Huang JZ et al (2006) The mean residual circulations in the Bohai Sea (in Chinese). Mar Sci 30:47–52

Zhang YL, Baptista AM, Myers EP (2004) A cross-scale model for 3D baroclinic circulation in estuary-plume-shelf systems: I. Formulation and skill assessment. Cont Shelf Res 24:2187–2214

Zhang LJ, Xia B, Gui ZS et al (2007) Contaminative conditions evaluation of 16 main rivers flowing into sea around Bohai Sea in Summer of 2005 (in Chinese). Environ Sci 28:2409–2415

Ocean Dynamics (2010) 60:1597–1619
DOI 10.1007/s10236-010-0321-3

Storm surge simulation along the U.S. East and Gulf Coasts using a multi-scale numerical model approach

Hongzhou Xu · Keqi Zhang · Jian Shen · Yuepeng Li

Received: 29 April 2010 / Accepted: 19 July 2010 / Published online: 10 August 2010

Abstract The effectiveness of simulating surge inundation using the Eulerian–Lagrangian circulation (ELCIRC) model over multi-scale unstructured grids was examined in this study. The large domain model grid encompasses the western North Atlantic Ocean, the Gulf of Mexico, and the Caribbean Sea to appropriately account for remote and resonance effects during hurricane events and simplify the specification of the open boundary condition. The U.S. East and Gulf Coasts were divided into 12 overlapping basins with fine-resolution (up to 30×30 m) grids to model overland surge flooding. These overlapping basins have different fine-resolution grids near the coastal region, but have an identical coarse-resolution grid in the offshore region within the large model domain. Thus, the storm surge prediction can be conducted without reducing computation efficiency by executing multiple model runs with local fine-resolution grids where potential hurricane landfalls may occur. The capability of the multi-scale approach was examined by simulating storm surge caused by Hurricanes Andrew (1992) and Isabel (2003) along the South Florida coast and in the Chesapeake Bay. Comparisons between simulated and observed results suggest that multi-scale models proficiently simulated storm surges in the Biscayne Bay and the Chesapeake Bay during two hurricanes. A series of sensitivity tests demonstrated that the simulation of surge flooding was improved when LiDAR topographic data and special bottom drag coefficient values for mangrove forests were employed. The tests also showed that appropriate representation of linear hydrologic features is important for computing surge inundation in an urban area.

Keywords Multi-scale grid model · Storm surge · Inundation prediction · LiDAR · ELCIRC · South Florida · Chesapeake Bay

Responsible Editor: Eric Deleersnijder

This paper was submitted to Ocean Dynamics' special issue on multi-scale modeling of coastal, shelf, and global ocean dynamics (code: MSM 2010).

H. Xu · K. Zhang · Y. Li
International Hurricane Research Center,
Florida International University,
Miami, FL 33199, USA

K. Zhang (✉)
Department of Earth and Environment,
Florida International University,
Miami, FL 33199, USA
e-mail: zhangk@fiu.edu

J. Shen
Virginia Institute of Marine Science,
College of William and Mary,
Gloucester Point, VA 23062, USA

1 Introduction

Coastal inundation caused by storm surge is the greatest threat from hurricanes to human life, coastal habitats, and ecosystems along the U.S. East and Gulf Coasts. To avoid loss of human life and damage of properties in the coastal zone, many numerical models have been developed to predict storm surge inundation; the models have proven to be very successful in disaster planning and mitigation, and coastal management (Flather and Proctor 1983; Flather et al. 1991; Luettich et al. 1992; Jelesnianski et al. 1992; Westerink et al. 1994; Zhang et al. 2008). In recent years, many studies have focused on improving the accuracy of numerical models, understanding physical processes, and predicting storm surge in real-time (Blain et al. 1994; Houston et al. 1999; Li et al. 2006; Peng et al. 2006; Shen

et al. 2006a, 2006b; Gong et al. 2007, 2009; Mattocks and Forbes 2008; Xie et al. 2008; Weisburg and Zheng 2008; Rego and Li 2010).

One important aspect of storm surge prediction is to compute the magnitude and extent of coastal inundation. Previous studies have shown that a fine-resolution model grid improves inundation simulation by resolving complex coastal topographic features affecting storm surge (Westerink et al. 1994; Shen et al. 2006b; Zhang et al. 2008). For example, the Sea, Lake, and Overland Surge from Hurricane (SLOSH) model developed by the National Oceanic and Atmospheric Administration (NOAA; Jelesnianski et al. 1992) has partitioned the U.S. East and Gulf Coasts into more than 30 overlapping basins to create high-resolution grids for modeling overland flooding (Glahn et al. 2009). Additionally, in order for a structured grid model such as SLOSH to cover a large area and maintain the fine-resolution near the coast without losing computational efficiency, a polar, elliptical, or hyperbolic grid with gradually varying cell size is usually chosen to represent the model domain. This allows the model grid to cover a basin extending from the inland area possibly flooded by storm surge to the deep water about 150 to 200 km offshore. The weakness of the structured grid models is that model domains often do not cover the open ocean sufficiently to account for the remote effect and cannot adequately represent complex shoreline geometry due to the inflexibility of grid cell shapes and sizes. The structured grid may also generate spurious flows at the grid corners near the shoreline area (Aldridge and Davies 1993; Davies and Jones 1996).

Davies and Lawrence (1995) emphasized that interaction of wind waves/current plays an important role in coastal physical processes. It especially affects surge height largely at coastal zones (Jones and Davies 1998). Besides, the remote wind effect has a considerable impact on storm surge development near the coast. Blain et al. (1994) simulated the effect of remote forcing at the open ocean on the coastal water by comparing domains with various sizes for modeling surge caused by Hurricane Kate. Mathew et al. (1996) and Shen and Gong (2009) have presented evidence on the importance of remote effects and shown the improvement of storm surge prediction when using a large model domain. Shen and Gong (2009) indicated that a large model domain encompassing the U.S. East Coast was needed to account for the effect of the Ekman transport by modeling the surge caused by Hurricane Ernesto in 2006.

There is considerable uncertainty in forecasting hurricane tracks and intensity that determine the wind and pressure fields driving storm surge models. The longer a forecast period is, the larger the cone of uncertainty. It is not uncommon that a 72-h cone of forecasted tracks covers several hundred to 1,000 km of shorelines along the U.S. East and Gulf Coasts. Obviously, all coastal areas influenced by cone of tracks have to be included in the model grids to compute possible surge inundation. In addition, any portion of low-lying areas along the U.S. East and Gulf Coasts can be flooded by storm surge during a hurricane season thus, the fine-resolution model grids covering the entire U.S. East and Gulf Coasts are required to perform storm surge forecast. However, it is not feasible to run the model over such a large domain with a fine-resolution grid due to the limitation of current computation power required to solve the governing equations numerically.

The requirement of the large domain for modeling the surge forerunner caused by the wind at the open ocean and the fine-resolution grid near the coast for modeling overland flooding poses a difficulty for surge simulation. One way to overcome this difficulty is to use a multi-scale modeling approach in which coarse-resolution grids are usually created with boundaries at the open ocean to derive relatively simple boundary conditions and fine-resolution grids are created in coastal areas to specify complex topographic features. Recent applications of a multi-scale model were found at the northeastern Gulf of Mexico (Blain et al. 1998; Rego and Li 2010), South Florida (Shen et al. 2006b; Zhang et al. 2008), South Carolina (Peng et al. 2006; Xie et al. 2008), North Carolina (Mattocks and Forbes 2008), and the Chesapeake Bay (Li et al. 2006; Shen et al. 2006c).

Both nested-grid and unstructured-mesh approaches have been applied extensively to multi-scale modeling (Pietrzak et al. 2005, 2006; Deleersnijder and Lermusiaux 2008; Ham et al. 2009). A nested-grid model usually has a large domain with a coarse-resolution grid and a nested small domain with a fine-resolution grid where the simulation over the coarse-resolution grid provides open boundary conditions for the model over the fine-resolution grid. However, model errors are often generated when data are interpolated from the cells of the large domain to those of the small domain at the interface of the two domains. Additional errors can also be generated when the hydrodynamics of the small domain affects that of the large domain in a bi-directionally coupled, nested-grid model. The errors are especially large at regions of abrupt topography (Davies and Hall 2002). Such errors can be reduced by using a multi-scale unstructured model because the unstructured-mesh approach makes smooth transitions between the offshore coarse-resolution grid and the nearshore fine-resolution grid. Unlike the nested-grid approach, in which the model runs over the coarse-resolution and fine-resolution grids separately, the model runs over one grid, which integrates the coarse-resolution and fine-resolution grids, in a multi-scale unstructured approach. However, the application of the multi-scale unstructured model approach to simulate storm surge along the U.S. East and Gulf Coasts has not been investigated extensively.

Coastal topographic elevation plays an important role in determining the overland inundation caused by storm surge. Topographic elevations are usually derived from topographic maps or digital elevation models (DEMs) provided by the U. S. Geological Survey (USGS) in the United States. The vertical resolution of USGS DEMs is about 1.5 m (5 feet) for low-lying areas. This vertical interval transfers horizontally from several to tens of kilometers in low-relief urban areas such as South Florida or the Florida Panhandle and can result in large errors in flooding prediction. Also, linear topographic and hydrologic features, and man-made structures such as sand ridges, major highways, flood walls, levees, and canals that can either block or facilitate surge flows needed to be measured and represented in the model grid for modeling overland flood. However, these linear features are often not represented well by the USGS contour maps. Recent advances in airborne light detection and ranging (LiDAR) technology allow rapid and inexpensive measurements of the Earth's surface topography and bathymetry over large areas. This technology is becoming the primary method for generating high-resolution DEMs and for capturing the elevations of linear features with a vertical accuracy of 0.15 m and a horizontal accuracy of 0.45 m (Whitman et al. 2003). For example, the State of Florida has collected LiDAR data for coastal areas vulnerable to storm surge flooding at a cost of $25 million. However, the effect of LiDAR DEMs and linear features on surge simulation has not been examined.

The main objective of this study is to explore the feasibility of using multiple multi-scale, unstructured models to perform surge flooding forecasts for the U.S. East and Gulf Coasts, to examine the influence of the uncertainty in hurricane track forecasting on the simulation of surge flooding, to evaluate the effect of the DEM accuracy on model performance, and to test the sensitivity of model prediction to the bottom friction force. The unstructured Eulerian–Lagrangian CIRCulation model (ELCIRC; Zhang et al. 2004) was selected for modeling tides along the U.S. East and Gulf Coasts and surges from Hurricanes Andrew and Isabel in order to save simulation time because the ELCIRC model required less computation resources compared to other unstructured models and can be run on a PC workstation. The structure of this paper is as follows: The numerical model is described in Section 2, tidal and storm surge simulations are shown in Sections 3 and 4, respectively, and discussion and summary are presented in Section 5.

2 Model description

2.1 Hydrodynamic model

The ELCIRC model was developed by Zhang et al. (2004), using numerical algorithms that are similar to the Unstructured Tide, Residual Intertidal Mudflat (UnTRIM) model (Casulli and Zanolli 1998; Casulli 1999). ELCIRC is a general three-dimensional model capable of simulating both two-dimensional (vertically averaged) and three-dimensional hydrodynamics and transport processes. Both triangular and four-sided polygon cells can be used in the model grid, providing flexibility to fit linear topographic and man-made features. The model uses a semi-implicit Eulerian–Lagrangian finite volume/finite difference algorithm to solve the shallow water equation that allows a large time step for simulation. ELCIRC includes a robust wetting-and-drying method to simulate overland surge flooding (Cho 2009; Gong et al. 2009). The water depth is given at the edge of a grid cell rather than the cell center in order to better fit the shoreline and morphological features. For each time step (n+1), the total water depth at edge j of a grid cell is computed as:

$$H_j^{n+1} = \max(h_{\min}, h_j + \eta_{i(j,1)}^{n+1}, h_j + \eta_{i(j,2)}^{n+1}) \quad (1)$$

Where $h_{\min}$ is the minimal water depth, h_j is the depth on the jth edge of the cell polygon, and $\eta_{i(j,1)}^{n+1}$ and $\eta_{i(j,2)}^{n+1}$ are water surface displacements on the centers of the cell polygons at the left and right sides of edge j, respectively. The grid cell is defined as a 'wet' cell when $H_j^{n+1} > h_{\min}$. The cell is defined as dry only if the total depths at all edges are less than $h_{\min}$. In this study, $h_{\min}$ of 0.02 m is chosen as the wet cell criterion.

A vertically averaged 2-D model was employed for tide and storm surge simulations. The surface wind stress is calculated by the following quadratic equation:

$$\overrightarrow{\tau}_w = \rho_a C_w \left|\overrightarrow{U_w}\right| \overrightarrow{U_w} \quad (2)$$

in which $\overrightarrow{U_w}$ is wind vector, ρ_a is air density, and C_w is the drag coefficient based on the formula of Garratt (1977),

$$C_w = (0.75 + 0.067\left|\overrightarrow{U_w}\right|) \times 10^{-3} \quad (3)$$

The bottom stress is calculated by the following equation:

$$\overrightarrow{\tau_b} = \rho_0 C_D \left|\overrightarrow{U_b}\right| \overrightarrow{U_b} \quad (4)$$

where $\overrightarrow{U_b}$ is current vector and ρ_0 is water density. There are several formulations for computing the bottom drag coefficient. We used Manning's formula, in which the roughness can vary based on land cover types. The bottom drag coefficient C_D is calculated by:

$$C_D = \frac{gn^2}{\sqrt[3]{H}} \quad (5)$$

where n is Manning's friction coefficient. H is the sum of water depth (H_0) and surface water elevation (η). Equation 5 suggests that C_D varies with H and n. A constant C_D of 0.0025 was used for water depths greater than 500 m for tide simulation.

2.2 Model domain and grid

In order to accommodate the remote wind effect at the open ocean, a large domain model grid was created first (referred to R0 hereafter), as shown in Fig. 1. This domain encompasses the U.S. East Coast, a portion of the Northern Atlantic Ocean, the Gulf of Mexico, and the Caribbean Sea (Fig. 1b). The model grid with 57,344 nodes and 106,389 edges extends from −98°W to −60°W in longitude and from 8°N to 46°N in latitude. Resolutions of grid cells range from 300 m near the coast to 25 km along the open ocean. Model depths at the open ocean were interpolated from the ETOPO1 global relief dataset from NOAA, which has a resolution of 1 min (1.5~2 km). Model depths in coastal areas were interpolated from the U.S. coastal relief dataset with 3-s resolution from NOAA (http://www.ngdc.noaa.gov/mgg/gdas/gd_desi-gnagrid.html). Both relief datasets are referenced to mean sea level.

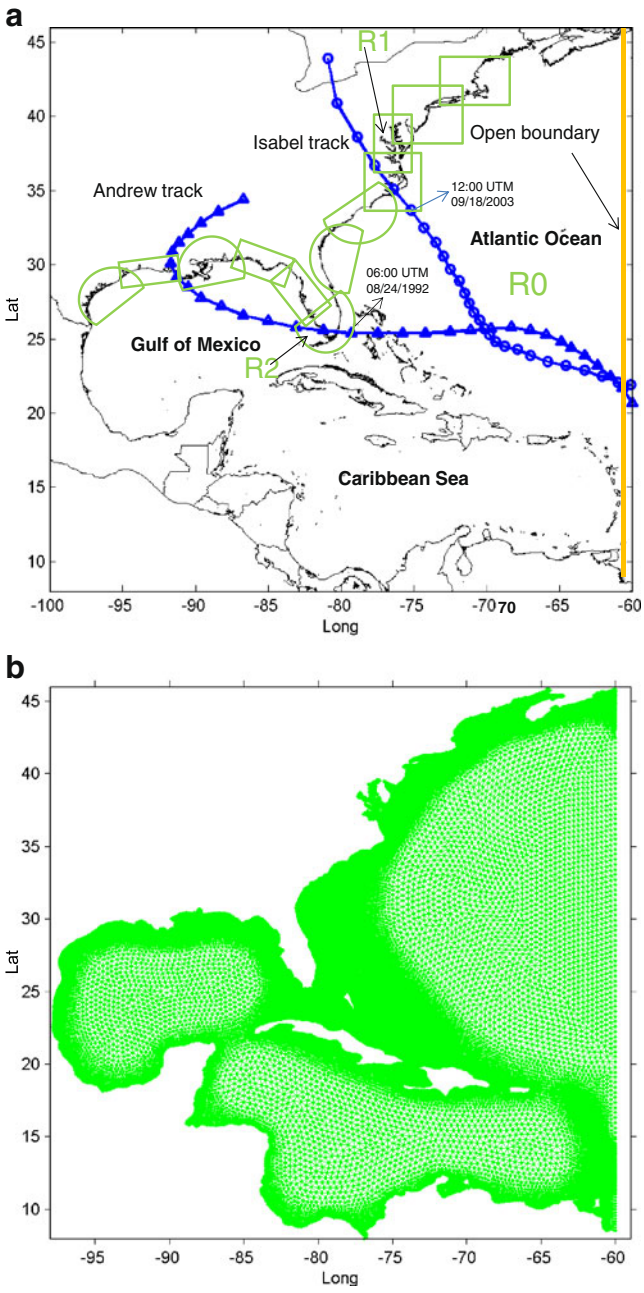

Fig. 1 **a** Twelve overlapping basins with fine-resolution grids covering the U.S. East and Gulf coasts and tracks of Hurricanes Andrew (1992) and Isabel (2003). R1 and R2 represent the basins for simulating surge induced by Isabel and Andrew. Note that R1 and R2 cover the same spatial area as R0 does, but with fine-resolution grid cells in coastal basins. The time interval for the hurricane center positions is 6 h. The time of track position before the hurricane makes landfall is also marked. **b** Coarse-resolution grid (R0) for the large model domain

To reach a good compromise between CPU cost and accuracy, a static mesh refinement strategy is adopted here to create grid cells for modeling coastal flooding. The coastal region of R0 was divided into 12 basins (Fig. 1a). The coarse grids within each basin were refined to resolve the complex coastal topography based on available fine-resolution topographic data. Therefore, 12 independent model grids with the same spatial extent as R0 but with a different resolution in each basin (Fig. 1a) were created. Because of the use of unstructured grids, the fine resolution of model grids in each basin can be gradually reduced toward the offshore region for a smooth transition between coarse grids and local fine grids. These twelve overlapping basins were used to model overland flooding and accommodate the uncertainty in the forecast of hurricane tracks (Fig. 1a). For a particular hurricane track, multiple model runs were conducted simultaneously over the basins that are influenced by the track. With twelve basins covering the entire U.S. East and Gulf Coasts, the overland inundation caused by any landfall hurricanes can be computed during a hurricane season. The fine-resolution grids for the Chesapeake Bay region (referred to R1 hereafter) and the South Florida Coast (referred to R2 hereafter) were employed to test the performance of the ELCIRC model for Hurricanes Isabel and Andrew (Fig. 1a).

The domain R1 for the Chesapeake Bay basin was comprised of triangular cells with totals of 120,658 nodes and 220,662 segments. In R1, the fine-resolution grid cells, which cover the Chesapeake Bay with sizes ranging from 200 m to 1 km, replaced the coarse-resolution portion of R0 for the Bay and connected with the remaining portion of R0 in the area approximately 200 km offshore from the Bay mouth through a varying-size grid transition zone (Fig. 2a).

The domain R2 has fine-resolution grid cells along the South Florida Atlantic Coast, especially in Miami (Fig. 2b). A low-lying area along the coast that has the potential to be flooded by storm surge was identified by analyzing topographic sheets. The 30 m×30 m USGS DEM were used to generate elevation values for grid cells over the land. The corresponding 30 m×30 m DEM from the

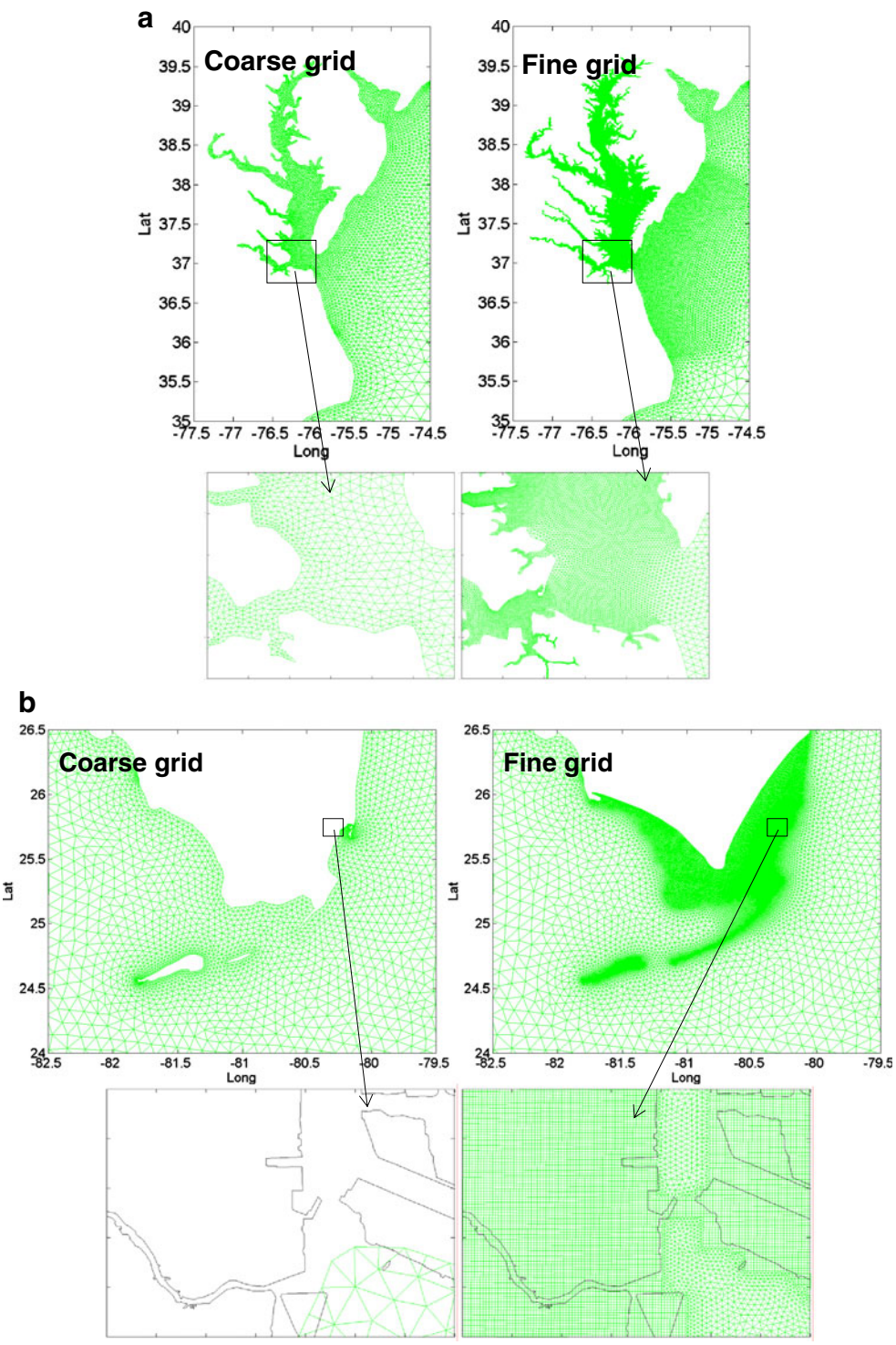

Fig. 2 Comparisons of fine-resolution and coarse-resolution model grids for **a** the Chesapeake Bay and **b** the South Florida. *Black line* represents the coastline

LiDAR surveys performed by the International Hurricane Research Center from 1998 to 2003 (Whitman et al. 2003) were also used to extract elevation values for grid cells for a comparison purpose. Most coastal areas vulnerable to storm surge flooding in R2 were covered by LiDAR data. The coverage map for the LiDAR DEM in South Florida can be found in Zhang (2010).

The nodes and edges of the 30 m×30 m DEM was directly converted into the four-sided polygon cells of the model grid covering the low-lying area in order to use the full resolution of the topographic data. Using this simple approach, the model grid is able to resolve topographic changes resolved by the DEM as indicated by an example for the portion of the Miami downtown area in Fig. 3. The

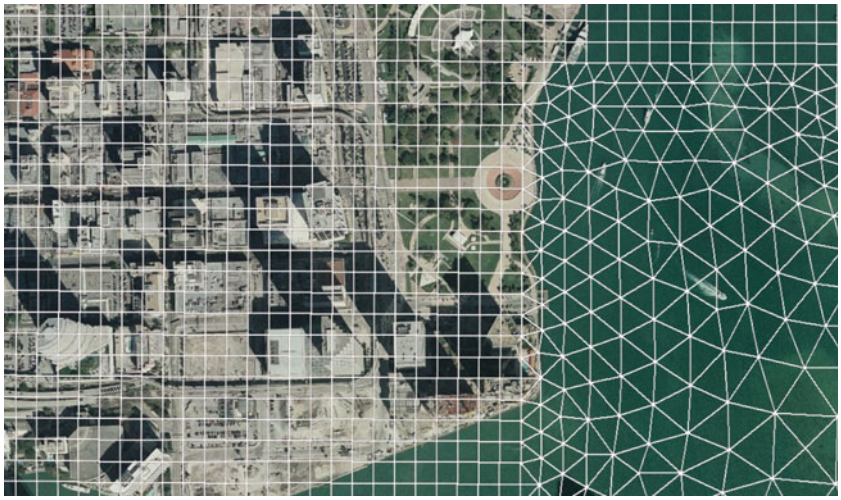

Fig. 3 Fine-resolution quadrangle grid (*white line*) in the eastern Miami downtown area. The quadrangles in the grid were derived from grid cells of the 30 m×30 m USGS DEM

approach has the potential to generate a street-level grid when the higher-resolution data become available. Since the topographic DEMs referred to the NAVD 88 vertical datum, the elevation values of grid cells were adjusted from the NAVD 88 datum to mean sea level using a 0.14 m offset between two vertical data, which was derived based on the water level records of tide gauges along the Miami Coast. The four-sided polygon cells over the land were first connected to the triangular cells for the coastal ocean, and then grid resolution was gradually decreased and merged with the coarse-resolution cells in R0 for the open ocean (Fig. 2b). The resolution of the R2 grid with totals of 609,152 nodes and 706,269 edges reduces gradually from onshore toward offshore.

3 Case study of tidal propagation

Prior to storm surge simulation, the simulation of tides along the U.S. East and Gulf Coasts was carried out to verify the ELCIRC model by using the R0 grid. The model was forced by 9 main tidal constituents at its open boundary, namely M_2, S_2, K_2, N_2, L_2, K_1, O_1, P_1, and Q_1, which were obtained from the ADCIRC 2DDI tidal database (Mukai et al. 2002). Tidal potentials from the tidal database were also applied to the interior cells of the model domain. The simulation of tide using a time step of 3 min started from 1/1/2000, 00:00 Eastern Standard Time and lasted for 60 days. The simulated hourly tidal levels of the last 29 days (Boon 2004) were used for comparison with predicted tides at 18 tidal gauge stations from the NOAA CO-OPS dataset (http://tidesandcurrents.noaa.gov; Fig. 4). Figure 5 shows both computed and NOAA predicted water levels during the last 10 days of the simulation. It can be seen that the model adequately simulated tidal elevation along the U.S. East Coast with correlation coefficients greater than 0.96 (Fig. 5a). The simulated tides along the Gulf Coast do not match the predicted ones as well as those along the East Coast with

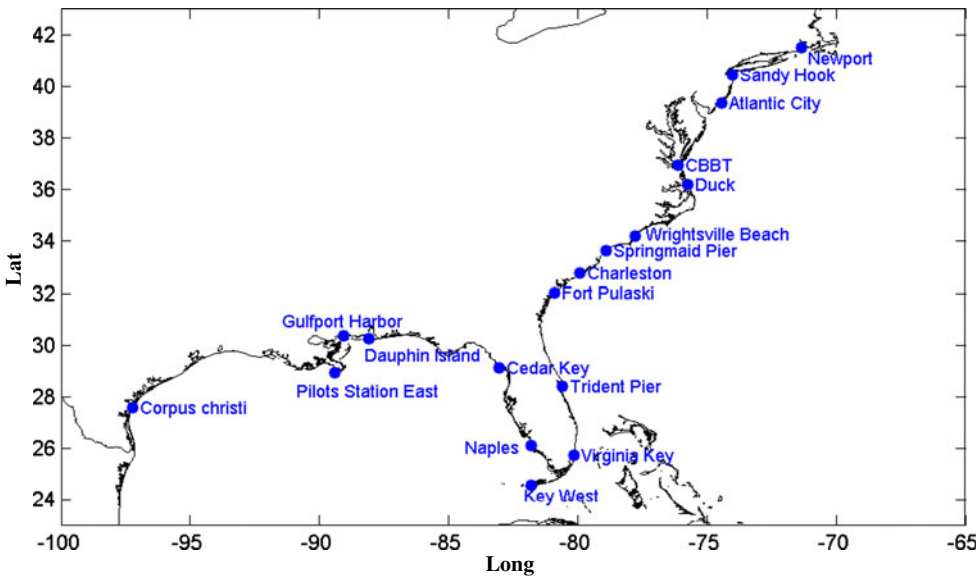

Fig. 4 Locations of 18 tidal stations along the U.S. East and Gulf Coasts for calibrating tide simulation over the coarse-resolution grid for the large domain

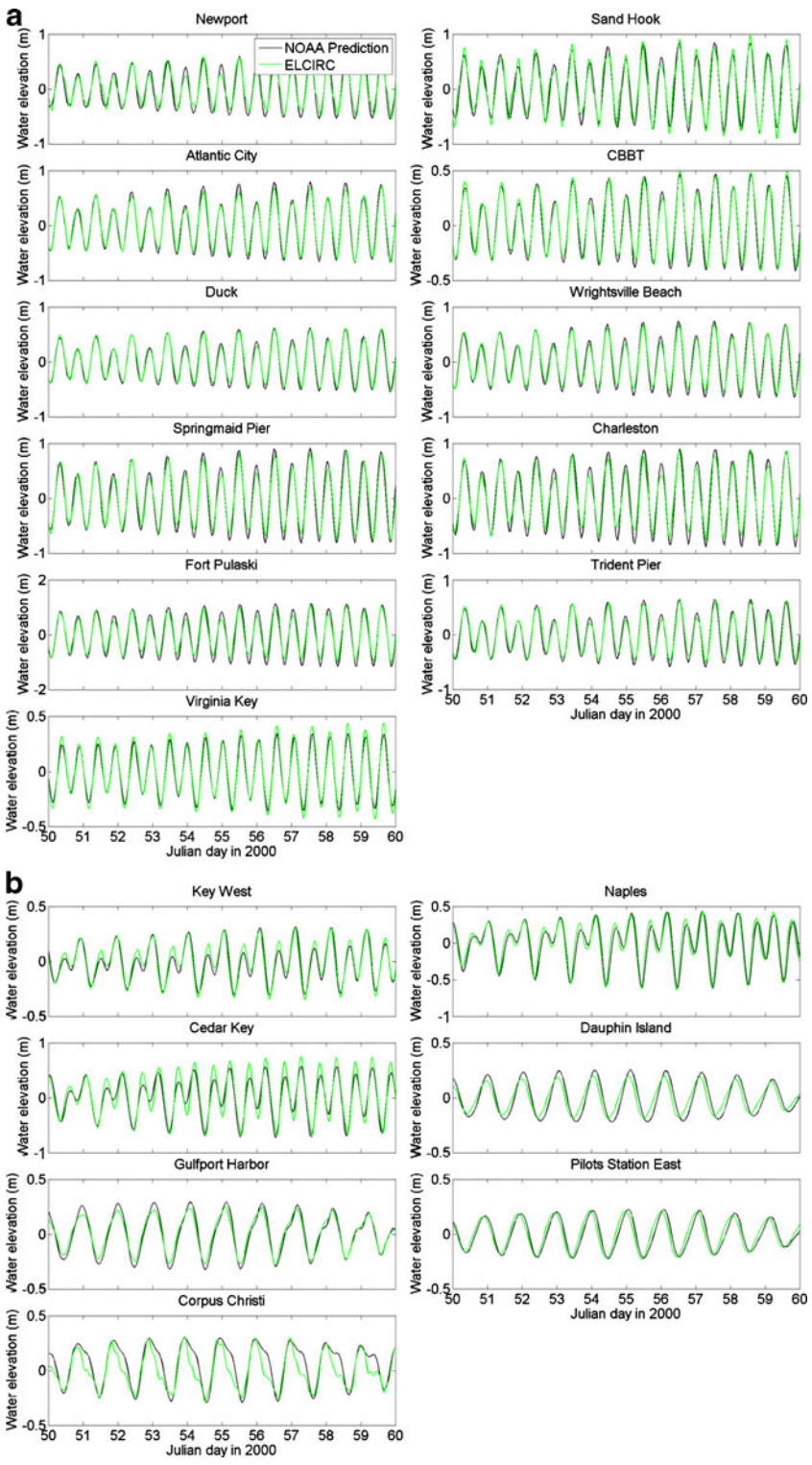

Fig. 5 Comparisons of computed tidal level (*green line*) and predicted tidal level from the NOAA (*black line*) for **a** 11 stations along the East Coast and **b** 7 stations along the Gulf Coast. Both computed and predicted NOAA tides are referenced to mean sea level and the unit of the tidal level is in meters

correlation coefficients ranging from 0.97 to 0.86 (Fig. 5b). The lowest correlation coefficient of 0.86 is located at Dauphine Island, probably as a result of poor representation of the complex coastal bathymetry by the R0 grid and the complex tidal resonance occurring in the nearly enclosed Gulf of Mexico. The harmonic analysis of computed hourly tidal levels during the last 29 days of the simulation at 18 stations was also conducted to compare simulated amplitudes and phases of six major tidal components with those in the NOAA data set (Tables 1 and 2). The vectorial

differences between simulated and observed harmonic constants are calculated by the formula (Tsimplis et al. 1995):

$$\text{Diff} = \left[(A_o \cos P_o - A_m \cos P_m)^2 + (A_o \sin P_o - A_m \sin P_m)^2\right]^{1/2} \quad (6)$$

where A_o, A_m, P_o, and P_m are the observed and simulated amplitudes and phases. The results of harmonic analysis indicate that the U.S. East Coast is dominated by semidiurnal tides (M_2, S_2 and N_2), while the Gulf Coast is dominated by diurnal tides (K_1 and O_1). The vectorial differences in M_2 tide along the East Coast are less than 0.06 m at most stations except for the Sandy Hook station in New Jersey, Charleston station in South Carolina, the Fort Pulaski station in Georgia, and the Naples station in Florida; the vectorial differences are 0.11, 0.11, 0.16 and 0.11 m, respectively (Table 1). These four stations are located at areas with convoluted shorelines, which the coarse-resolution grid is not fully able to resolve. The differences in M_2 phase along the East Coast are less than 10° while the differences at several stations such as Naples, Dauphin Island and Corpus Christi, along the Coast of the Gulf of Mexico, are greater than 20° (Table 2). Both mean and root-mean-square (RMS) differences in M_2 tidal amplitudes at all stations are 0.01 m (Table 3). The mean and RMS differences in M_2 tidal phase are 2.50° and 2.44°, respectively. Most of the vectorial differences between computed S_2 tidal amplitudes and those in the NOAA data set are less than 0.03 m, with large errors occurring at stations along the North Carolina, South Carolina, Georgia, and Florida Coasts (Table 1), where the errors of phases are also large (Table 2). Both mean and RMS differences for S_2 tidal amplitude are 0.02 m, and S_2 tidal phase are 5.58° and 7.28°, respectively (Table 3). The vectorial differences between computed K_1 and O_1 constituents and those in the NOAA data set are relatively larger along the Gulf Coast than those along the East Coast (Tables 1 and 2). The mean differences for K_1 and O_1 tidal amplitudes are the same, with values of 0.01 m (Table 3). The RMS differences for K_1 and O_1 tidal amplitudes are also the same, with values of 0.02 m. These results indicate that the simulation of tide by the ELCIRC model is overall satisfactory along the U.S. East and Gulf Coasts.

The simulations of tides were also performed over basins R1 and R2. The results at three tidal gauge stations (CBBT, Duck, and Virginia Key) show that the refined grids improve the tide simulation along the East Coast. For example, the vectorial differences in M_2 tide were reduced from 0.03 and 0.06 m (Table 1) to 0.01 and 0.02 m at the Duck and Virginia Key stations, respectively. However, the tide along the Gulf coast at Key West is not improved, possibly due to the effect of tide resonance in the Gulf of Mexico. The locally refined grid does not improve the simulation of the basin-wide tide resonance in the Gulf of Mexico.

4 Case study of storm surge simulation

The simulations of storm tide, the combination of storm surge and astronomical tide (Boon 2004) generated by Hurricanes Andrew and Isabel, were carried out to examine the effectiveness of the multi-scale approach. Storm surge is an abnormal water level change caused by atmospheric pressure drop and strong wind field induced by hurricanes. The parameter wind model of SLOSH, which was developed by the National Weather Service (Myers and Malkin 1961; Jelesnianski et al. 1992), was employed to generate both pressure and wind fields for simulating surges. The wind and pressure fields were computed in terms of hurricane track, atmospheric pressure drop, and the radius of maximum wind based on a stationary, circularly symmetric storm by the balance of the force along and normal to a surface wind trajectory with correction of storm movement (Jelesnianski et al. 1992). The wind speed profile for a stationary storm is described as:

$$V(r) = V_{\max} \frac{2 \times R_{\max} \times r}{R^2{}_{\max} + r^2} \quad (7)$$

where $V_{\max}$ is the maximum wind speed, r is the radius, and $R_{\max}$ is the radius of maximum wind, respectively. To deal with the territorial impacts on hurricane wind, the SLOSH wind model generates different wind fields over the land and ocean, which are referred as the lake and ocean winds by using different friction coefficient in the wind model (Jelesnianski 1967; Jelesnianski et al. 1992). In this study, wind and pressure fields during Hurricanes Andrew (1992) and Isabel (2003) were computed based on the best track records by following the same procedure used by Shen et al. (2006b). The central pressure at each time step was interpolated based on reported hourly data. In addition, nine tidal constituents at the open boundary, namely M_2, S_2, K_2, N_2, L_2, K_1, O_1, P_1, and Q_1, were used to force the model. Because the open boundary is far away from the coast and storm surge is often relatively small at the open ocean, it is assumed that there is no storm surge at the model open boundary during the initial stage.

4.1 Hurricane Isabel (2003)

Hurricane Isabel originated from a tropical wave from the Coast of Africa on September 1, 2003. It developed to a tropical storm on September 6 and became a Category 5 hurricane on September 11 as the storm moved toward the

Table 1 Comparisons of tidal amplitudes (in meters) derived from harmonic analysis of simulated and NOAA predicted tidal levels. The difference was calculated in terms of Equation 6

Station	M2			S2			N2			K1			O1			K2		
	Mod	Obs	Diff	Mod	Obs	Diff	Mod	Obs	Diff	Mod	Obs	Diff	Mod	Obs	Diff	Mod	Obs	Diff
Newport	0.48	0.50	0.02	0.08	0.10	0.02	0.12	0.12	0.01	0.08	0.06	0.02	0.06	0.05	0.02	0.02	0.03	0.01
Sandy Hook	0.71	0.69	0.11	0.11	0.13	0.02	0.16	0.16	0.00	0.08	0.10	0.02	0.06	0.05	0.01	0.03	0.04	0.01
Atlantic City	0.59	0.60	0.01	0.08	0.11	0.03	0.13	0.14	0.02	0.08	0.11	0.03	0.06	0.07	0.01	0.02	0.03	0.01
CBBT	0.39	0.38	0.01	0.05	0.07	0.02	0.09	0.09	0.01	0.06	0.06	0.01	0.05	0.05	0.01	0.01	0.02	0.01
Duck NC	0.49	0.49	0.03	0.06	0.09	0.03	0.11	0.11	0.02	0.09	0.09	0.01	0.06	0.06	0.01	0.02	0.02	0.00
Wrightsville	0.57	0.59	0.03	0.06	0.10	0.04	0.13	0.13	0.02	0.09	0.09	0.01	0.07	0.07	0.01	0.02	0.03	0.01
Springmaid	0.71	0.74	0.06	0.07	0.12	0.05	0.17	0.18	0.04	0.10	0.10	0.02	0.07	0.07	0.00	0.02	0.03	0.01
Charleston	0.73	0.78	0.11	0.07	0.12	0.05	0.17	0.17	0.00	0.10	0.10	0.03	0.07	0.08	0.01	0.02	0.03	0.01
Fort Pulaski	0.92	1.01	0.16	0.09	0.16	0.07	0.21	0.22	0.01	0.11	0.11	0.04	0.08	0.08	0.01	0.02	0.04	0.03
Trident Pier	0.51	0.52	0.06	0.05	0.08	0.03	0.12	0.12	0.03	0.10	0.10	0.04	0.08	0.08	0.00	0.01	0.02	0.01
Virginia Key	0.37	0.30	0.06	0.05	0.05	0.00	0.08	0.07	0.01	0.04	0.03	0.01	0.04	0.03	0.01	0.01	0.01	0.00
Key West	0.19	0.19	0.02	0.05	0.05	0.01	0.04	0.04	0.00	0.08	0.09	0.01	0.08	0.09	0.04	0.01	0.01	0.01
Naples	0.28	0.29	0.11	0.11	0.10	0.07	0.05	0.06	0.01	0.12	0.16	0.04	0.11	0.14	0.05	0.03	0.03	0.03
Cedar Key	0.41	0.39	0.03	0.16	0.13	0.03	0.05	0.06	0.01	0.14	0.18	0.06	0.11	0.16	0.05	0.01	0.04	0.04
Dauphin Island	0.02	0.02	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.10	0.14	0.04	0.09	0.14	0.07	0.01	0.01	0.00
Gulfport Harbor	0.03	0.04	0.01	0.03	0.03	0.00	0.01	0.01	0.00	0.14	0.17	0.03	0.11	0.16	0.07	0.01	0.01	0.00
Pilots Station	0.01	0.02	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.12	0.13	0.01	0.10	0.13	0.06	0.01	0.01	0.00
Corpus Christi	0.08	0.08	0.03	0.02	0.02	0.00	0.02	0.02	0.01	0.13	0.160	0.03	0.11	0.16	0.08	0.01	0.01	0.00

Table 2 Comparisons of tidal phases (in degrees) derived from harmonic analysis of simulated and NOAA predicted tidal levels. The difference was calculated by subtracting the modeled phase from the observed phase

Station	M2			S2			N2			K1			O1			K2		
	Mod	Obs	Diff	Mod	Obs	Diff	Mod	Obs	Diff	Mod	Obs	Diff	Mod	Obs	Diff	Mod	Obs	Diff
Newport	226.84	227.28	−0.44	28.04	29.44	−1.40	337.84	332.27	5.57	194.76	205.25	−10.49	40.59	62.19	−21.60	237.80	238.91	−1.11
Sandy Hook	221.71	231.28	−9.57	40.11	41.66	−1.55	334.37	333.55	0.82	196.71	196.29	0.42	46.32	34.13	12.19	278.65	260.30	18.35
Atlantic City	219.48	220.67	−1.19	26.24	25.27	0.97	329.74	320.81	8.93	188.62	189.94	−1.32	40.26	27.80	12.46	212.02	242.34	−30.32
CBBT	246.64	246.26	0.38	56.74	54.58	2.16	356.47	345.89	10.58	206.77	189.25	17.52	55.66	69.72	−14.06	239.97	273.33	−33.36
Duck NC	226.41	223.25	3.16	30.65	28.81	1.84	335.57	321.91	13.66	192.42	197.31	−4.89	41.43	54.04	−12.61	224.35	246.56	−22.21
Wrightsville	221.54	218.44	3.10	29.53	22.59	6.94	331.44	320.21	11.23	199.34	188.45	10.89	49.74	57.03	−7.29	238.37	237.95	0.42
Springmaid	227.65	222.84	4.81	38.61	26.19	12.42	338.07	323.08	14.99	201.43	187.94	13.49	52.04	54.24	−2.20	266.29	245.29	21.00
Charleston	227.60	235.74	−8.14	42.01	50.61	−8.60	338.67	339.03	−0.36	202.26	184.32	17.94	52.79	65.11	−12.32	299.14	282.68	16.46
Fort Pulaski	235.16	243.19	−8.03	57.57	59.76	−2.19	347.34	346.29	1.05	206.56	185.67	20.89	57.26	68.25	−10.99	345.44	288.37	57.07
Trident Pier	240.44	232.55	7.89	39.91	37.28	2.63	348.25	329.68	18.57	214.38	186.57	27.81	65.31	68.38	−3.07	226.67	261.90	−35.23
Virginia Key	264.00	266.26	−2.26	90.11	82.25	7.86	14.50	4.79	9.71	268.31	246.77	21.54	127.69	150.34	−22.65	323.81	304.37	19.44
Key West	284.90	292.32	−7.42	80.46	92.72	−12.26	36.84	32.20	4.64	345.53	339.61	5.92	186.74	214.15	−27.41	226.91	294.66	−67.75
Naples	−13.39	9.73	−23.12	118.20	160.11	−41.91	106.23	113.28	−7.05	5.67	−5.48	11.15	206.29	224.73	−18.44	265.18	342.43	−77.25
Cedar Key	51.03	55.07	−4.04	224.30	234.15	−9.85	177.55	167.22	10.33	39.74	20.36	19.38	244.42	249.62	−5.20	−47.98	80.32	−128.30
Dauphin Island	338.92	358.92	−20.00	120.82	135.89	−15.07	137.40	101.13	36.27	35.62	38.55	−2.93	232.03	263.86	−31.83	19.88	62.42	−42.54
Gulfport Harbor	34.42	34.44	−0.02	156.90	171.84	−14.94	162.84	146.35	16.49	37.45	28.46	8.99	233.43	254.46	−21.03	342.19	352.90	−10.71
Pilots Station	343.97	348.81	−4.84	85.89	115.59	−29.70	148.08	128.65	19.43	8.88	9.39	−0.51	206.27	234.17	−27.90	299.83	316.11	−16.28
Corpus Christi	149.91	125.24	24.67	279.65	277.38	2.27	252.13	223.26	28.87	12.65	13.29	−0.64	209.09	238.97	−29.88	86.341	127.49	−41.15

Table 3 Mean and RMS differences at all stations for tidal constitutes

	Mean difference		RMS difference	
	Amplitude (m)	Phase (deg)	Amplitude (m)	Phase (deg)
M2	0.01	2.50	0.01	2.44
S2	0.02	5.58	0.02	7.28
N2	0.01	−11.3	0.01	−9.58
K1	0.01	−8.62	0.02	−8.25
O1	0.01	13.52	0.02	15.9
K2	0.01	20.75	0.01	10.5

U.S. East Coast. Isabel reduced to a Category 2 hurricane and made landfall along the Eastern Coast of North Carolina at 17:00 Universal Time Coordinate (UTM) on September 18 with a maximum wind speed of 43 m/s and a 56 mb pressure drop. The storm surge flooding induced by Hurricane Isabel caused severe flood damage in the low-lying areas of the Chesapeake Bay. The maximum water level reached 1.86 m above the mean sea level at the CBBT station near the bay mouth, decreased to 1.48 m at Lewisetta in the middle portion of the Bay, and increased to 2.10 m at Baltimore in the upper Bay (Figs. 6 and 7).

The model simulation was conducted for 6 days from September 15 to 22 (referred to as base case EI0 hereafter), producing the maximum Chesapeake Bay storm tides (water levels) as shown in Fig. 6a. The model simulation shows that high storm tides occurred in the upper Bay north of Baltimore, along the south side of the Bay entrance near Sewells Pt., and tributaries along the west side of the Bay, while relatively low storm tides occurred along the east side of the Bay below the Potomac River. The comparison of computed and observed time series of storm tides at ten stations (Fig. 6b) indicates that the model adequately simulated the peaks of storm tides with a mean correlation coefficient of 0.97 and a mean RMS error of 0.11 m, but the phases between computed and observed maximum tides were off by several hours in the upper Bay (Fig. 7), which is partially due to the error in the wind field generated by an analytical wind model (Shen 2009).

4.2 Hurricane Andrew (1992)

Hurricane Andrew formed along the West Coast of Africa and drifted to the Northern Atlantic Ocean on August 14, 1992, as a tropical wave. It became a tropical storm on August 20 and

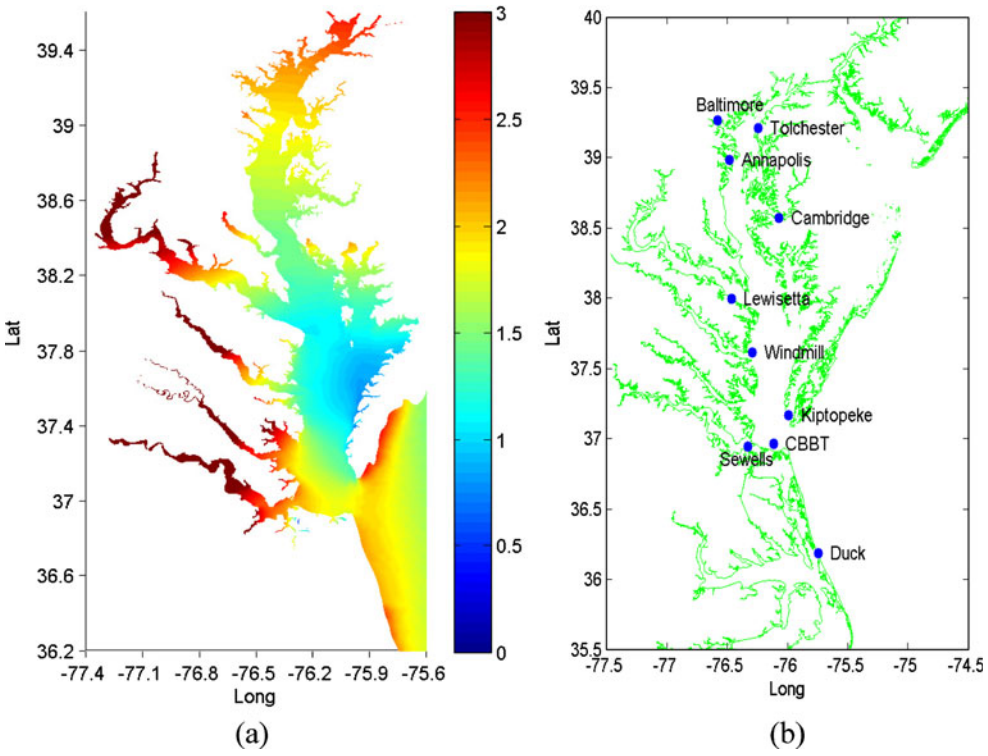

Fig. 6 **a** Computed maximum storm tides referenced to means sea level in the Chesapeake Bay (EI0), and **b** the locations of tidal stations in the Chesapeake Bay. The unit of storm tide is in meters

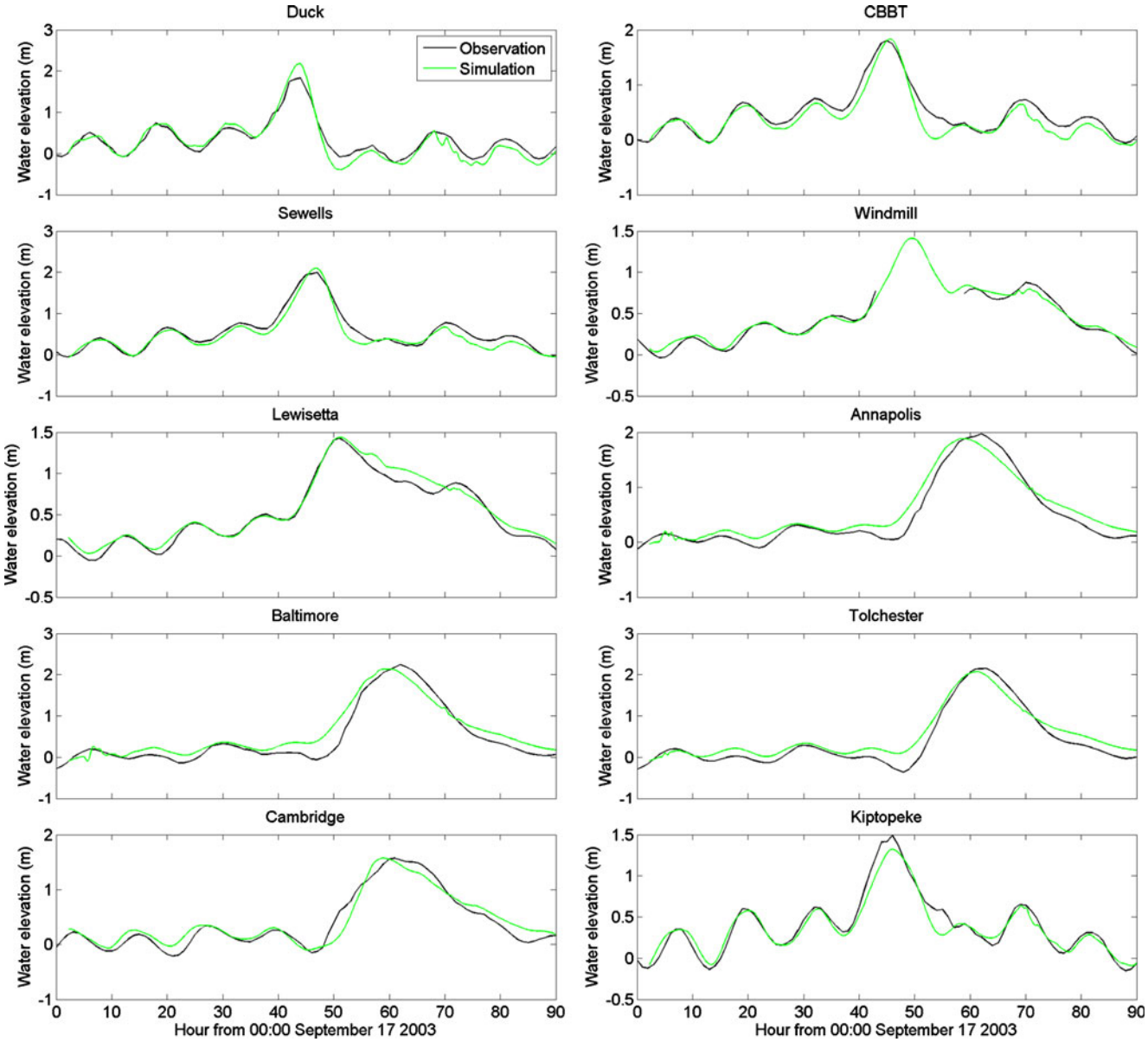

Fig. 7 Computed (*green line*) and observed (*black line*) storm tides for ten stations in the Chesapeake Bay (EI0). Both computed and observed tides are referenced to mean sea level and the unit of storm tide is in meters

reached hurricane strength on August 22. It made landfall at Miami–Dade County, Florida on August 24 with a minimum center pressure of 992 mb and a peak wind speed of 63 m/s. The maximum storm tide caused by Andrew reached 5.2 m at the old headquarters of Burger King Inc., located in the central coast of Biscayne Bay (Zhang et al. 2008). The simulation of storm tides was conducted from August 15 to 25 (referred to as base case EA0 hereafter). Two types of field observed data, high-water-mark elevations, and debris lines, were used to compare with the model results. The elevations of high-water marks that were generated by storm tides on buildings and trees represent the highest water levels reached at those locations during a hurricane, while the debris lines represent the landward extents of storm inundation (Zhang et al. 2008).

The large values of computed maximum storm tides are located along the central coast of Biscayne Bay and maximum storm tides decrease toward southern and northern coasts (Fig. 8a), which is similar to the results of Zhang et al. (2008). There are considerable areas along the central coast of the Bay where maximum storm tides reached about 4.0 m above mean sea level. However, such high storm tides only inundated very limited land areas as indicated by debris lines because the Atlantic Coastal Ridges with elevations of 5–6 m and widths of 8–10 km

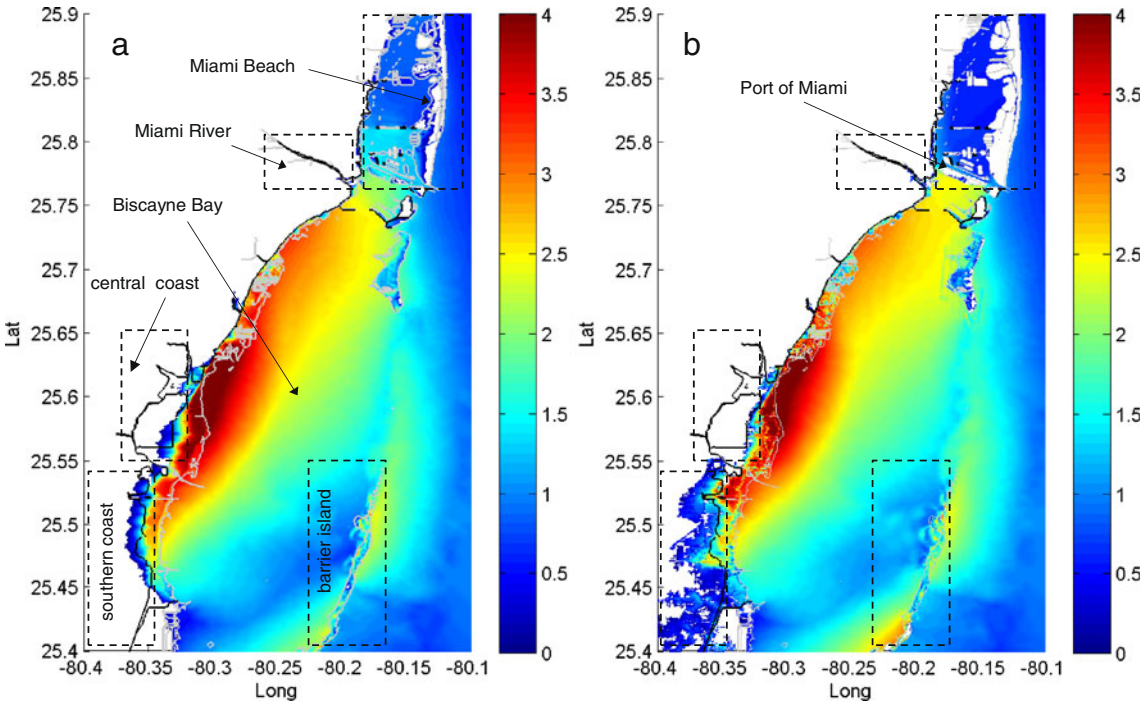

Fig. 8 Simulated maximum storm tides in the Biscayne Bay using **a** the 30 m×30 m USGS DEM (EA0) and **b** 30 m × 30 m LiDAR DEM (EA5). The *gray line* represents coast line and *black line* represents observed debris line. The maximum storm tides are referenced to mean sea level and the unit of storm tide is in meters

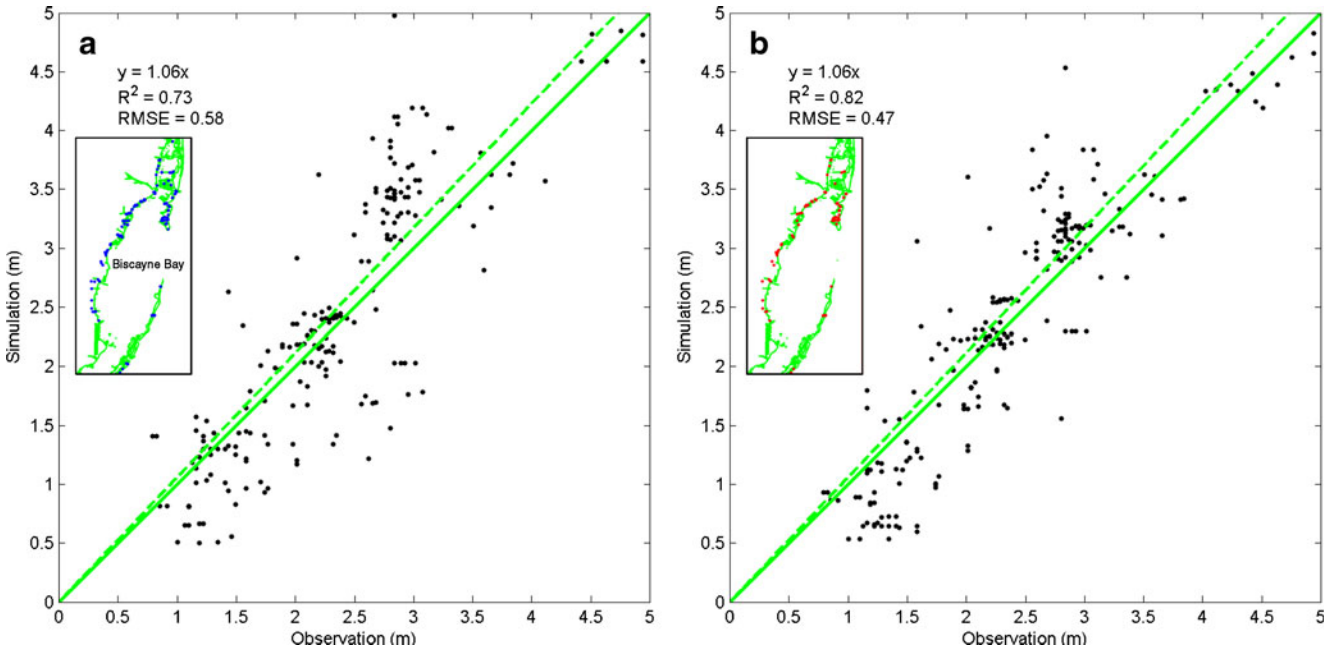

Fig. 9 Scatter plots (*black dots*), linear fits (*dashed green line*), and perfect predictions (*solid line*) for computed maximum storm tides and observed elevations of high-water marks referenced to the NGVD 29 vertical datum. **a** The simulations using USGS topographic data (EA0) and **b** the simulations using LiDAR topographic data (EA5). *Blue points* in **a** represent the locations of all high-water marks and red points in **b** represent high-water mark locations where simulated surges based on the LiDAR DEM are better than those based on USGS DEM

(Hoffmeister 1974; Zhang et al. 2008) blocked surge from further flooding the inland areas (Fig. 12a). More than 300 high-water-mark measurements, most of which are located on the coastal area of Biscayne Bay (Fig. 9a), were utilized to examine the computed maximum storm tides. The scatter plot for the elevations of high-water marks and computed maximum storm tides at corresponding locations shows that the model underestimated storm tides in the range of 0–2.5 m while it overestimated storm tides in the range of 2.5–4 m (Fig. 9a). Statistical analysis indicated that the *R*-square (R^2) value for the linear fit between observed and computed maximum storm tides and the RMS error were 0.73 and 0.58, respectively. Additionally, the extent of surge inundation computed by the model agreed with debris line (black line in Fig. 8a), indicating that the model simulated surge inundation well during Hurricane Andrew.

5 Discussion and summary

5.1 Effect of tide–surge interaction

The tide–surge interaction, which was recognized several decades ago (Proudman 1957; Rossiter 1961), not only affects the amplitude of surge height, but also alters the surge phase in the coastal zone (Davies and Lawrence 1995; Jones and Davies 1998; 2003). Although the linear superposition of surge and tide has been used for surge prediction, the nonlinear effect caused by bottom friction and momentum advection cannot be ignored in coastal regions (Davies and Jones 1992; Davies et al. 2000; Bernier and Thompson 2006; Horsburgh and Wilson 2007; Jones and Davies 2007; 2008; Brown and Wolf 2009; Nicolle et al. 2009). Jones and Davies (2007) used both the de-tidal method and harmonic analyses to remove the tide from storm tide to estimate the influence of nonlinearity on the surge level along the west coast of Britain. Horsburgh and Wilson (2007) applied the de-tidal method to investigate the tide–surge interaction in the North Sea. Similarly, we also employed the de-tidal method to examine the effect of tide–surge interaction on surge height in the Chesapeake and Biscayne Bays. First, two additional water level dataset were generated by running ELCIRC for a storm with tidal forcing only and with atmospheric forcing only, respectively. Then, storm surge η_1, calculated by subtracting tide η' (tide forcing) from water level η_0 (tide+atmospheric forcing), were compared with storm surge η_2 simulated by atmospheric forcing.

For Hurricane Isabel, two additional model experiments were named EI1 (only with tide forcing) and EI2 (only with atmospheric forcing). Figure 10a shows storm surge η_1 computed used the de-tide method (i.e., EI0 minus EI1) in the Chesapeake Bay at 22:00 UTM September 18, 2003, when the surge wave approached the lower Bay. The storm surge η_2 computed, based only on the atmospheric forcing from the EI2 simulation, is similar to η_1, but peak surge increases (Fig. 10b). A comparison of time history of storm surge at selected stations shows that the model overestimated storm surge by about 0.1 m with a phase shift of one and one-half hours without considering tide–surge interactions inside the Bay (Fig. 11). The change of the

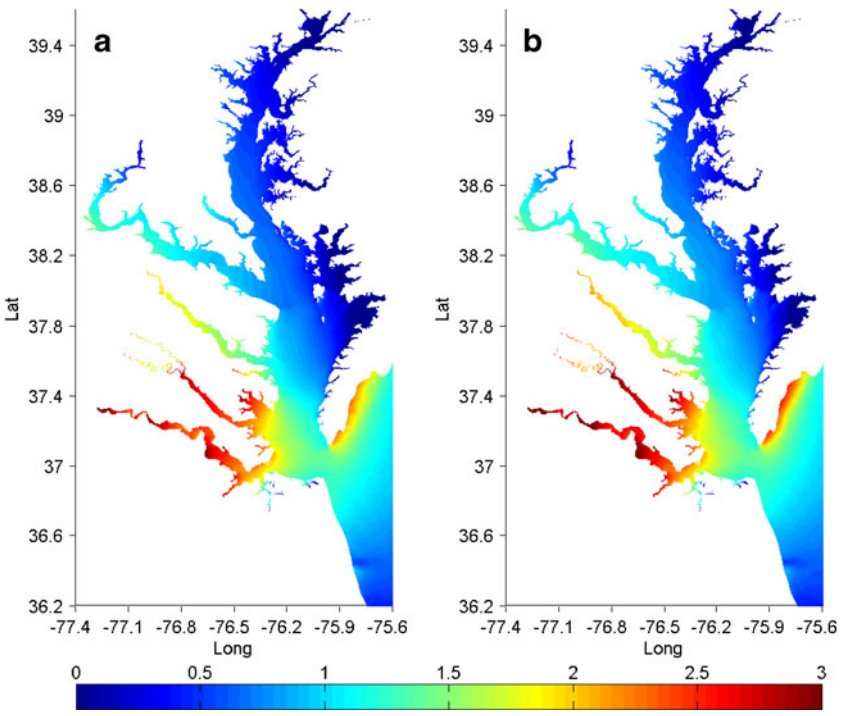

Fig. 10 **a** Storm surge from EI0–EI1 and **b** storm surge from EI2 at 22:00 UTM September 18, 2003 for Hurricane Isabel in the Chesapeake Bay

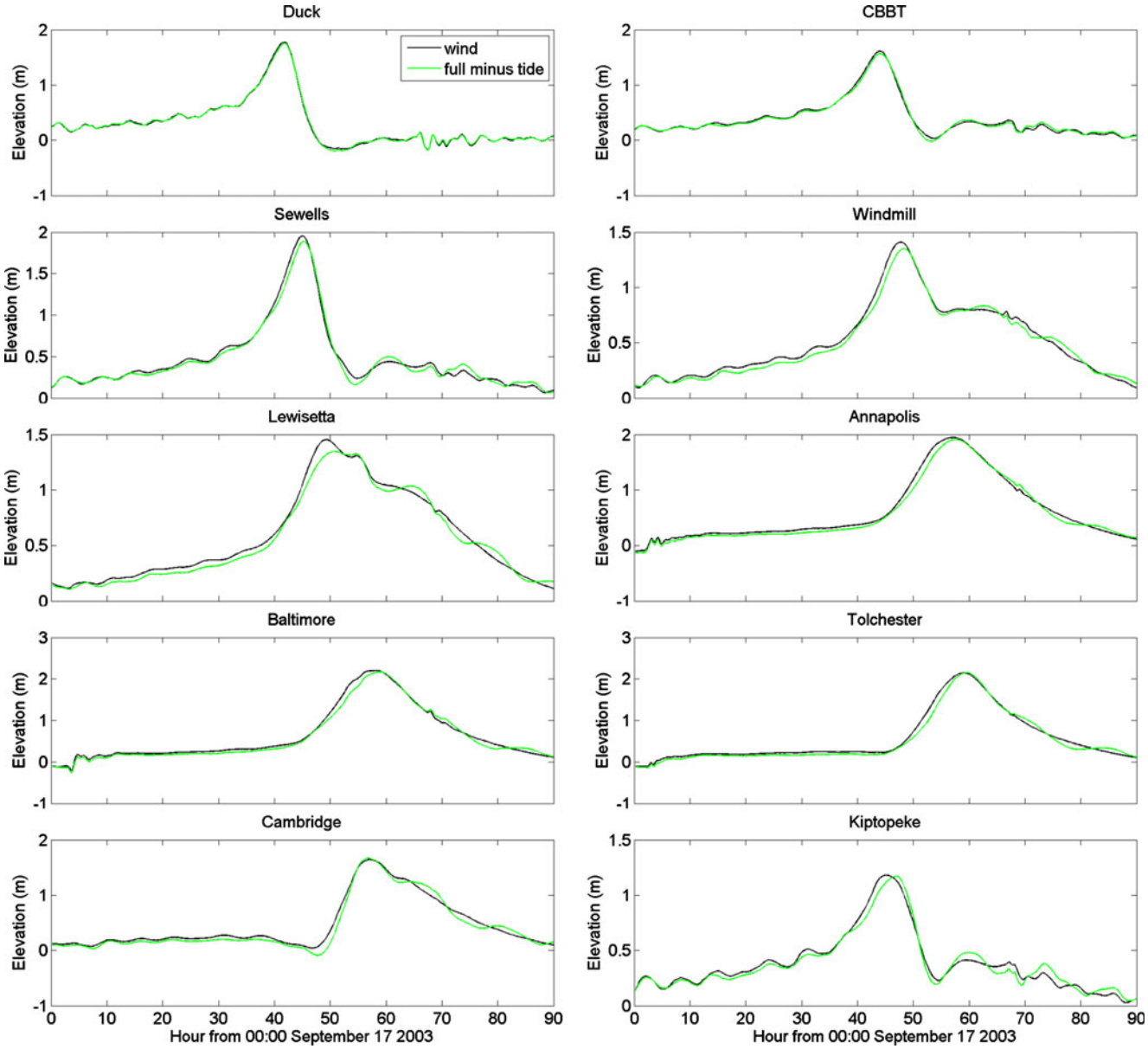

Fig. 11 Comparisons of storm surges from EI0–EI1 and storm surges from EI2 for 10 stations in the Chesapeake Bay

magnitude and timing of peak surges is consistent with previous findings (Flather 1994; Heaps 1983; Shen et al. 2006a), which are mainly caused by quadratic friction. Similar model simulations were conducted for Hurricane Andrew and results for the Biscayne Bay at 9:00 UTM August 24, 1992, are shown in Figs. 12 and 13. The differences between the two model runs are minor because of a low tidal range in the Biscayne Bay. Nevertheless, the simulation with tide and surge interaction produced more inundated area (about 12.5 km^2) than that without tide and surge interaction, indicating that the tide and surge interaction not only affects surge amplitude and phase but also influences the inundation area caused by the surge.

5.2 Effect of uncertainty in forecasting hurricane tracks

The uncertainty in forecasting hurricane track and intensity can affect surge prediction. It is easy to understand that a more intense hurricane causes higher surge at landfall areas if the track positions remain unchanged. However, the effect of the variation in track position on storm surge flooding is more complicated when compared with the change of intensity because of the intricate relationship between track and coastal bathymetry. In order to evaluate the effect of uncertainty in track forecasting on surge prediction, two simulations of storm surge were performed for two potential tracks from Hurricane Andrew (Fig. 14a).

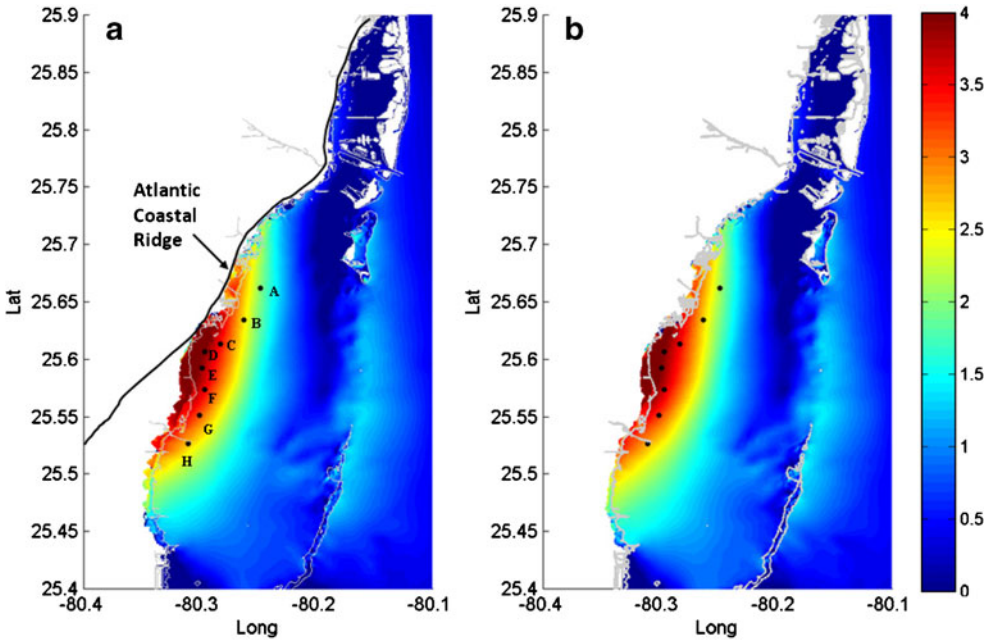

Fig. 12 **a** Storm surge from EA0–EA1 and **b** storm surge from EA2 at 9:00 UTM August 24, 1992 for Hurricane Andrew in the Biscayne Bay. The location of the Atlantic Coastal Ridge is also displayed in **a**

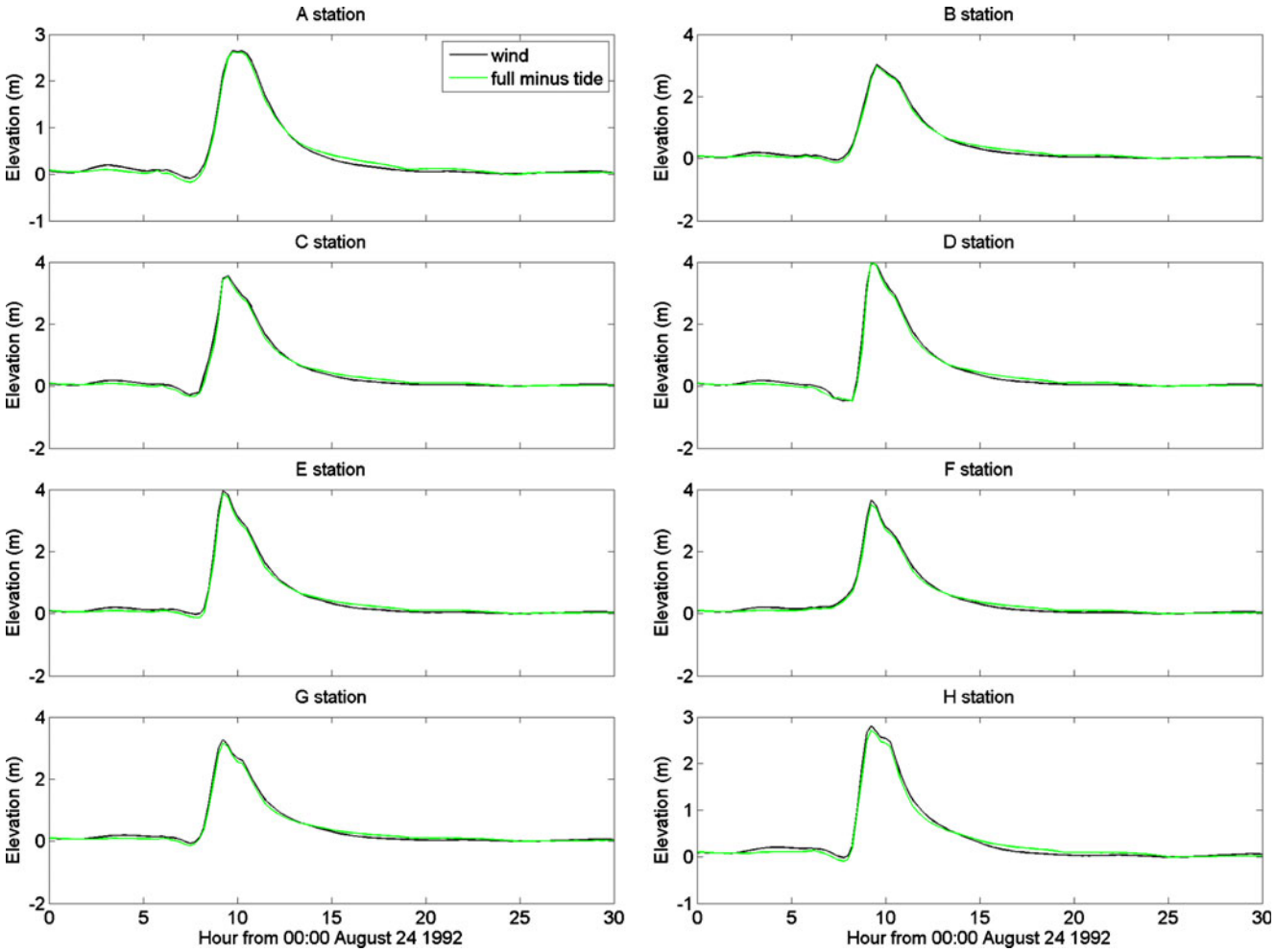

Fig. 13 Comparisons of simulated surges from EA0–EA1 and those from EA2 for eight stations in the Biscayne Bay

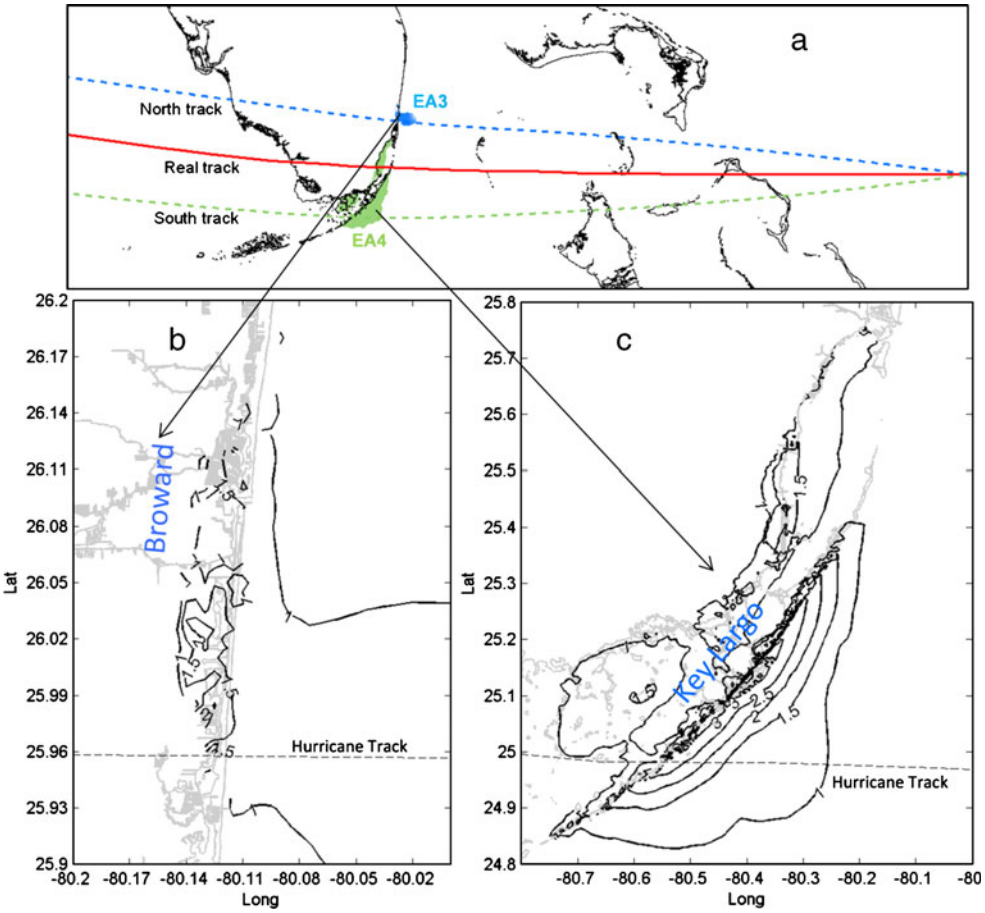

Fig. 14 **a** Hypothetical hurricane tracks (*blue and green dashed lines*) derived by rotating the track of Hurricane Andrew (*red solid line*) by 5° toward north and south from the location 24 h prior to making landfall at Miami. *Blue and green shaded areas* represent the areas with maximum storm tides higher than 1 m for the hypothetical tracks EA3 and EA4, respectively. **b** Contours of maximum storm tides along the Broward Coast generated using the EA3 simulation and **c** contour of maximum storm tides along the Key Largo Coast generated using the EA4 simulation. The maximum storm tides are referenced to mean sea level and the *gray line* represents the coastline

The two hypothetical tracks were created by deviating Andrew's track by about 5° at the location 24 h prior to landfall at Miami. Such a deviation is not uncommon in the 24-h forecast of hurricane tracks. The numerical experiments, namely EA3 and EA4, were conducted based on northerly and southerly deviated tracks with the same wind and pressure parameters for Hurricane Andrew.

The area of EA3 (Fig. 14a), with maximum storm tide higher than 1 m, is smaller than that of EA0 (Fig. 8b) and EA4 (Fig. 14a). The highest maximum storm tides in EA3 and in EA4 reach, respectively, approximately 2.0 m along the Broward Coast and about 3.0 m along the Key Largo Coast (Fig. 14b and c), much lower than that of maximum storm tides in EA0. The major reason for this difference is that there is a large-sized shallow water area next to and inside Biscayne Bay (maps.google.com), and the track of Andrew is located at the southern side of the Bay (Fig. 14). Strong onshore winds from the right side of Andrew's track pushed the shallow ocean water entering into the Bay and the water inside the Bay toward the shore, resulting in large storm tides at the northern Biscayne Bay. The size of the shallow water zone is reduced when the track is shifted to south Biscayne Bay, leading to a smaller maximum storm tide along the Key Largo Coast. The size of the shallow water zone is further reduced when the track is shifted to north Biscayne Bay, where the continental shelf is extremely narrow, causing a limit of only 2.0 m maximum storm tides at the Broward Coast, even with a Category 5 hurricane. This indicates that the magnitude and extent of surge inundation are greatly influenced by coastal and continental bathymetry and the position of a hurricane track.

5.3 Effect of accuracy of topographic elevation

In order to examine the effect of topographic data accuracy on surge modeling, an additional simulation of storm surge from Hurricane Andrew (referred to EA5 hereafter) was conducted by replacing the elevation values of land cells of R2 from the 30×30 m USGS DEM with those from the 30×30 m LiDAR DEM. Figure 15 shows that the LiDAR DEM data clearly specified more topographic features than the USGS DEM data. The hurricane parameters and boundary conditions used for the EA5 simulation were the same as those for the EA0 simulation. The simulated maximum storm tides matched well with observed values from high-water marks with R^2 and RMS error values of 0.82 and 0.47, respectively (Fig. 9b). The comparison of

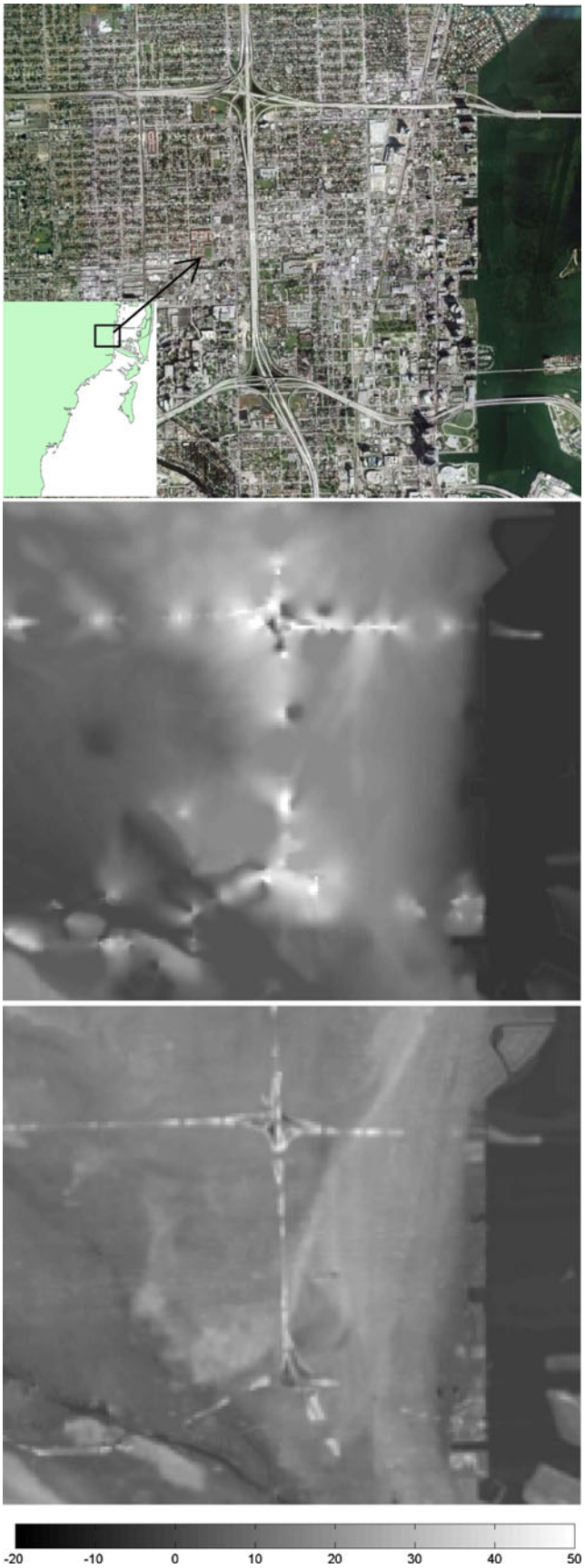

Fig. 15 The aerial photograph (*upper panel*), the 30×30 m USGS DEM (*middle panel*), and 30×30 m LiDAR DEM (*lower panel*) for the southern coastal downtown area of Miami. The DEMs are referenced to the NAVD88 vertical datum and the unit of elevation values is in meters. The *black color* represents low elevation, while the *white color* represents high elevation

linear fits for EA0 and EA5 shows that the model simulation was apparently improved by using the LiDAR DEM, implying that the accuracy of the topographic data is critical for modeling surge and inundation. The stations of better simulated maximal storm surge by the LiDAR DEM than that by the USGS DEM are marked in Fig. 9b (red points). The spatial pattern of maximum storm tides along the coast of Biscayne Bay generated using LiDAR DEM (Fig. 8b) is in general similar to that generated using USGS DEM (Fig. 8a). This is expected because the storm surges from Hurricane Andrew were blocked by the Atlantic Coastal Ridge with a width of 8–10 km and elevation of 5–6 m (Fig. 12a). The general shape of the Atlantic Coastal Ridge is well represented in both USGS and LiDAR DEMs because of its large size, although the details of ridge topography at a specific location from the two data sets can be quite different.

A close examination reveals that the maximum storm tides from the EA0 and EA5 simulations exhibit considerable differences in many areas. For example, the contours

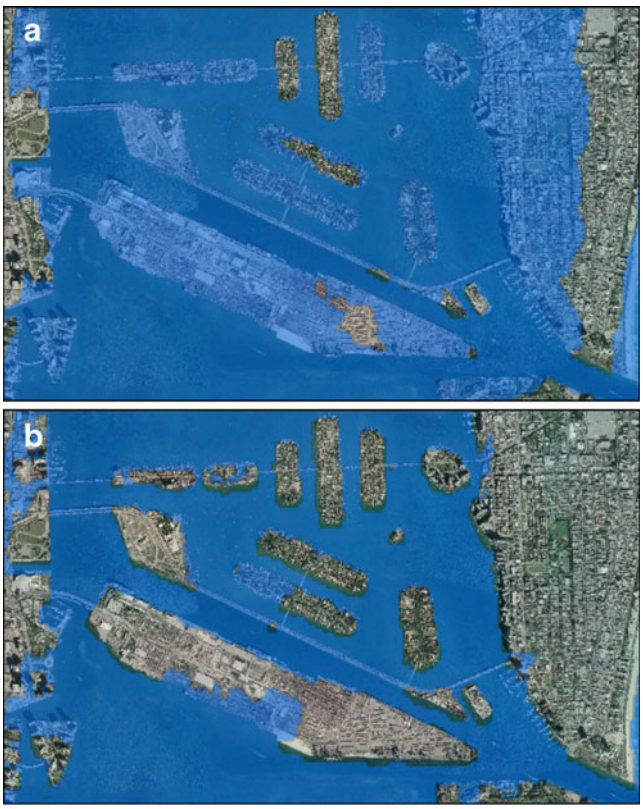

Fig. 16 Simulated maximum storm tides (*blue*) of EA0 (*upper panel*) and EA5 (*lower panel*) over the aerial photograph for the port of Miami

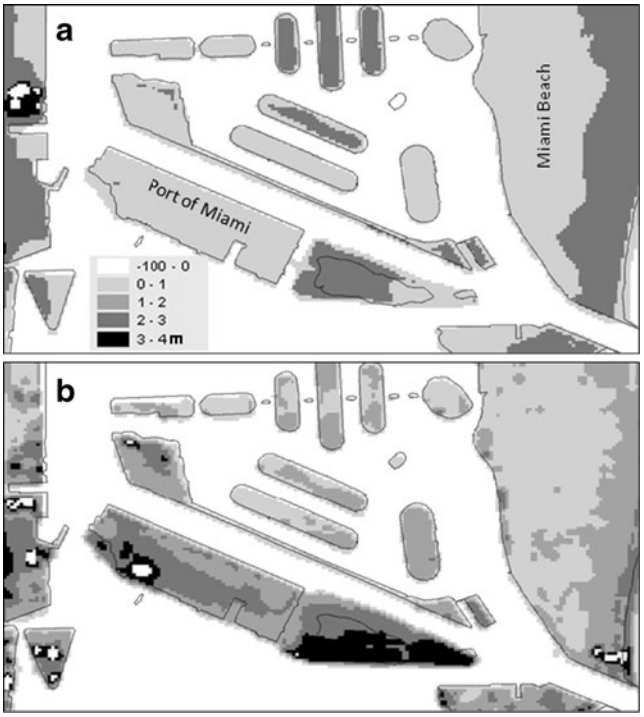

Fig. 17 Topographic data for the Port of Miami from **a** USGS and **b** LiDAR surveys

of the maximum storm surges of EA5 in the upper Bay bend more toward land than those of EA0. The maximum storm tides of EA5 are higher than those of EA0 along the bayside of the barrier island south of Biscayne Bay entrance (Fig. 8a and b). This difference is caused by using the LiDAR DEM dataset that has better representation of the topographic elevations of the narrow barrier island.

Another two examples with significant differences between the EA0 and EA5 simulations occur at Miami Beach and the southern coast of Biscayne Bay. In EA0, most of the Port of Miami and a large size of the bayside portion of the Miami Beach barrier island is inundated (Fig. 16a). In contrast, much less inundation occurs at the Port of Miami and the bayside of the Miami Beach barrier island in EA5 (Fig. 16b). Figure 17 shows that the topographic elevations at the Port of Miami and the bayside of the Miami Beach barrier island from the LiDAR DEM are significantly higher than those from the USGS DEM. Most topographic elevations from the LiDAR DEM are higher than 2 m at the Port of Miami, while the topographic elevations from the USGS DEM are lower than 2 m. The USGS DEM also lacks details of the local topography due to a large vertical resolution (~1.5 m). At the southern coast of Biscayne Bay, however, EA5 over-predicted surge inundation compared with EA0 (Fig. 8), which is probably due to ignoring the effects of vegetation. There are mangrove zones with widths of 1–4 km and tree heights of 1–20 m along the southern coast of Biscayne Bay (Fig. 18a). It is well-known that the mangrove forest can attenuate surge flooding (Das and Vincent 2009). Therefore, it can be concluded by comparison of the simulations based on the USGS and LiDAR DEMs that topographic data with high vertical accuracy are critical to predict the inundation induced by storm surge at the street level.

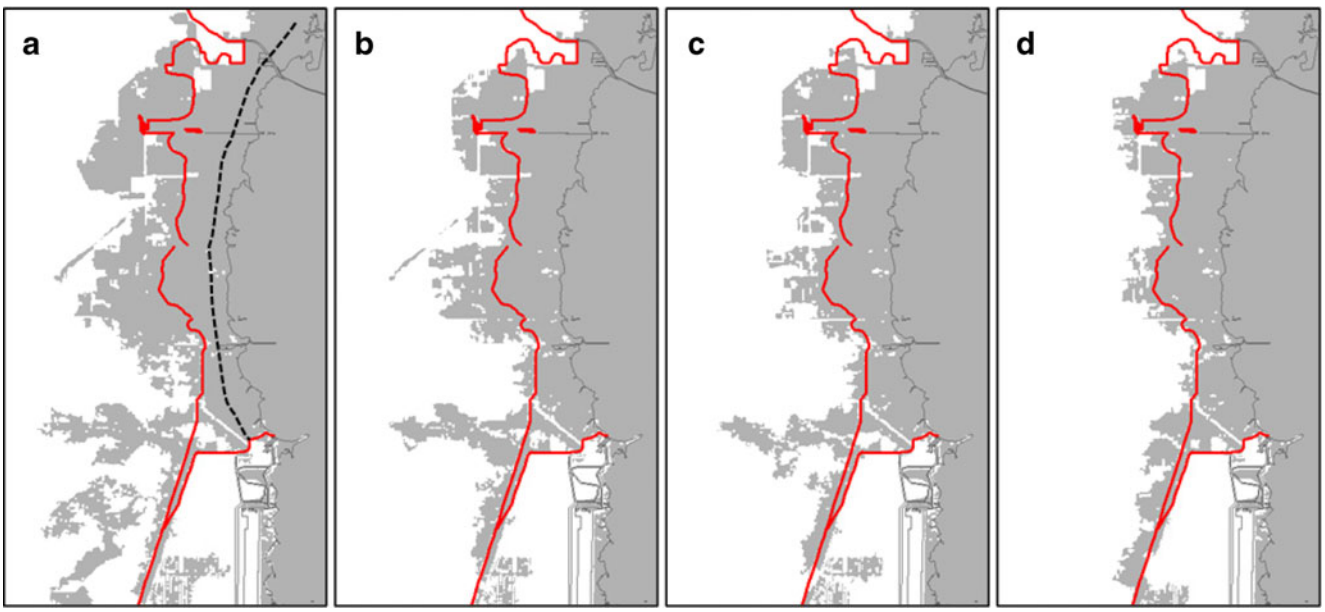

Fig. 18 The storm surge inundation extents (*shaded areas*) that include the land area with surge depth greater than 0.3 m, observed debris lines (*red line*), and the landward boundary of mangrove forest (*black dashed line*) at the southern coast of Biscayne Bay. The inundation extents were generated suing the simulations of **a** EA5 with n=0.02, **b** EA6 with n=0.05, **c** EA7 with n=0.1, and **d** EA8 with n=0.15. *Gray line* represents coastline

Table 4 Idealized manning's friction coefficient based on Chow (1959) and Mattocks and Forbes (2008)

Landuse type	Manning's friction coefficient
Open water/sand	0.02
Scattered brush/shrub/scrub	0.05
Woody wetlands/mixed forest	0.10
Dense woods	0.15

5.4 Effect of mangrove forest

The overestimation of maximum storm tides at the Southern Coast of Biscayne Bay by the EA5 simulation suggests that land cover types may have significant effects on inundation computation. The previous studies show that bed forms and different types of the bottom can affect the flow and surge (Davies and Lawrence 1995; Kagan et al. 2005, Nicolle and Karpytchev 2007). To improve the accuracy of storm surge simulation, the effect of land covers needs to be incorporated into the model. A common approach is to use different bottom roughness to represent the effects of various land cover features. Many studies suggested that overland flooding can be reduced by the vegetation through complex processes (Nepf 1999; Green 2005; Loder et al. 2009). Mattocks and Forbes (2008) emphasized the effect of changes in bottom friction caused by different land uses on storm surge forecasting based on Manning's formula.

To examine the role of the bottom friction on the storm surge prediction, we conducted three sensitivity experiments (EA6–EA8) with respect to three different Manning's friction coefficients of 0.05, 0.1, and 0.15 based on friction values shown in Table 4 for the mangrove area along the Southern Coast of Biscayne Bay. Figure 18 shows overland surge inundation extents with surge depth >0.3 m from Hurricane Andrew at the southern coast of Biscayne Bay for simulations of EA0, EA6, EA7, and EA8. The inundation extent lines gradually approach the observed debris line with an increase of Manning's friction coefficient. Although the exact increment of Manning's friction coefficient due to the mangrove forest is unknown owing to the lack of field measurements, this experiment suggests that change of friction coefficients due to vegetation can significantly influence the local inundation pattern of storm surge.

5.5 Effect of rivers and canals

One of the shortcomings of LiDAR topographic data is that they only provide accurate elevation values for land, but do not provide reliable measurements for water depths of rivers and canals because the topographic LiDAR systems use near infrared laser that cannot penetrate water. The depths and boundaries of rivers and canals are often

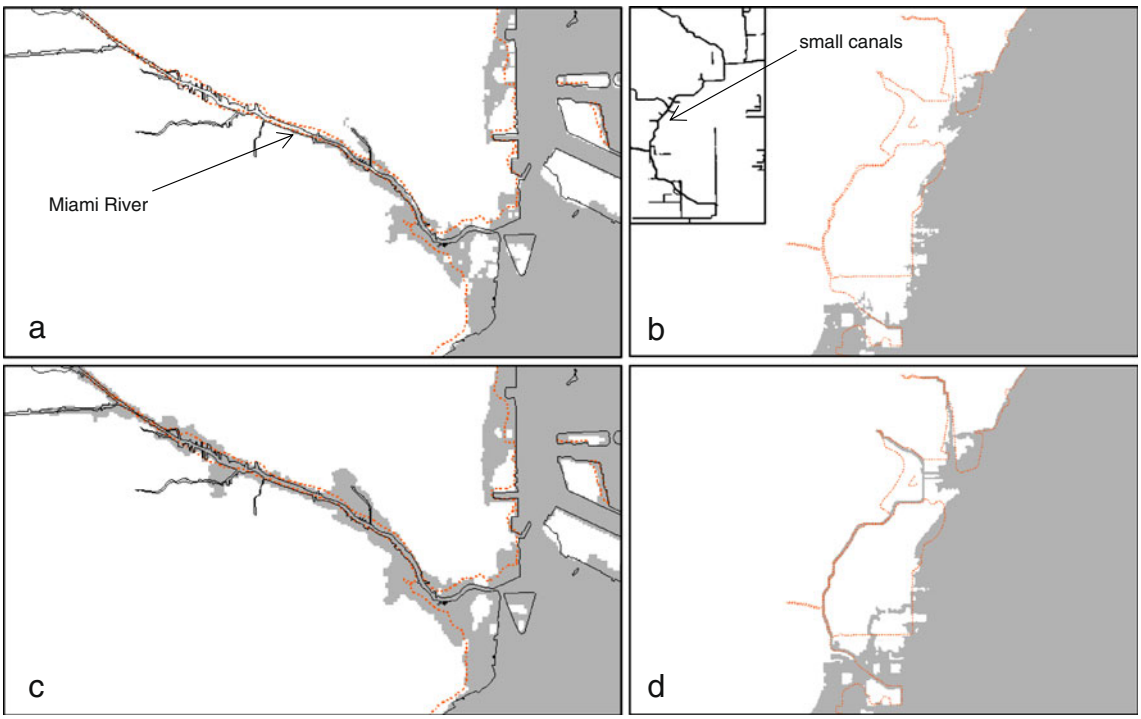

Fig. 19 The inundation extents for **a** the Miami River area and **b** the central coast of Biscayne Bay from the EA5 simulation, and the inundation extents **c** for the Miami River area and **d** the central coast of Biscayne Bay from the EA9 simulation. *Dashed red line* represents the observed debris line

distorted in the DEM that is generated by interpolating the LiDAR point measurements for the terrain, leading to incorrect simulations of surge flooding around these hydrologic features. For example, the model results from EA5 show that the surge inundation in the Miami River and the canals at the central coast of Biscayne Bay are underestimated (Fig. 19a and b). Thus, corrections of water depths for the river and canals need to be performed before conducting the simulation.

The shapes of rivers and canals in the Miami area can be extracted from recent aerial photographs, but there is no detailed recent survey for water depths. Therefore, a constant depth value of 0.5 m was used for the rivers and canals in the Miami area. A new simulation (EA9) with updated shapes and water depths for rivers and canals was conducted to examine the effect of the linear hydrologic features on surge simulation. The model results show that surge inundation surrounding the Miami River, especially the upstream area, and the canal areas at the central coast of Biscayne Bay was improved (Fig. 19c and d) compared with simulation EA5. This indicates that it is important to resolve the linear hydrologic features such as rivers and canals in simulating surge flooding in urban areas.

5.6 Summary

The successful application of multi-scale unstructured models to simulate the coastal flooding caused by Hurricanes Andrew and Isabel in South Florida and the Chesapeake Bay demonstrates the potential of the multi-scale approach in predicting storm surges along the entire U.S. East and Gulf Coasts. Since considerable uncertainty is associated with the hurricane track forecast and one hurricane can make landfall more than once along the U.S. East and Gulf Coasts, as in the case of Hurricane Andrew, it is not feasible to generate one fine-resolution model grid to cover all areas possibly impacted by the hurricane. Moreover, future hurricanes can make landfalls at any location along the U.S. East and Gulf Coasts. Our approach to predict the overland flooding caused by hurricanes along the U.S. East and Gulf Coasts is to generate multiple multi-scale model grids based on shoreline configuration. First, a coarse-resolution large domain grid was created to cover the Northern Atlantic Ocean, the Gulf of Mexico, and the Caribbean Sea over which the ELCIRC model was well calibrated for the tides along the U.S. East and Gulf Coasts. Then, the entire coast was divided into 12 overlapping basins with fine-resolution grids, as shown in Fig. 1a. Each basin has a fine-resolution grid for simulating overland flooding in which the fine grids gradually reduce to the same coarse-resolution grid in the offshore region so that each model grid has an identical spatial coverage for a consistent boundary condition.

With the use of a multi-scale modeling approach, grid cells with resolution close to "street level" can be generated for populated urban areas such as Miami where the surge inundations are of great concern, while the remote effects can be fully simulated without difficulty in specifying the open boundary condition. The advantage of the multiple multi-scale grid approach for storm surge prediction is that it can handle more efficiently the uncertainty in the forecast of the hurricane track by executing models over the multiple domains along the potential passages of the hurricane simultaneously. For a test grid with a total of 609,152 nodes for the R2 basin, a 5-day simulation requires 3 h of computation using a single 2.93 Hz CPU. Therefore, hundreds of simulations can be conducted within a short time period for predicting storm surges caused by the cone of hurricane tracks using the computer cluster. Another advantage of the multi-scale grid approach is that it is more convenient to update basin grids when new topographic and bathymetric data for the basin become available.

In summary, the multi-scale models with local refinement at different regions are implemented to perform simulation of storm surge flooding along the U.S. East and Gulf Coasts. The application of the multi-scale approach to Hurricanes Andrew and Isabel indicates that the ELCIRC model simulated the tides along the U.S. East and Gulf Coasts and hurricane-induced storm surges successfully. A series of experiments were also conducted to examine the effects of the accuracy of topographic data, uncertainty in hurricane track forecast, and bottom drag coefficients on the computation of coastal flooding. The results indicate that accurate topographic data such as those generated from LiDAR measurements are essential for modeling the surge flooding at the street level. Changes in bottom drag coefficients due to different land cover types also play an important role in the simulation of surge inundation and deserve further investigation through field observations. Appropriate representation of rivers and canals in the model is critical to model surge inundation surrounding these linear hydrologic features.

Acknowledgments This project was supported by the Florida Hurricane Alliance Research Project sponsored by NOAA. We thank Mr. Mac Sisson of VIMS and three anonymous reviewers for reviewing the manuscript and providing valuable comments. We also thank Dr. Joseph Zhang of the Center for Coastal Margin Observation & Prediction at Oregon Health and Science University for supporting our numerical modeling.

References

Aldridge JN, Davies AM (1993) A high resolution three-dimensional hydrographic tidal model of the eastern Irish Sea. J Phys Oceanogr 23:207–224

Bernier NB, Thompson KR (2006) Predicting the frequency of storm surges and extreme sea levels in the northwest Atlantic. Journal of Geophysical Research, 111, C100009, doi:10.1029/2005JC003168.

Blain CA, Westerink JJ, Luettich RA (1994) The influence of domain size on the response characteristics of a hurricane storm surge model. J Geophys Res 99(C9):18,467–18,497

Blain CA, Westerink JJ, Luettich RA (1998) Grid convergence studies for the prediction of hurricane storm surge. Int J Numer Methods Fluids 26:369–401

Boon J (2004) Secrets of the Tide. Horwood Publishing Limited, Chichester (212p)

Brown J, Wolf J (2009) Coupled wave and surge modeling for the eastern Irish Sea and implications for model wind-stress. Cont Shelf Res 29:1,329–1,342

Casulli V (1999) A semi-implicit numerical method for non-hydrostatic free surface flows on unstructured grid. Proceeding of International Workshop on Numerical Modeling of Hydrodynamic System (Zaragoza, Spain): 175–193.

Casulli V, Zanolli P (1998) A three-dimensional semi-implicit algorithm for environmental flows on unstructured grids. Proceedings of Conference on Numerical Methods for Fluid Dynamics, University of Oxford.

Cho K (2009) A numerical modeling study on barotropic and baroclinic responses of the Chesapeake Bay to hurricane events. PhD Dissertation, Virginia Institute of Marine Science, College of William and Mary.

Chow VT (1959) Open channel hydraulics. McGraw-Hill Book Co, NY, 680pp

Das S, Vincent JR (2009) Mangroves protected villages and reduced death toll during Indian super cyclone, Proceedings of the National Academy of Sciences of the United States of America 106 (18): 7357–7360.

Davies AM, Hall P (2002) Numerical problems associated with coupling models in shelf edge regions. Appl Math Model 26:807–831

Davies AM, Jones JE (1992) A three-dimensional wind driven circulation model of the Celtic and Irish Seas. Cont Shelf Res 12:159–188

Davies AM, Jones JE (1996) Sensitivity of tidal bed stress distributions, near bed currents, overtides and tidal residuals to frictional effects in the eastern Irish Sea. J Phys Oceanogr 26:2,553–2,575

Davies AM, Lawrence J (1995) Modelling the effect of wave-current interaction on the three-dimensional wind driven circulation of the eastern Irish Sea. J Phys Oceanogr 25:29–45

Davies AM, Kwong SC, Flather RA (2000) On determining the role of wind wave turbulence and grid resolution upon computed storm driven currents. Cont Shelf Res 20:1,825–1,888

Deleersnijder E, Lermusiaux P (2008) Multi-scale modeling: nested-grid and unstructured-mesh approaches. Ocean Dyn 58:335–336

Flather RA (1994) A storm-surge prediction model for the northern Bay of Bengal with application to cyclone disaster in April 1991. J Phys Oceanogr 24:172–190

Flather RA, Proctor R (1983) Prediction of North Sea storm surge using numerical models: recent developments in the U.K. In: Sundermann J, Lenz W (eds) North Sea Dynamics. Berlin, Springer, pp 299–317

Flather RA, Proctor R, Wolp J (1991) Oceanographic forecast models. In: Farmer DJ, Rycroft MJ (eds) Computer Modelling in the Environmental Science. Clarendon, Oxford, pp 15–30

Garratt J (1977) Review of drag coefficients over oceans and continents. Mon Weather Rev 105:915–929

Glahn B, Taylor A, Kurkowski N, Shaffer W (2009) The role of the SLOSH model in national weather service storm surge forecasting. Natl Weather Dig 33(1):3–14

Gong W, Shen J, Reay WG (2007) The hydrodynamic response of the York River estuary to Tropical Cyclone Isabel, 2003. Estuar Coast Shelf Sci 73:695–710

Gong W, Shen J, Cho KH, Wang H (2009) A numerical model study of barotropic subtidal water exchange between estuary and subestuaries (tributaries) in the Chesapeake Bay during nor'easter events. Ocean Model 26(3–4):170–189

Green J (2005) Modeling flow resistance in vegetated streams: review and development of new theory. Hydrol Process 19:1245–1259

Ham D, Pain C, Hanert E, Pietrzak J, Schröter J (2009) Special Issue: The sixth international workshop on unstructured mesh numerical modelling of coastal, shelf and ocean flows. Imperial College London, September 19-21, 2007. Ocean Model 28(1–3):19–21

Heaps NS (1983) Storm surges, 1967–1982. Geophys J R Astron Soc 74:331–376

Hoffmeister E (1974) Land from the sea. Miami. Florida, University of Miami Press, p 143

Horsburgh KJ, Wilson C (2007) Tide–surge interaction and its role in the distribution of surge residuals in the North Sea. J Geophys Res 112:C08003. doi:10.1029/2006JC004033

Houston S, Shaffer W, Powell M, Chen J (1999) Comparisons of HRD and SLOSH surface wind fields in hurricanes: implications for storm surge and wave modeling. Weather Forecast 14:671–686

Jelesnianski CP (1967) Numerical computations of storm surge with bottom stress. Mon Weather Rev 95:740–756

Jelesnianski CP, Chen J, Shaffer W (1992) SLOSH: Sea, Lake, and Overland Surges from Hurricane. National Weather Service, Silver Springs, MD, (71p). (check fontsize)

Jones JE, Davies AM (1998) Storm surge computations for the Irish Sea using a three-dimensional numerical model including wave–current interaction. Cont Shelf Res 18:201–251

Jones JE, Davies AM (2003) Processes influencing storm-induced currents in the Irish Sea. J Phys Oceanogr 33:88–104

Jones JE, Davies AM (2007) Influence of non-linear effects upon surge elevations. Ocean Dyn 57:401–416

Jones JE, Davies AM (2008) On the modification of tides in shallow water regions by wind effects. Journal of Geophysical Research, 113, doi:10.1029/2007JC004310.

Kagan BA, Alvarez O, Izquierdo A (2005) Weak wind-wave/tide interaction over fixed and moveable bottoms: a formulation and some preliminary results. Cont Shelf Res 25:753–773

Li M, Zhong L, Boicourt WC, Zhang S, Zhang DL (2006) Hurricane induced storm surges, currents and destratification in a semi-enclosed bay. Geophys Res Lett 33:L02604. doi:10.1029/2005GL024992

Loder N, Irish J, Cialone M, Wamsley T (2009) Sensitivity of hurricane surge to morphological parameters of coastal wetlands. Estuar Coast Shelf Sci 84:625–636

Luettich RA Jr, Westerink J, Scheffner NW (1992) ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves, Coasts and Estuaries, Report 1: Theory and Methodology of ADCIRC-2DDI and ADCIRC-3DL", DRP Technical Report DRP-92-6, Department of the Army, US Army Corps of Engineers, Waterways Experiment Station, Vicksburg, MS., November, p 137

Mathew J, Mahadevan R, Bharatkumar B, Subramanian V (1996) Numerical simulation of open coast surges. Part I: experiments on offshore boundary conditions. J Coast Res 12:112–122

Mattocks C, Forbes C (2008) A real-time, event-triggered storm surge forecasting system for the state of North Carolina. Ocean Model 25:95–119

Mukai A, Westerink J, Luettich R (2002) Guidelines for using the East Coast 2001 Database of tidal constitutes within western North Atlantic Ocean, Gulf of Mexico and Caribbean. US Army Corps of Engineers. ERDC/CHL CHETN-IV-40, p 20

Myers V, Malkin W (1961) Some properties of hurricane wind fields as deduced from trajectories. National hurricane research project report, no. 49, NOAA, US Department of Commerce, p 43

Nepf H (1999) Drag, turbulence, and diffusion in flow through emergent vegetation. Water Research Resource 35(2):479–489

Nicolle A, Karpytchev M (2007) Evidence for spatially variable friction from tidal amplification and asymmetry in a shallow semi-diurnal embayment: the Pertuis Breton Bay of Biscay, France. Cont Shelf Res 27:2345–2356

Nicolle A, Karpytchev M, Benoit M (2009) Amplification of the storm surges in shallow waters of the Pertuis Charentais (Bay of Biscay, France). Ocean Dyn 59:921–935

Peng M, Xie L, Pietrafesa L (2006) A numerical study on hurricane-induced storm surge and inundation in Charleston Harbor, South Carolina. J Geophys Res 111:C08017. doi:10.1029/2004JC002755

Pietrzak J, Deleersnijder E, Schröter J (2005) Special issue: the second international workshop on unstructured mesh numerical modelling of coastal, shelf and ocean flows Delft, The Netherlands, September 23–25, 2003. Ocean Model 10(1–2):1–3

Pietrzak J, Iskandarani M, Schröter J, Lyard F (2006) Special issue: the third international workshop on unstructured mesh numerical modelling of coastal, shelf and ocean flows Toulouse, France, September 20–22, 2004. Ocean Model 15(1–2):1–2

Proudman J (1957) Oscillations of tide and surge in an estuary of finite length. J Fluid Mech 2:371–382

Rego J, Li C (2010) Nonlinear terms in storm surge prediction: Effect of tide and shelf geometry with case study from Hurricane Rita. Journal of Geophysical Research doi:10.1029/2009JC005285, in press

Rossiter JR (1961) Interaction between tide and surge in the Thames. Geophys J R Astron Soc 6:29–53

Shen J, Gong W (2009) Influence of model domain size, wind direction and Ekman transport in storm surge development inside the Chesapeake Bay: a case study if extratropical cyclone Ernesto, 2006. J Mar Syst 75:198–215

Shen J, Gong W, Wang H (2006a) Water level response to 1999 Hurricane Floyd in the Chesapeake Bay. Cont Shelf Res 26:2484–2502

Shen J, Zhang K, Xiao C, Gong W (2006b) Improved prediction of storm surge inundation with a high-resolution unstructured grid mode. J Coast Res 22(6):1309–1319

Shen J, Wang H, Sisson M, Gong W (2006c) Storm tide simulation in the Chesapeake Bay using an unstructured grid model. Estuar Coast Shelf Sci 68:1–16

Shen T (2009) Development of a Storm Surge Model Using a High-Resolution Unstructured Grid over a Large Domain. MS Thesis, The College of William and Mary. Gloucester Point VA 23062. p 279

Tsimplis MN, Proctor R, Flather RA (1995) A two-dimensional tidal model for the Mediterranean Sea. J Geophys Res 100 (C8):16,223–16,239

Weisburg R, Zheng L (2008) Hurricane storm surge simulations: comparisons three-dimensional with two-dimensional formulations based on an Ivan-like storm surge over the Tampa Bay, Florida region. Journal of Geophysical Research 113 (C12001): doi:10.1029/2008JC005115.

Westerink J, Luettich R, Muccino J (1994) Modeling tides in the western North Atlantic using unstructured graded grids. Tellus 46A:178–199

Whitman D, Zhang K, Leatherman SP, and Robertson W (2003) An airborne laser topographic mapping application to hurricane storm surge hazard. In G. Heiken, R. Fakundiny, and J. Sutter (editors), Earth Science in the Cities p 363-376.

Xie L, Liu H, Peng M (2008) The effect of wave-current interactions on the storm surge and inundation in Charleston Harbor during Hurricane Hugo 1989. Ocean Model 20:252–269

Zhang K (2010) Analysis of non-linear inundation from sea-level rise using LIDAR data: a case study for South Florida. Climatic Change, in press.

Zhang K, Xiao C, Shen J (2008) Comparison of the CEST and SLOSH models for storm surge flooding. J Coast Res 24(2):489–499

Zhang Y, Baptista A, Myers E (2004) A cross-scale model for 3D baroclinic circulation in estuary-plume-shelf systems: I. Formulation and skill assessment. Cont Shelf Res 24:2187–2214

Ocean Dynamics (2010) 60:1621–1637
DOI 10.1007/s10236-010-0348-5

Multi-scale modeling of Puget Sound using an unstructured-grid coastal ocean model: from tide flats to estuaries and coastal waters

Zhaoqing Yang · Tarang Khangaonkar

Received: 30 April 2010 / Accepted: 5 October 2010 / Published online: 19 November 2010

Abstract Water circulation in Puget Sound, a large complex estuary system in the Pacific Northwest coastal ocean of the United States, is governed by multiple spatially and temporally varying forcings from tides, atmosphere (wind, heating/cooling, precipitation/evaporation, pressure), and river inflows. In addition, the hydrodynamic response is affected strongly by geomorphic features, such as fjord-like bathymetry and complex shoreline features, resulting in many distinguishing characteristics in its main and sub-basins. To better understand the details of circulation features in Puget Sound and to assist with proposed nearshore restoration actions for improving water quality and the ecological health of Puget Sound, a high-resolution (around 50 m in estuaries and tide flats) hydrodynamic model for the entire Puget Sound was needed. Here, a three-dimensional circulation model of Puget Sound using an unstructured-grid finite volume coastal ocean model is presented. The model was constructed with sufficient resolution in the nearshore region to address the complex coastline, multi-tidal channels, and tide flats. Model open boundaries were extended to the entrance of the Strait of Juan de Fuca and the northern end of the Strait of Georgia to account for the influences of ocean water intrusion from the Strait of Juan de Fuca and the Fraser River plume from the Strait of Georgia, respectively. Comparisons of model results, observed data, and associated error statistics for tidal elevation, velocity, temperature, and salinity indicate that the model is capable of simulating the general circulation patterns on the scale of a large estuarine system as well as detailed hydrodynamics in the nearshore tide flats. Tidal characteristics, temperature/salinity stratification, mean circulation, and river plumes in estuaries with tide flats are discussed.

Keywords Puget Sound · Numerical model · Tides · Estuarine circulation · Stratification · Tide flat

Responsible Editor: Pierre Lermusiaux

Z. Yang (✉) · T. Khangaonkar
Pacific Northwest National Laboratory,
Battelle Seattle Research Center,
1100 Dexter Avenue North, Suite 400,
Seattle, WA 98109, USA
e-mail: zhaoqing.yang@pnl.gov

1 Introduction

Puget Sound is a complex fjordal estuarine system located along the Pacific Northwest coast of the United States (Fig. 1). It is one of the most pristine estuarine systems in the United States and provides suitable habitats for salmon and marine wildlife. The circulation in Puget Sound is dominated by tides, but is also influenced by atmospheric forcing and river discharges. It is altered by geometric effects, such as shallow-water tide flats, complex shorelines, and multiple sills. Because of the interactions and balances of all the forcing mechanisms and geometric effects, circulation patterns in various sub-basins of Puget Sound exhibit distinct characteristics (Cannon 1983; Cokelet et al. 1990). The sub-basins are interconnected, and experiences from on-ground projects and previous assessments indicate that there is potential for site-specific restoration actions to have basin-wide effects or impacts in Puget Sound (Yang et al. 2010a, b). Many restoration actions aim at restoring agriculture farmlands to fully functional tidal wetlands and improving juvenile migration pathways through dike breaches and/or removal. A numerical model therefore becomes necessary to not only describe the overall circulation patterns of Puget Sound,

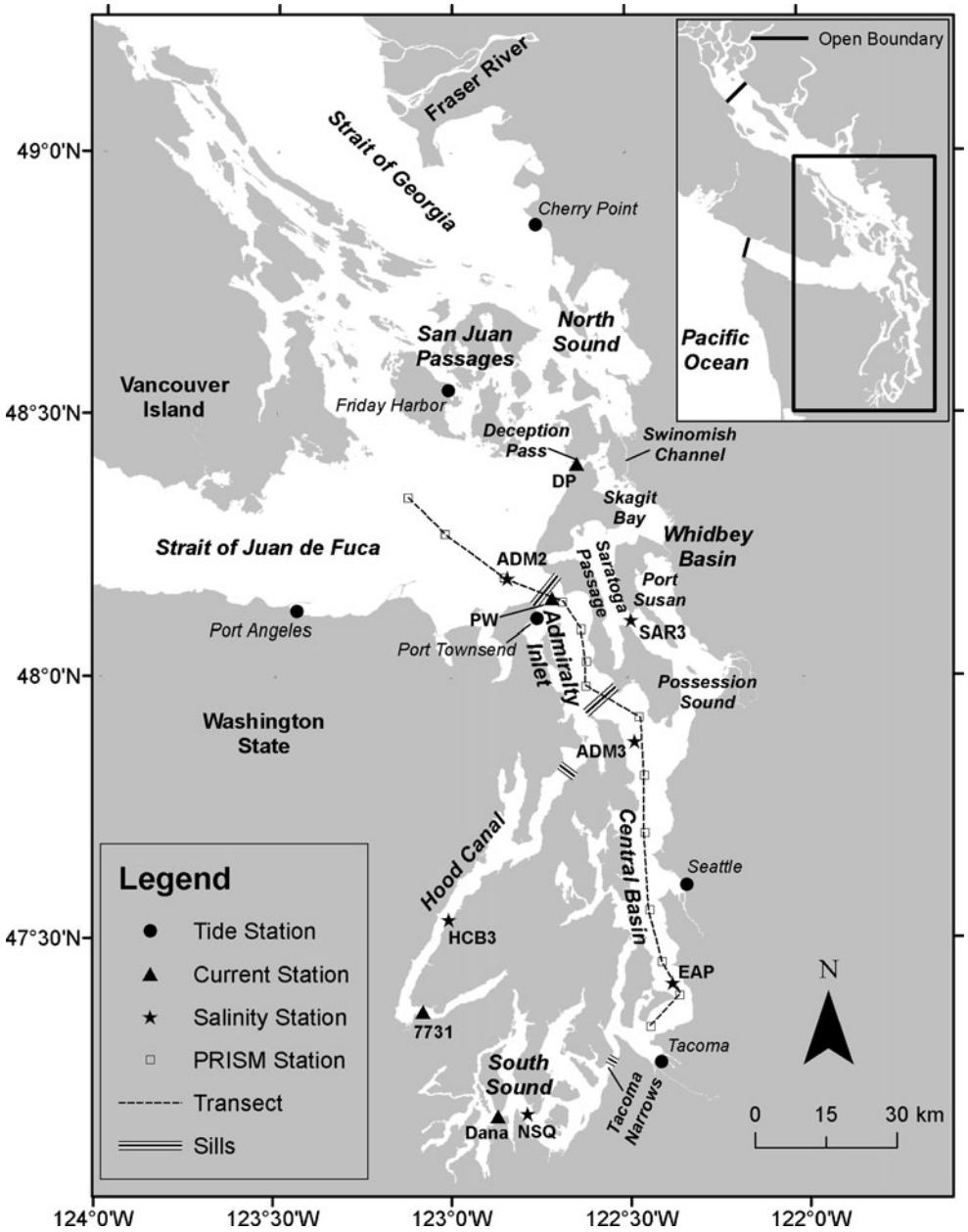

Fig. 1 Study domain—Puget Sound and the Pacific Northwest Straits

but also to better understand detailed circulation responses in the individual estuaries and sub-basins.

There have been many descriptive studies addressing circulation in Puget Sound and its sub-basins based on the analysis of observed data (Cannon and Ebbesmeyer 1978; Cannon 1983; Ebbesmeyer and Barnes 1980; Bretschneider et al. 1985; Cokelet et al. 1990; Moore et al. 2008a, b). Most of these data were collected in the 1970s, until a recent effort was initiated at gathering detailed oceanographic information in the main basin of Puget Sound in connection with an outfall relocation project (Ebbesmeyer et al. 2002), for example, Barnes and Ebbesmeyer (1978) described general circulation patterns and mixing processes in the Main Basin and the Whidbey Basin of Puget Sound, qualitatively based on historical observations. Ebbesmeyer and Coomes (1989) showed that low frequency (decadal) fluctuations in the period of 10 to 20 years were closely linked to the atmospheric low-pressure systems over the North Pacific Ocean and Puget Sound. Ebbesmeyer et al. (2002) characterized the mean current patterns in Elliott Bay located off Seattle, Washington, using historical conductivity, temperature, and depth observations as well as contemporary atmospheric data, along with observations of drift cards. Through these studies and assessments, a fairly robust understanding of the general circulation and behavior of the Puget Sound fjordal waterbody, bounded by shallow sills and with multiple interconnected basins with freshwater inflow, has emerged. Moore et al. (2008a)

conducted a comprehensive analysis of temporal and spatial distribution patterns of oceanographic properties in Puget Sound using long-term observed data. A separate study by Moore et al. (2008b) indicated that seasonal and interannual variability of Puget Sound oceanographic properties are closely related to local and large-scale climate forcings.

However, in contrast, numerical modeling studies of the circulation in Puget Sound are somewhat limited. Until this effort, a comprehensive hydrodynamic modeling effort covering the entire domain of Puget Sound and the Straits of Juan de Fuca and Georgia with a resolution to the sub-basin level had not been attempted. Many previous modeling studies focused on large-scale circulation patterns in the Pacific Northwest coastal waters and the Straits of Juan de Fuca and Georgia, for example, Foreman et al. (1993), Cummins and Oey (1997), Crawford et al. (1999), and Cummins et al. (2000) conducted modeling studies of circulation around Vancouver Island, British Columbia. Masson and Cummins (2000) examined the influence of fortnightly modulated mixing on the estuarine circulation in the Strait of Juan de Fuca. A three-dimensional (3-D) baroclinic model was implemented by Masson and Cummins (2004) to simulate the seasonal variation of the Straits of Juan de Fuca and Georgia using the Princeton Ocean Model (POM). A laterally averaged, vertical, 2-D model was used to study density intrusion into Puget Sound (Lavelle et al. 1991). Logutov and Lermusiaux (2008) simulated barotropic tides in the Strait of Juan de Fuca, Strait of Georgia, and Puget Sound using an inverse method in a depth-averaged, spectral, shallow-water, tidal model.

In recent years, there has been a resurgence of effort to develop numerical models allowing quantitative assessments. A box model of Puget Sound developed by Babson et al. (2006) was used to study the long-term variability in the circulation of Puget Sound. As part of the University of Washington's Puget Sound Regional Synthesis Model (PRISM) program, a 3-D structured-grid model was developed with POM to study the variability of currents with a focus on the complexities of the triple junction site at the confluence of Admiralty Inlet, Possession Sound, and the Main Basin (Fig. 1) (Nairn and Kawase 2002). Many other modeling studies on the smaller-scale estuarine circulation in individual estuaries and the various interactions in Puget Sound have been conducted in the past in connection with site-specific water-quality-management actions. Previous studies were conducted mostly with structured-grid approaches, typically using simplified representations of the nearshore geometry and neglecting the tide flat regions. Wetting and drying processes in tide flats are particularly important from the perspective of nearshore salmon habitat restoration. In recent years, modeling studies have been conducted with unstructured-grid models to simulate the tidal circulation and salinity intrusion in estuaries of Whidbey Basin of Puget Sound in connection with nearshore wetland restorations that address the above concerns (Yang and Khangaonkar 2009; Yang et al. 2010a, b) and to investigate the complex, small-scale, flow processes in the Snohomish River estuary (Wang et al. 2009).

The movement of river plumes and density-induced circulation in each individual estuary and sub-basin plays an important role and contributes to the larger scale estuarine circulation in Puget Sound. Accurate simulations of individual river plumes, stratification, and baroclinic motions are critical to accurately represent the large-scale physical processes and assess the nearshore ecological processes (de Brye et al. 2010). The small, estuarine-scale processes (in the range of 10 to 100 m) are highly influenced by the outer boundaries provided by large coastal-scale processes. Neglecting small-scale processes or the lack of model resolution could affect the large-scale estuarine circulation and stratification. Similarly, it is important to correctly address the larger scale behavior of the system to make certain that important processes, such as saline water intrusion into the individual estuaries and sub-basins, are simulated accurately.

Barnes and Collias (1958) considered the influence of coastal processes on dissolved oxygen in Puget Sound. This consideration is particularly true for Puget Sound where larger scale processes show considerable variability between the wet and dry seasons. These processes include wind-driven upwelling from the continental slope (Hickey and Banas 2003) and baroclinic mixing over the sills (Ebbesmeyer and Barnes 1980), resulting in stratification that varies with freshwater inflow. These processes in turn affect the finer scale response within the sub-basins.

To further understand and investigate circulation patterns in Puget Sound and assess the exchange interaction between estuaries and sub-basins, a high-resolution, multi-scale, estuarine and coastal circulation model for the entire Puget Sound was implemented. Here, the development of the multi-scale Puget Sound model using an unstructured-grid approach is described. The overall performance of the model, including tidal characteristics (elevation and current), freshwater plumes, wetting and drying processes on tide flats, temperature/salinity, and mean circulation in the sub-basins is also described.

2 Description of study domain—Puget Sound

Pacific Ocean water enters the Puget Sound estuarine system and the Strait of Georgia through the deep narrow Strait of Juan de Fuca. The Strait of Juan de Fuca connects the Strait of Georgia around Vancouver Island of British Columbia and Puget Sound (Fig. 1). A large freshwater discharge in the Strait of Georgia (primarily the Frasier River) and outflow

from Puget Sound set up the estuarine circulation patterns in the Strait of Juan de Fuca. The Strait of Juan de Fuca, therefore, is also the outlet for fresh water leaving Puget Sound and the Strait of Georgia. The Fraser River plume travels southward to the Strait of Juan de Fuca through a series of complex waterways of the San Juan Island Passages.

Puget Sound is commonly defined as the waterbody south of three entrances—the main entrance at Admiralty Inlet; a second entrance at Deception Pass; and a third entrance at the south end of the Swinomish Channel, which connects Skagit Bay and Padilla Bay (U. S. Geological Survey 1979). The Puget Sound estuarine system consists of four deep basins connected by shallower sills: Hood Canal, Whidbey Basin, South Puget Sound, and the Main Basin, which is further subdivided into Admiralty Inlet and the Central Basin (King County 2001). Exchanges between Puget Sound water and the saline Pacific Ocean water occur mainly through the Admiralty Inlet, which is defined as the entrance of Puget Sound (Collias 1972). A shallow 30-km-long double-sill at about 65 m depth in Admiralty Inlet (sometimes referred to as the "seaward sill") controls the water exchange between Puget Sound and the Strait of Juan de Fuca (Ebbesmeyer and Barnes 1980; Cannon 1983). Deception Pass and the Swinomish Channel are two narrow pathways that contribute a very small percentage of the water exchanges for the entire Puget Sound, but may have a significant effect on local estuarine circulation in the Skagit-Padilla Bay system. The Central Basin is the largest and deepest basin in Puget Sound. It connects with Admiralty Inlet to the north, South Puget Sound to the south, and Whidbey Basin to the northeast. South Puget Sound consists of a number of narrow inlets and is separated from the Central Basin by another shallow sill at the Tacoma Narrows at about 44 m depth (sometimes referred to as the "landward sill"). The Tacoma Narrows is the only passage between the Central Basin and the South Puget Sound. Hood Canal is the smallest and least complex basin in terms of geometry in Puget Sound. Another shallow sill approximately 53 m deep exists at the entrance of Hood Canal, which restricts water exchange between Admiralty Inlet and Hood Canal. Whidbey Basin connects to the Central Basin through Possession Sound. The Whidbey Basin contains the three largest rivers that discharge into Puget Sound: the Skagit River, Snohomish River, and Stillaguamish River. Unlike other sub-basins, there are no sills in Whidbey Basin.

3 Hydrodynamic model of Puget Sound

3.1 Hydrodynamic model

The 3-D unstructured-grid finite volume coast ocean model (FVCOM) (Chen et al. 2003) simulates water-surface elevation, velocity, temperature, salinity, sediment, and water-quality constituents in an integral form by computing fluxes between non-overlapping horizontal triangular control volumes. This finite volume approach combines the advantages of finite-element methods for geometric flexibility and finite-difference methods for simple discrete structures and computational efficiency. A sigma-stretched coordinate is used in the vertical direction to better represent the irregular bottom topography. Unstructured triangular cells are used in the horizontal plane. The detailed theoretical aspects of the model are presented in Chen et al. (2006). The model has been used to simulate the circulation and physical processes in estuaries, coastal waters, and open oceans (Chen et al. 2009; Xue et al. 2009; Li et al. 2008; Huang et al. 2008; Zhao et al. 2006; Isobe and Beardsley 2006; Aoki and Isobe 2007; Weisberg and Zheng 2006a); wetting and drying processes in tide flats (Zheng et al. 2003; Yang and Khangaonkar 2009); physical–biological interactions (Tian et al. 2009; Ji et al. 2008, 2009; Chen and Rawson 2005); storm surges (Weisberg and Zheng 2006b; Qi et al. 2009); and nearshore restoration (Yang et al. 2010a, b).

The 3-D governing equations for Reynolds-averaged turbulent flows with the Boussinesq approximation are represented in the following form in FVCOM:

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0 \tag{1}$$

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+w\frac{\partial u}{\partial z}-fv=-\frac{1}{\rho_o}\frac{\partial P}{\partial x}+\frac{\partial}{\partial z}\left(K_m\frac{\partial u}{\partial z}\right)+F_u \tag{2}$$

$$\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+w\frac{\partial v}{\partial z}+fu=-\frac{1}{\rho_o}\frac{\partial P}{\partial y}+\frac{\partial}{\partial z}\left(K_m\frac{\partial u}{\partial z}\right)+F_v \tag{3}$$

$$\frac{\partial P}{\partial z}=-\rho g \tag{4}$$

where (x, y, z) are the east, north, and vertical axes in the Cartesian coordinates; (u, v, w) are the three velocity components in the x, y, and z directions; (F_u, F_v) are the horizontal momentum diffusivity terms in the x and y directions; K_m is the vertical eddy viscosity coefficient; ρ is density; P is the pressure; and f is the Coriolis parameter.

The 3-D transport equations for temperature and salinity are solved simultaneously in FVCOM:

$$\frac{\partial T}{\partial t}+u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}+w\frac{\partial T}{\partial z}=\frac{\partial}{\partial z}\left(K_h\frac{\partial T}{\partial z}\right)+F_T \tag{5}$$

$$\frac{\partial S}{\partial t}+u\frac{\partial S}{\partial x}+v\frac{\partial S}{\partial y}+w\frac{\partial S}{\partial z}=\frac{\partial}{\partial z}\left(K_h\frac{\partial S}{\partial z}\right)+F_S \tag{6}$$

where T and S are temperature and salinity, K_h is the vertical eddy diffusivity coefficient, and (F_T, F_S) are the horizontal thermal and salt diffusivity terms.

Temperature and salinity are related to density through the equation of state:

$$\rho=\rho(T,S) \tag{7}$$

The bottom friction is described by the quadratic law with the drag coefficient determined by the logarithmic bottom layer as a function of bottom roughness (Blumberg and Mellor 1987). Wind stress is specified at the water surface (Large and Pond 1981). The model employs the Smagorinsky scheme for horizontal mixing (Smagorinsky 1963) and the Mellor–Yamada level 2.5 turbulent closure scheme for vertical mixing (Mellor and Yamada 1982). The effect of breaking, wind-driven, surface waves on turbulent energy (Mellor and Blumberg 2004) is also incorporated in FVCOM.

3.2 Model configurations

An unstructured grid of FVCOM for the entire Puget Sound and the Straits of Juan de Fuca and Georgia was developed with bathymetric data from the digital elevation model and light detection and ranging (LiDAR). The model open-boundary conditions were specified at the entrance of the Strait of Juan de Fuca and the north end of the Strait of Georgia. For model efficiency, the model grid was generated in a way such that coarse-grid resolution was used in the areas away from Puget Sound, and a fine-grid resolution was specified in the Puget Sound. In the tide flat regions, LiDAR data were used to improve the model bottom topography, which is important for simulating a freshwater front and salinity dispersion in the river deltas. The model element size varied from approximately 2,000 m in the open boundaries to around 50 m in the estuaries and tide flats and as small as less than 20 m in the tidal channels and the wetland inside the estuaries. The model consists of about 214,000 elements and 118,000 nodes in the horizontal plane. Twenty vertical layers with power distribution were specified in the water column in a sigma-stretched coordinate system. The model grid was set up in Universal Transverse Mercator North American Datum 83 (Zone 10) in the horizontal plane, with reference to NAVD 88 in the vertical direction. From close-ups of the model grids in the sub-domains, including the Snohomish Basin, Skagit Bay, South Puget Sound, and the San Juan Islands (Fig. 2), tide flats and shallow-water tidal channels in the river deltas are clearly visible.

Tidal boundary conditions were specified along the two open boundaries using eight harmonic tidal constituents (S2, M2, N2, K2, K1, P1, O1, and Q1) based on National Oceanic and Atmospheric Administration (NOAA) observations. An example of the composite tidal elevations along the open boundaries is demonstrated (Fig. 3). Open-boundary temperature and salinity are important to the mean temperature and salinity distribution and circulation in the Straits and Puget Sound. In this study, constant temperature and salinity values along open boundaries (7.4°C and 33.5 ppt on eastern boundary; 9.3°C and 30.6 ppt on northern boundary) were estimated based on historical monthly observations conducted by the Department of Fisheries and Oceans, Canada near the open boundaries and were adjusted as part of model calibration. Initial temperature and salinity conditions were specified uniformly as 9°C and 31 ppt, respectively, in the entire model domain. The model was run with year 2005 forcings to obtain the final initial condition for the year 2006 model run.

There are 19 major rivers that discharge to the Straits and Puget Sound region. River inflows are specified in all the major rivers in the Straits and Puget Sound. River data in the United States were obtained from either real-time or historical ensemble average flow data from the U.S. Geological Survey (USGS) and the Water Survey of Canada. The annual means of daily flows in the Straits and each sub-basin of Puget Sound based on long-term historical data are summarized in Table 1. To characterize the relative importance of the freshwater discharge in each sub-basin, volumes of sub-basins and the ratio of total annual river flow to the basin volume were also calculated based on the model grid (Table 1). As can be seen, the Whidbey Basin consists of nearly 70% of the total inflow into Puget Sound while its basin volume is only about 16.6% of Puget Sound. The ratio of total annual river flow to the basin volume of 1.02 is about an order of magnitude greater than that in other sub-basins. South Puget Sound has the least river inflow with 6% of the total inflow. Although Hood Canal only consists of 10% of the total flow discharge into Puget Sound, the ratio of the total annual river flow to the basin volume is higher than the Main Basin and South Sound because of the relatively small basin volume.

Seasonal variations of the freshwater discharges in the sub-basins of Puget Sound are similar (Fig. 4), with high

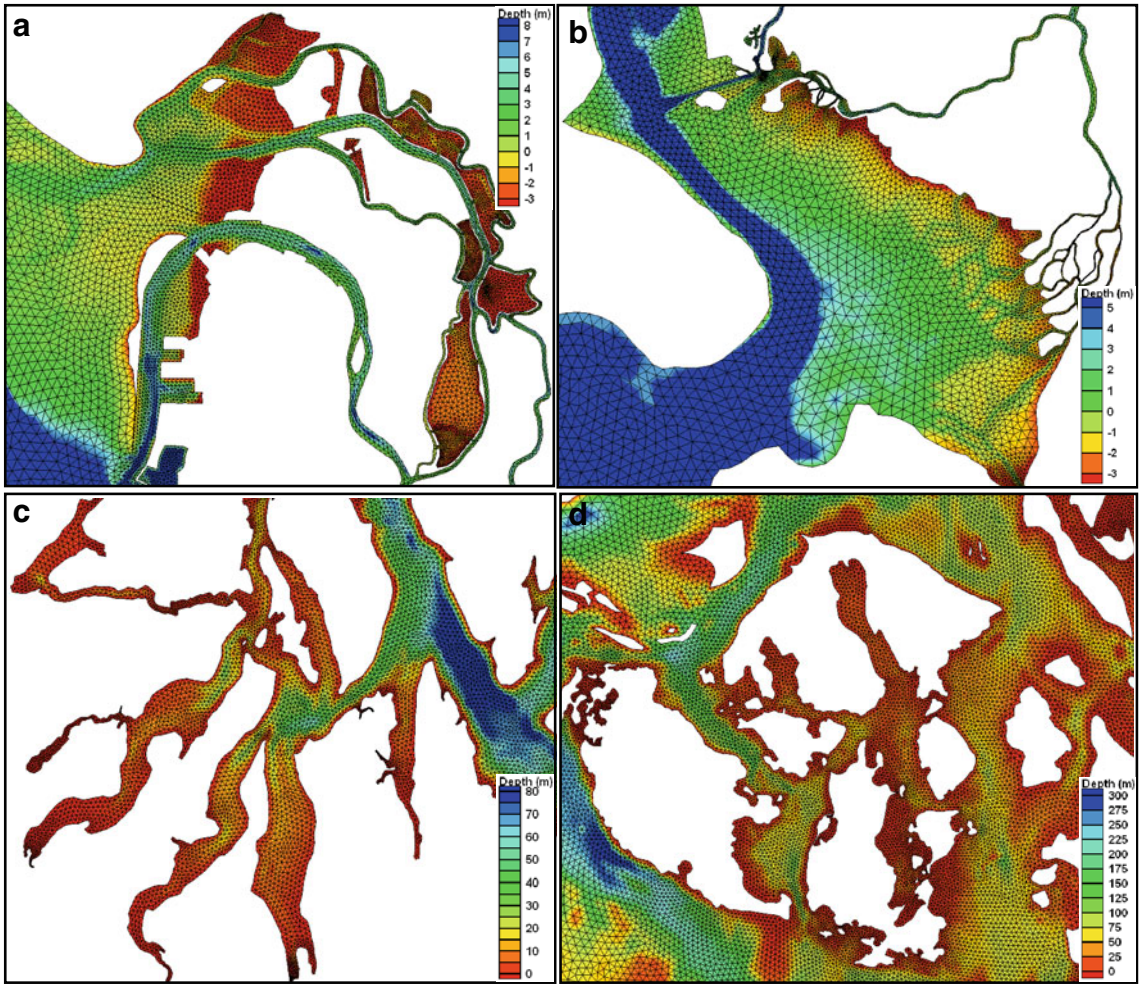

Fig. 2 Model grids in different regions of Puget Sound: **a** Snohomish Basin, **b** Skagit Bay, **c** South Puget Sound, **d** San Juan Islands

flows during late fall and winter periods and relatively low flows in late summer. In contrast, the Fraser River inflow, which is significantly higher than the rest of the inflows into Puget Sound and the Straits, shows a different seasonal distribution pattern with a high flow in late spring and summer and a low flow in the fall and winter. River temperatures are only available for a limited number of major rivers (including the Fraser River) and often do not cover the entire year. Comparisons of monthly measured temperatures from seven major rivers with average monthly values recorded at the USGS station on the Cedar River indicated that the Cedar River data show similar patterns to the seven major rivers. Therefore, as an approximation, the Cedar River values were assigned to all rivers (except the Fraser River) in Puget Sound river temperatures in the model setup.

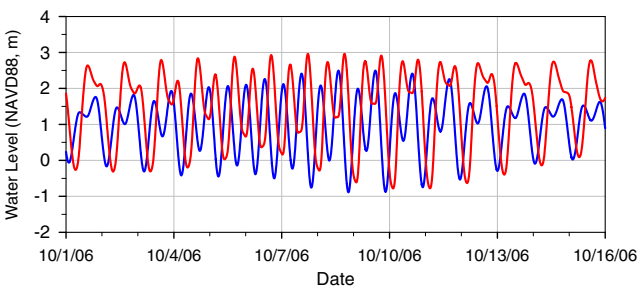

Fig. 3 Tidal elevations at open boundaries: entrance of the Strait of Juan de Fuca (*blue*) and north end of the Strait of Georgia (*red*)

Wind effects were also included in this study. Wind could be an important factor that affects the local mean circulation patterns in the sub-basins of Puget Sound. For simplicity, time series of wind vectors were applied uniformly to the entire model domain at hourly intervals. Wind stress within FVCOM is calculated based on the well-known method by Large and Pond (1981). Several long-term NOAA meteorological data stations around the Straits and Puget Sound were examined. The observations of meteorological variables were not consistent and available at all the stations. However, the Seattle station, which is located near the center of the model domain, has the most compelete data and thus was used in the model simulation. A pre-calculated net heat flux was specified at the water

Table 1 Summary of river inflows

River and basin name	Annual mean flow (m^3/s)	Volume ($\times 10^{10}$ m^3)	River flow/basin volume (yr^{-1})
Straits/North Sound	2,541.0	156.2	0.05
Fraser (Canada)	2,351.6		
Dungeness	14.8		
Elwha	52.1		
Nooksack	109.8		
Samish	12.5		
Whidbey Basin	919 (68%)[a]	2.85 (16.6%)[a]	1.02
Skagit	475.5		
Stillaguamish	144.6		
Snohomish	298.8		
Main Basin	214 (16%)[a]	10.16 (59.1%)[a]	0.07
Lake Washington	48.8		
Green/Duwamish	52.3		
Puyallup	112.7		
Hood Canal	139 (10%)[a]	2.60 (15.1%)[a]	0.17
Tahuya	10.6		
Skokomish	56.9		
Duckabush	16.7		
Dosewallips	20.9		
Hamma Hamma	23.5		
Big Quilcene	10.0		
South Sound	76 (6%)[a]	1.59 (9.2%)[a]	0.15
Nisqually	58.3		
Deschutes	17.8		

[a] Percentage is relative to the total inflows in Puget Sound

surface. Meteorological parameters for calculating net heat flux in FVCOM include (1) downward and upward shortwave radiation, (2) downward and upward longwave radiation, (3) latent heat flux, and (4) sensible heat flux. These meteorological parameters were obtained from the North Americal Regional Reanalysis (NARR) data generated by NOAA National Centers for Environmetal Prediction. Details of net heat flux calculation and specificiation in the model is presented in Yang et al. (2011).

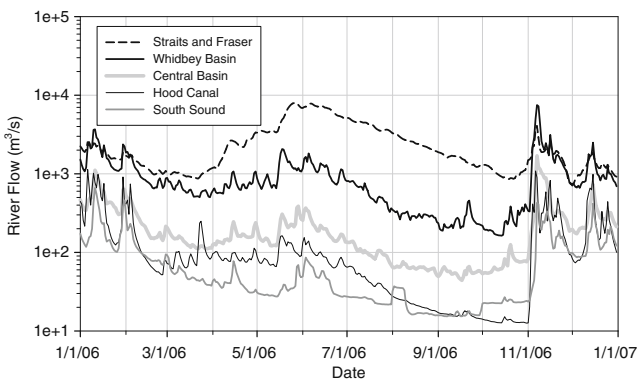

Fig. 4 River inflows in the straits and major basins of Puget Sound (shown in log scale because of large range of inflow rates)

The simulation of tides in a large region is generally sensitive to the magnitude of the bottom friction coefficient and roughness (Crean 1978; Lavelle et al. 1988). A bottom friction coefficient of 0.005 and a bottom roughness of 0.002 m were used in the model. These values are comparable to the values in previous studies (Sternberg 1968; Mofjeld and Larsen 1984). The Smagorinsky multiplicative coefficient was set to 0.2, and a background value of vertical eddy viscosity of 10^{-6} m^2/s was used. The minimum depth for wetting–drying simulation was specified to 0.05 m.

4 Model results and discussion

4.1 Tidal elevations and currents

Accurate simulation of tidal propagation from the Pacific Ocean into Puget Sound is a fundamental requirement in developing the multi-scale model of Puget Sound. There are six real-time tidal stations maintained by NOAA throughout the Strait of Juan de Fuca and Puget Sound. They are located at Port Angeles, Friday Harbor and Cherry Point in the Strait of Juan de Fuca and Port Townsend,

Table 2 Comparison of observed and modeled harmonic constituents of tidal elevations (magnitude [meter]/phase [degree])

Station	M2	S2	N2	K1	O1	P1
Port Angeles	0.49/104.2	0.10/112.6	0.08/40.5	0.72/141.4	0.46/168.4	0.29/158.1
	0.51/101.6	**0.11/120.6**	**0.09/34.8**	**0.67/143.9**	**0.45/170.9**	**0.30/160.6**
Friday Harbor	0.55/167.5	0.17/176.2	0.09/109.1	0.83/159.0	0.50/184.3	0.35/177.0
	0.56/147.1	**0.18/166.8**	**0.08/86.2**	**0.77/157.6**	**0.48/184.3**	**0.35/176.5**
Cherry Point	0.72/179.3	0.23/185.3	0.12/120.7	0.88/161.2	0.53/186.0	0.37/180.3
	0.65/159.6	**0.23/172.6**	**0.10/101.4**	**0.81/160.6**	**0.50/186.5**	**0.37/179.8**
Port Townsend	0.66/147.3	0.18/160.1	0.11/84.6	0.82/149.6	0.52/176.7	0.35/170.1
	0.69/138.4	**0.20/154.4**	**0.11/74.7**	**0.75/150.8**	**0.50/178.5**	**0.34/169.0**
Seattle	1.04/168.1	0.32/175.2	0.17/105.8	0.88/156.5	0.53/180.3	0.40/176.8
	1.07/158.9	**0.35/173.6**	**0.17/97.6**	**0.82/158.0**	**0.52/183.6**	**0.39/178.8**
Tacoma	1.11/168.8	0.33/176.2	0.18/106.4	0.90/156.7	0.54/180.2	0.39/175.5
	1.13/160.1	**0.38/174.9**	**0.18/99.1**	**0.83/158.2**	**0.53/183.8**	**0.39/179.4**

Data in boldface are values for modeled harmonic constituents of tidal elevations

Seattle, and Tacoma inside Puget Sound (Fig. 1). Tidal elevation magnitudes and phases for six major tidal constituents (M2, S2, N2, K1, O1, and P1) are compared at all six stations (Table 2); the dominant tidal constituents are M2 and K1. The tide around Vancouver Island is classified as mixed tide with the form of the ratio of major diurnal to semi-diurnal constituents [F=(K1+O1)/(M2+S2)] in the range of 0.25 to 3.0 with a break value of 1.5 between mixed semi-diurnal and diurnal tides (Pond and Pickard 1983). Based on the observed data and model results, the tide in Port Angeles (and thus the Strait of Juan de Fuca) is a mixed diurnal tide (F_{data}=2.0 and F_{model}=1.8), often with only one high tide of water per day. As the tide propagates into Puget Sound, it becomes mixed semi-diurnal (e.g., Tacoma station) (F_{data}=1.0 and F_{model}=0.91), which shows large inequalities in range and time between high tides and low tides. Tidal magnitudes for the dominant constituent M2 and K1 increase significantly from the Strait of Juan de Fuca (Port Angeles) to South Sound (Tacoma), with 122% for M2 and 24% for K1, respectively. Station-averaged root mean square error (RMSE) statistics for all the constituents are calculated (Table 3). The station-averaged modeled phase for major semi-diurnal tide M2 has an error of 13.2° (27.3 min). The M2 tidal phase was underpredicted at all stations, especially at Friday Harbor and Cherry Point (Table 1). The large phase errors at Friday Harbor and Cherry Point are likely caused by the approximation of the northern open boundary in the Strait of Georgia. Comparisons of time series of observed and modeled tidal elevations at selected stations demonstrate that the model reproduces the tidal amplitudes, phases, spring-neap tide cycle, and inequalities (Fig. 5). The correlation coefficient R^2 at each station is also calculated and provided in the figure. Close correlations between modeled results and observed data are shown with R^2 values greater than 0.9.

Subtidal variations of water-surface elevations were calculated and compared to observed data at all six stations. The comparison of modeled and observed subtidal surface

Table 3 RMSE statistics for tidal elevation calibration

Constituent	M2	S2	N2	K1	O1	P1
Amplitude (m)	0.04	0.03	0.01	0.06	0.02	0.01
Phase (degree)	13.2	7.6	13.8	1.6	2.4	2.1

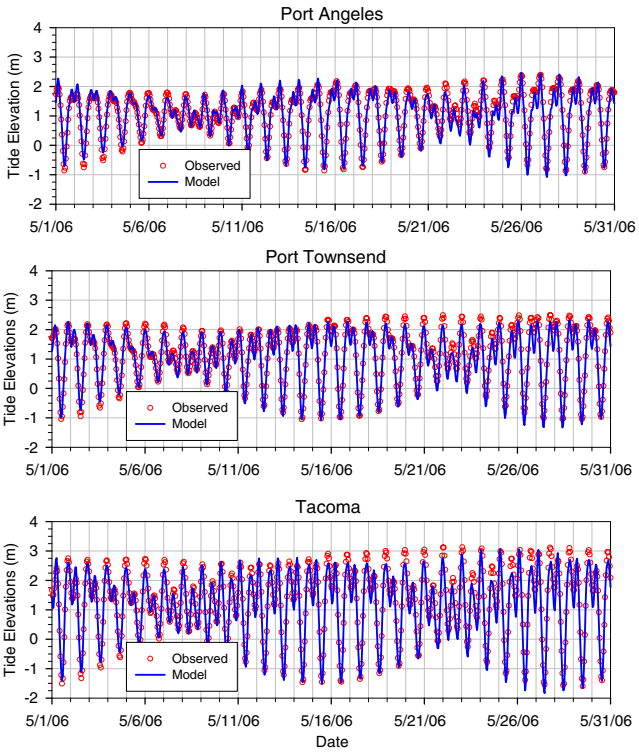

Fig. 5 Modeled and observed water-surface elevations at selected stations (correlation coefficients, R^2, are provided in the *plots*)

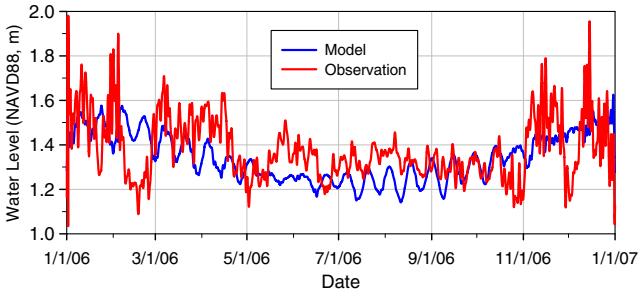

Fig. 6 Modeled and observed subtidal water-surface elevations at Tacoma Station

elevations at Tacoma station was presented here because the time series distribution patterns at all stations were quite similar for both observed data and model results. Observed data showed subtidal anomalies in winter and spring, but the model did not reproduce such seasonal anomalies (Fig. 6). Further analysis of NOAA observed data at Neah Bay station near the model open boundary at the Strait of Juan de Fuca suggested that the seasonal variations were likely induced by seasonal meteorological forcing variations and propagated into Puget Sound.

Validating modeled velocities in a large complex estuary system like Puget Sound is a challenge due to the paucity of synoptic velocity measurements. Here, the model results are compared against acoustic Doppler current profiler data at selected locations in Puget Sound. These locations include Deception Pass (DP), Admiralty Inlet (PW) at the main entrance of Puget Sound (Public Utility District No. 1 of Snohomish County 2007), Dana Passage (Dana) in South Puget Sound (Washington Department of Ecology 2008), and Hood Canal (7731) (U. S. Geological Survey 2007) (Fig. 1). Similar to tidal elevations, comparisons of tidal current magnitudes and phases between model results and observed data were made for six major tidal constituents at the surface and bottom levels of the water column (Table 4). Model results in general match observations well in terms of velocity magnitudes and phases. Compared to bottom velocities, modeled surface velocities tend to have slightly larger errors.

Tidal elevations are dominated by diurnal tides (K1) in the Straits and near the entrance of Puget Sound and become dominant semi-diurnal tides (M2) inside Puget Sound. Unlike tidal elevations, the major semi-diurnal component (M2) in tidal currents is generally much greater than the diurnal component (K1) at all locations. Velocities are the strongest in Deception Pass, which is a narrow waterway connecting Skagit Bay and the Strait of Juan de Fuca. In contrast, velocities in Hood Canal are the smallest, with all the constituent components less than 0.1 m/s. Time series of modeled and observed tidal currents along the main axis in Admiralty Inlet, Deception Pass, Hood Canal, and South Sound are compared (Fig. 7). The spring-neap tidal cycles occur in both modeled results and observed data, except in Hood Canal where tidal currents are extremely weak. The correlation coefficient, R^2, at each station was also calculated and provided in Fig. 7. High correlation values were obtained at locations where tides are dominant. Relatively low correlation coefficients were found in the low-current Hood Canal basin. Station-

Table 4 Comparison of observed and modeled harmonic constituents of surface and bottom tidal currents (magnitude [meter per second]/phase [degree])

Station	M2	S2	N2	K1	O1	P1
Deception Pass (surface)	2.69/40.4	0.63/38.3	0.31/9.1	0.51/14.7	0.30/27.6	0.17/19.9
	2.44/49.4	**0.58/35.7**	**0.31/15.1**	**0.48/16.5**	**0.28/40.6**	**0.09/87.9**
Deception Pass (bottom)	1.82/40.9	0.44/40.8	0.21/7.9	0.34/16.8	0.22/35.5	0.09/25.8
	1.83/51.3	**0.39/31.0**	**0.26/17.4**	**0.27/12.4**	**0.21/36.2**	**0.10/78.5**
Admiralty Inlet (surface)	1.54/14.7	0.42/105.7	0.24/53.4	0.65/67.0	0.41/7.9	0.18/73.0
	1.43/6.2	**0.35/118.3**	**0.24/46.7**	**0.57/78.5**	**0.38/12.0**	**0.16/108.3**
Admiralty Inlet (bottom)	0.91/7.6	0.20/113.2	0.19/54.7	0.34/88.9	0.21/15.4	0.07/158.3
	0.87/2.5	**0.18/114.9**	**0.16/42.3**	**0.37/80.5**	**0.25/16.0**	**0.08/110.8**
South Sound (surface)	0.82//53.0	0.28/94.9	0.19/117.8	0.40/104.8	0.16/62.2	0.16/123.9
	0.82/74.3	**0.42/92.7**	**0.20/142.4**	**0.54/118.7**	**0.16/69.6**	**0.32/139.4**
South Sound (bottom)	0.70/57.4	0.22/102.2	0.15/122.4	0.31/108.7	0.12/63.4	0.10/124.4
	0.65/75.4	**0.34/88.3**	**0.16/140.5**	**0.42/120.7**	**0.12/72.4**	**0.25/141.5**
Hood Canal (surface)	0.09/104.8	0.03/140.1	0.01/51.3	0.04/133.2	0.02/251.9	0.02/171.6
	0.09/74.4	**0.03/162.3**	**0.01/38.6**	**0.06/118.7**	**0.03/93.6**	**0.05/176.4**
Hood Canal (bottom)	0.06/105.5	0.01/145.9	0.01/13.0	0.04/69.4	0.02/104.8	0.01/96.3
	0.05/106.3	**0.01/140.8**	**0.01/27.7**	**0.02/128.9**	**0.01/74.1**	**0.01/168.2**

Data in boldface are values for harmonic constituents of surface and bottom tidal currents

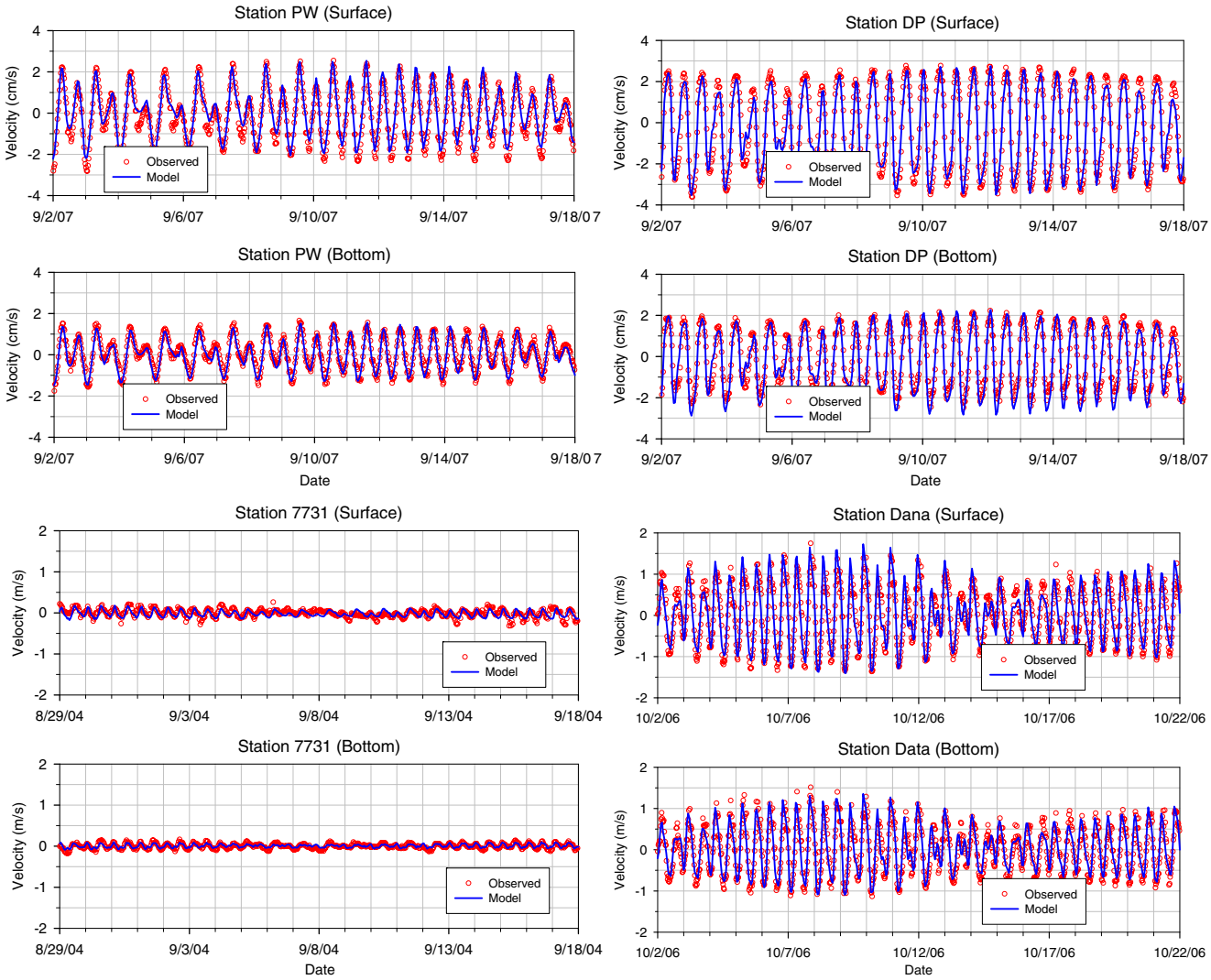

Fig. 7 Modeled and observed currents (*main axis*) in Admiralty Inlet (*PW*), Deception Pass (*DP*), Hood Canal (*7731*), and the South Puget Sound (*Dana*) (correlation coefficients, R^2, are provided in the plots)

averaged RMSE statistics for all tidal current constituents are also calculated for all stations (Table 5). RMSEs for modeled magnitudes are all within 0.1 m/s. The modeled phase errors for major semi-diurnal tide M2 and diurnal tide K1 are relatively large, primarily due to the difficulty in matching small current magnitudes in the Hood Canal station (Table 4).

4.2 Temperature/salinity profiles and mean circulation

While the tides are the dominant force in driving the circulation in the hourly to daily time scale in Puget Sound, temperature and salinity stratification and gradients in estuaries generally play a more important role in the long-term (from days to year time scale) mean circulation patterns and water exchanges in the estuary. Hence, it is important that the Puget Sound model also properly simulates the behavior of the salinity intrusion and density stratification in Puget Sound. While long-term synoptic temperature and salinity time series are very limited throughout Puget Sound, monthly temperature and salinity data at various locations of Puget Sound are available through the Washington Department of Ecology's ongoing ambient monitoring program. To illustrate the capability for

Table 5 RMSE statistics for velocity calibration

Constituent	M2	S2	N2	K1	O1	P1
Amplitude (m/s)	0.10	0.07	0.02	0.08	0.02	0.08
Phase (degree)	15.8	11.1	14.3	23.2	57.4	45.6

modeling temperature and salinity stratification, modeled temperature and salinity profiles are compared to monthly observed data at six different locations in Puget Sound for high river flow (February) and low river flow (September) conditions, respectively (Fig. 8).

Overall, modeled temperature and salinity profiles follow the general trends of observed data at all stations in both high-flow and low-flow conditions. Model temperatures follow the observed patterns with surface cooling in February (late winter) and warming in September (early fall). Stratification at station ADM2 near the entrance of Admiralty Inlet (Fig. 1) is very weak with temperature and salinity slightly increased during low-flow conditions in September. Station ADM3 is located in Admiralty Inlet near the entrance of Possession Sound where most of the freshwater from Whidbey Basin flows out of Puget Sound. Salinity stratification is seen in the surface layer, especially in the high-flow season (February). Salinity distributions show a weakly stratified water column in both high-flow and low-flow seasons at station EAP in Central Sound. Salinity profiles at station SAR3 in Saratoga Passage are highly stratified, especially in the upper surface layer, because of the influence of Skagit River, which is the largest river discharge to Puget Sound (Table 1). Similarly, high-salinity stratification also occurs in Hood Canal in both high-flow and low-flow conditions. This occurrence of high-salinity stratification is mainly because of low tidal energy and a relatively high ratio of river flow to the basin volume in Hood Canal (Table 1). Salinity at station NSQ in South Sound tends to be more affected by local river flows through the water column because of shallow-water depth. RMSE statistics for temperature and salinity profiles between modeled results and observations at all six stations indicate that the model can simulate temperature and salinity distributions in Puget Sound (Table 6).

Horizontal 2-D distributions of monthly averaged surface salinity and velocity during the high river flow period (February) in Admiralty Inlet, Possession Sound, Saratoga Passage, and Hood Canal were examined (Fig. 9). A strong surface mean current induced by density gradient because of high freshwater runoff from Whidbey Basin flows out through Possession Sound and immediately turns north, joining the outgoing surface current from Central Basin and exiting Puget Sound through Admiralty Inlet. In Hood Canal, an outgoing (flowing north) surface mean current, originating from the Skokomish River, is also clearly seen. However, the surface mean current east of the Skokomish River flows to the end of Hood Canal as the Skokomish River splits part of its runoff to the east.

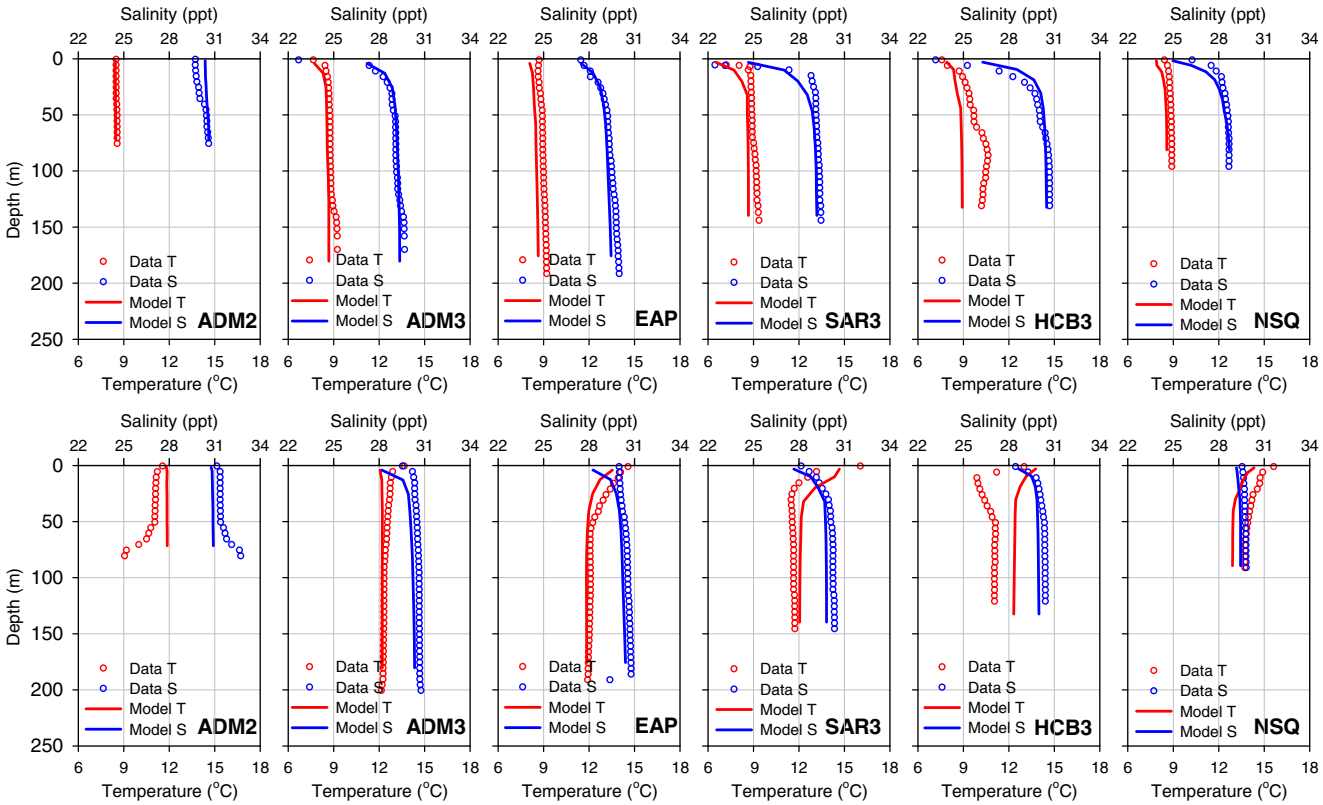

Fig. 8 Modeled (*line*) and observed (*circle*) temperature (*red*) and salinity (*blue*) profiles in February high-flow condition (*upper panel*) and September low-flow condition (*lower panel*)

Table 6 RMSE statistics for temperature and salinity profiles

Stations	ADM2	ADM3	EAP	SAR3	HCB3	NSQ
Temperature (°C)	0.69	0.41	0.53	0.83	1.52	0.77
Salinity (ppt)	0.43	0.30	0.42	1.22	0.87	0.54

Since 1998, the University of Washington has conducted cruises around 40 stations within Puget Sound as part of the Puget Sound Regional Synthesis Model (PRISM) project. The cruises are designed to collect a synoptic snapshot of the oceanographic conditions in the major sub-basins of Puget Sound in early summer (June) and winter (December) (http://www.nanoos.org/nvs/nvs.php?path=NVS-Assets-Cruises-PRISM). In 2006, oceanographic data were collected along transects in the Main Basin, Hood Canal, and Whidbey Basin. Modeled tidally averaged temperature, salinity, and longitudinal velocity were further examined along the Main Basin transect and compared to the PRISM cruise data for the periods of June 26 to 29 and December 15 to 18, 2006. In the June 2006 low-flow condition, a thin layer of brackish water with a warm temperature due to river runoff and surface heating resides on top of the saline bottom water (Fig. 10), which matches the general distribution pattern along the PRISM transect (Fig. 11). However, the model underpredicted the bottom salinity in the Strait of Juan de Fuca as well as the surface temperature. The misfit is likely caused by the specification of constant open-boundary salinity and the use of pre-calculated net heat flux from NARR data. Corresponding to the stratified temperature and salinity distribution is the well-defined two-layer mean circulation shown in the model results (Fig. 10; negative indicates outgoing current). The strong outgoing surface flow is compensated for by the thicker and weaker bottom saline inflow from the Strait of Juan de Fuca. Compared to June 2006, both model results and observed data showed stronger salinity stratification in December 2006, which corresponded to a high river flow condition (Figs. 12 and 13). Both model results and observed data show colder surface temperatures in December due to the effect of surface cooling. The two-layer mean circulation is also clearly seen in the model result along the Main Basin transect in December 2006 (Fig. 12).

4.3 River plumes in tide flat estuaries

Many of the estuaries in Puget Sound consist of shallow tide flats near the mouths. The physical processes, such as wetting and drying in tide flats in the intertidal zone, are important factors locally, but could also affect the overall circulation and salinity stratification in Puget Sound. Hence, it is critical to include the effects of wetting and drying processes in the model simulation. To accurately represent the tide flats and interconnected tidal channels near the mouths of estuaries, model bottom topography in the tide flat regions is defined by LiDAR data with ground-truth measurements and high-resolution aerial photographic

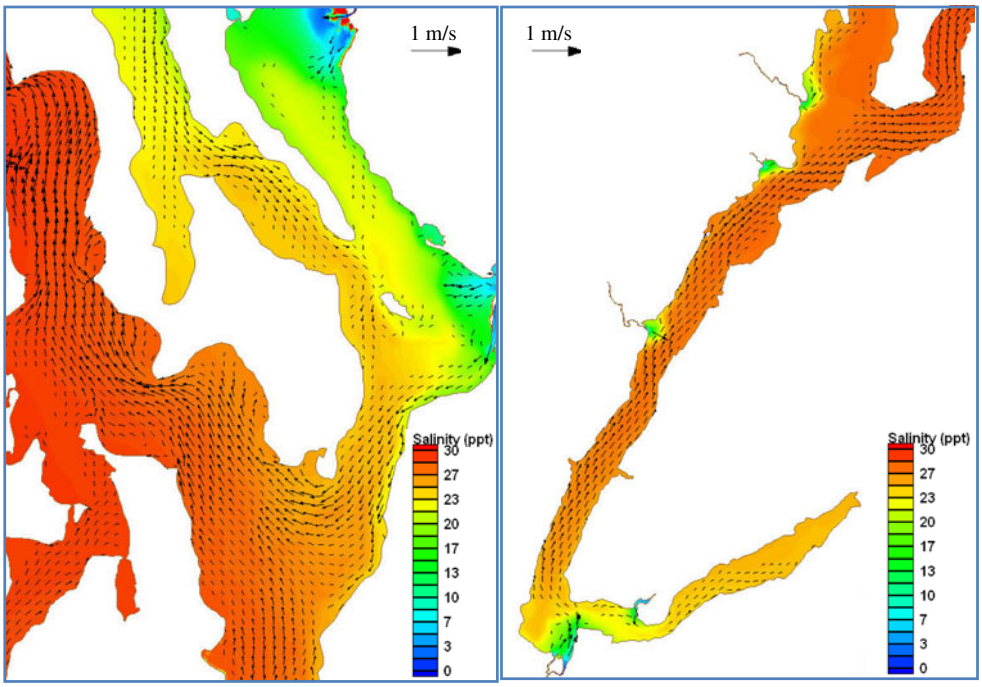

Fig. 9 Mean surface salinity and current in Admiralty Inlet, Possession Sound and Saratoga Passage (*left*), and Hood Canal (*right*). Velocity was plotted at rectangular grid because of the high density of unstructured-grid cells

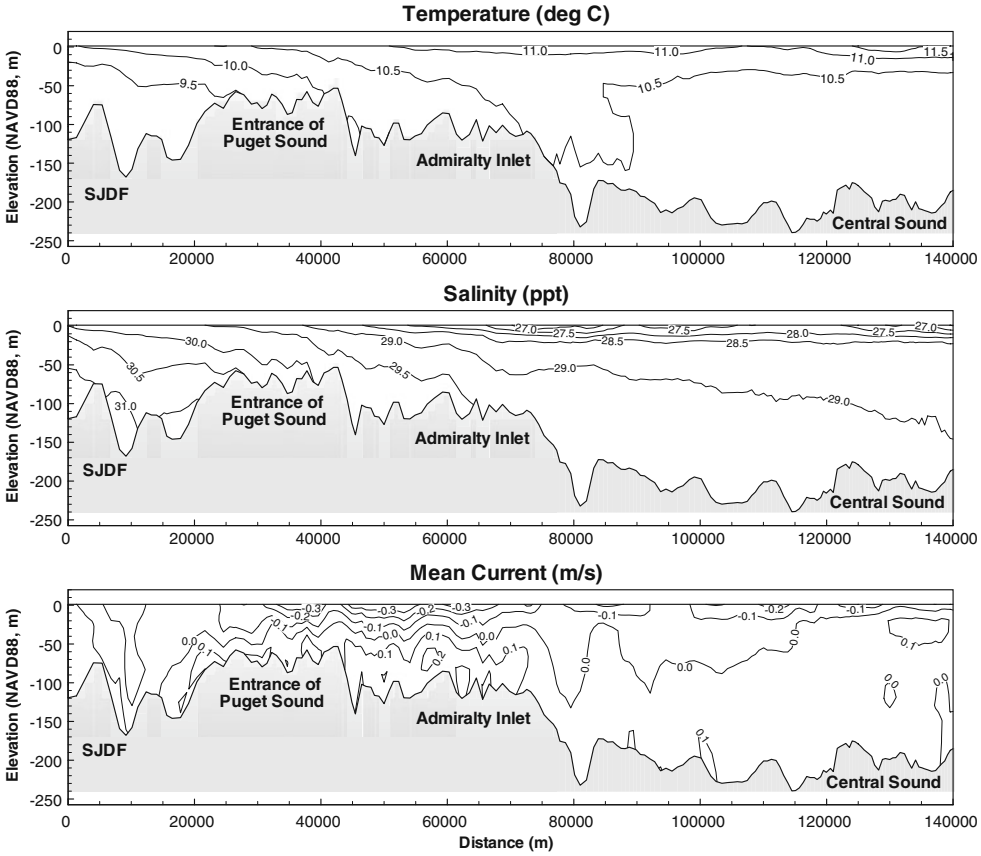

Fig. 10 Tidally averaged temperature (degree Celsius), salinity (parts per thousand) and longitudinal currents (meters per second) along the Main Basin transect for the period of June 26 to 29, 2006

images. Model grid sizes in the complex river distributaries and tide flat areas are as small as 20 m. The surface salinity and velocity distributions in the two largest tide flat estuaries in Puget Sound—Skagit Bay and Port Susan Bay in Whidbey Basin—are examined (Fig. 14). Large areas of tide flats become dry during low tide near the river mouths. During low tide, surface salinity distribution in an intertidal zone is mainly influenced by freshwater from the river. On the other hand, salinity distribution is mixed with incoming saline seawater during flood tide because of shallow-water depths and strong mixing in the intertidal zones. The stratification and destratification processes in

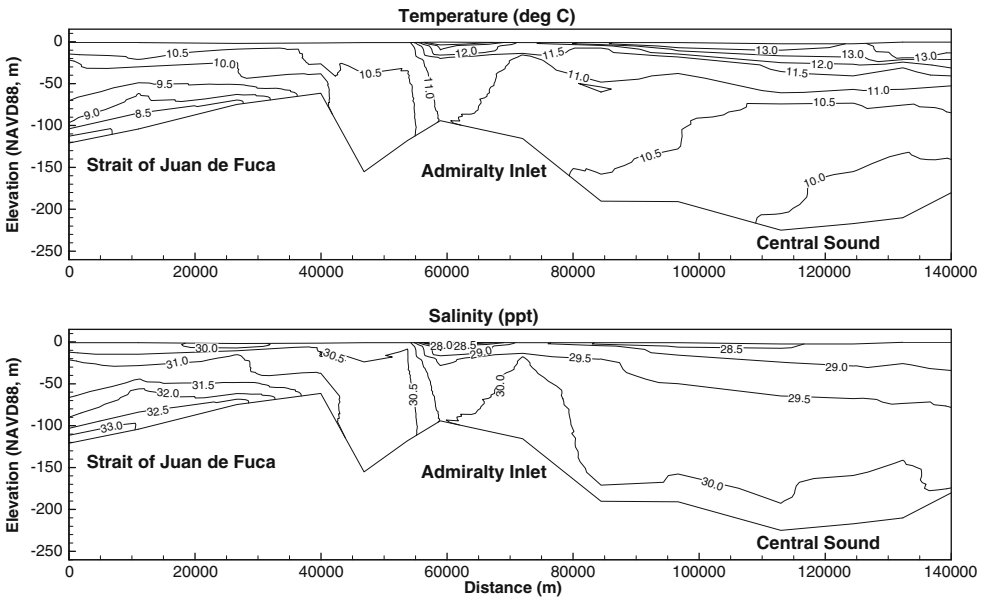

Fig. 11 Temperature (degree Celsius) and salinity (parts per thousand) distributions from PRISM cruise along the Main Basin transect for the period of June 26 to 29, 2006

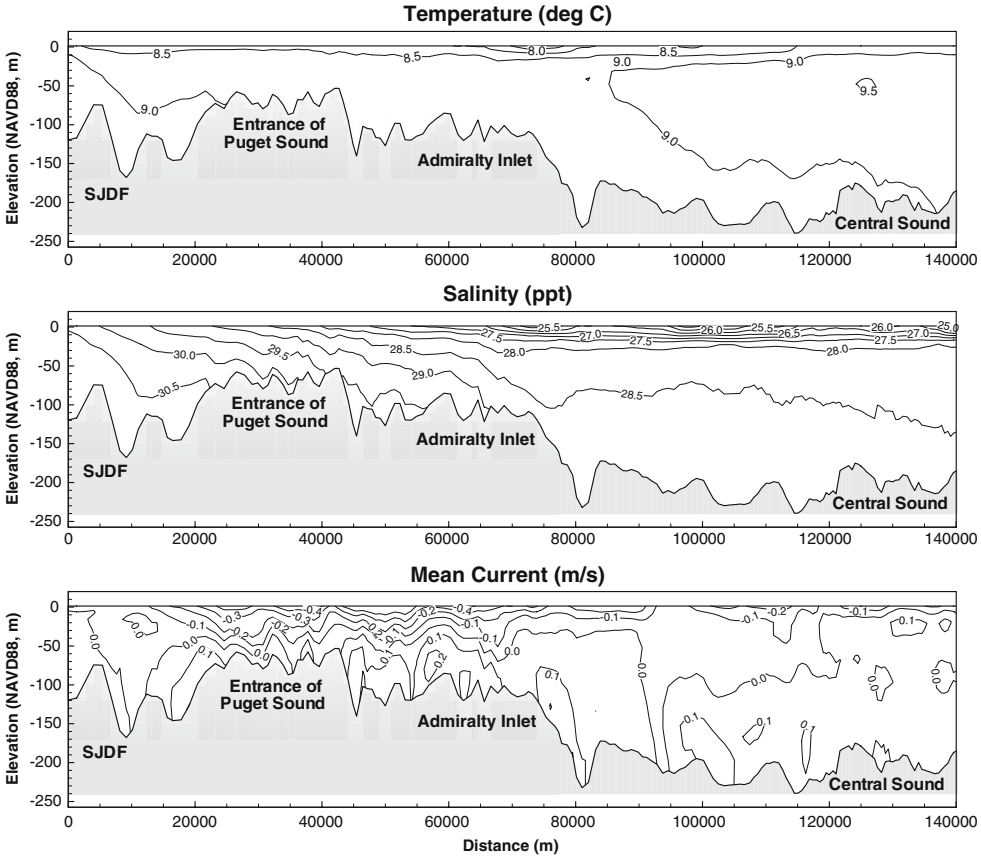

Fig. 12 Tidally averaged temperature (degree Celsius), salinity (parts per thousand), and longitudinal currents (meters per second) along the Main Basin transect for the period of December 15 to 18, 2006

the Skagit Bay intertidal zone are shown more clearly in the time series distributions of surface and bottom salinities in Skagit Bay (Fig. 15). Model results were extracted at a location with a depth of about 6.0 m near the tide flat regions in Skagit Bay. Strong temporal variations occur in the surface-layer salinity associated with each tidal cycle. The difference between bottom salinity and surface salinity indicates that vertical salinity variation can be as large as 18.0 ppt during low tide and as small as 1.0 ppt during high tide. Overall, the present Puget Sound model qualitatively simulates the freshwater plumes and wetting and drying processes in tide flat estuaries in Puget Sound.

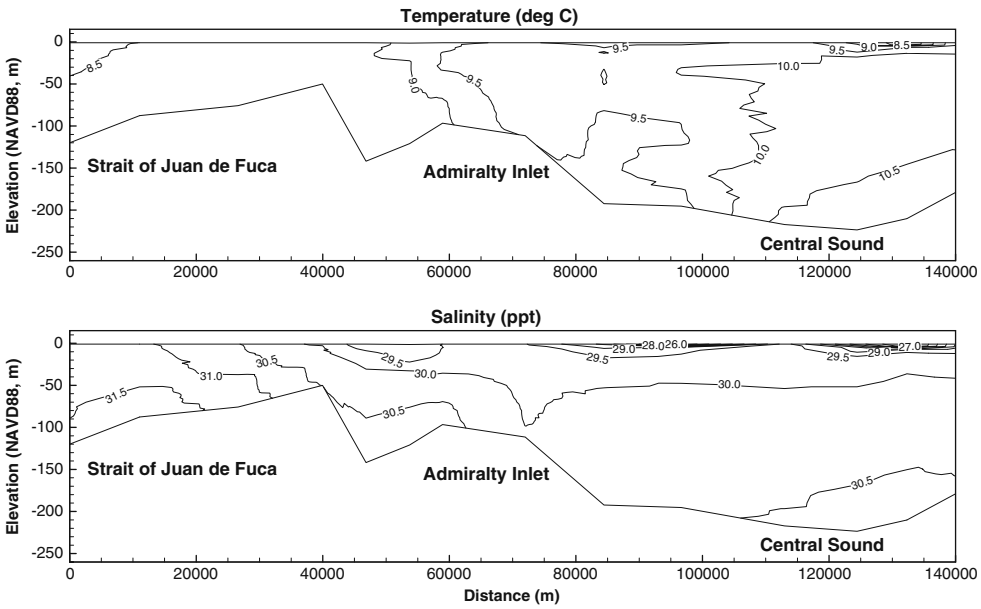

Fig. 13 Temperature (degree Celsius) and salinity (parts per thousand) distributions from PRISM cruise along the Main Basin transect for the period of December 15 to 18, 2006

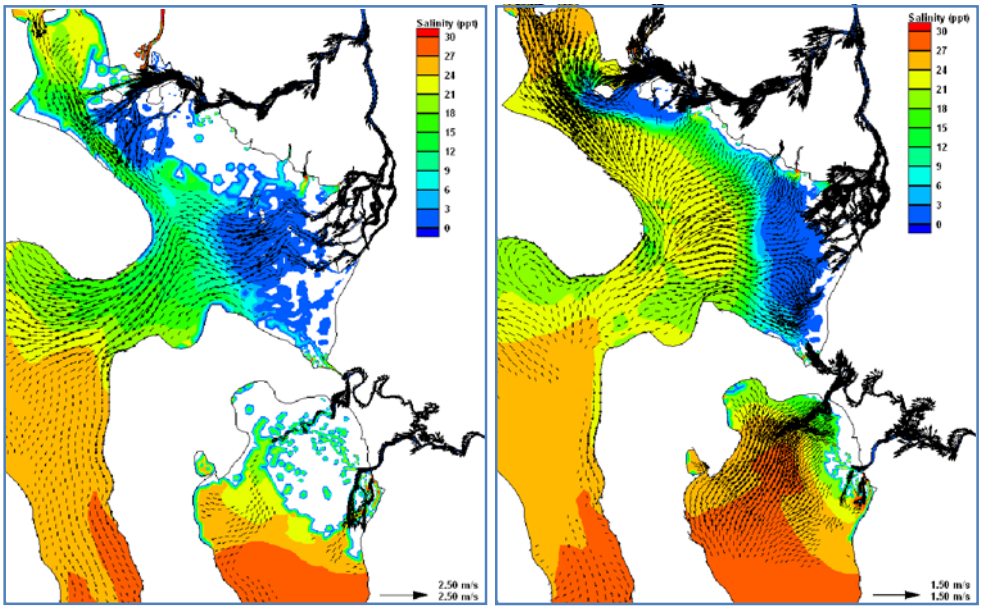

Fig. 14 Surface salinity and velocity distributions in Skagit and Port Susan Bays of Whidbey Basin in spring tide: ebb tide (*left*) and flood tide (*right*)

5 Summary

A multi-scale estuarine and coastal ocean model has been implemented for Puget Sound and the Northwest Straits using the unstructured-grid model FVCOM. Simulations of tides, estuarine circulation, temperature, and salinity in the entire Puget Sound as well as river plumes and wetting and drying processes in tide flat estuaries were conducted in a single modeling framework, which has not been done before. Comparisons of model results to observed data for tidal elevation, current, temperature, and salinity demonstrate that the model is capable of simulating the multi-scale processes in Puget Sound and the Straits. At a larger scale, the model reproduced the tidal characteristics from diurnal dominant mixed tides in the Straits that are transformed to semi-diurnal dominant mixed tides inside Puget Sound. The model also reproduced the amplification of tidal magnitudes and inequalities as tides propagate from the Pacific Coast to Puget Sound.

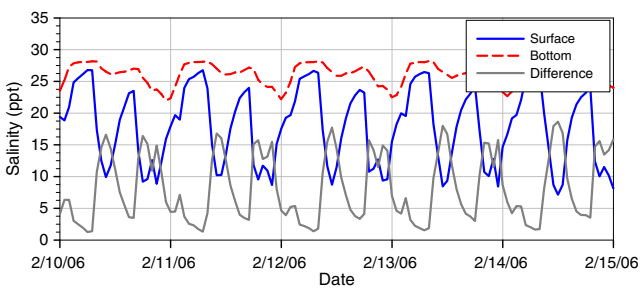

Fig. 15 Simulated surface and bottom salinities at intertidal zone of Skagit Bay

Although tidal forcing dominates the short-term circulation in Puget Sound, tidally averaged mean circulation plays a bigger role in terms of long-term exchange and renewal of Puget Sound water with the Pacific coastal ocean water. Tidally averaged mean temperature and salinity as well as mean longitudinal velocity distributions that showed well-defined two-layer circulation patterns in the Main Basin have been presented.

The geometric setting of an estuary and the magnitude of river inflow also have a significant influence on the circulation patterns and salinity distributions, for example, the presence of large tide flats near the mouths of estuaries in Whidbey Island play a role in tidal mixing and the interactions between the freshwater plume and the saline water from the Main Basin of Puget Sound. The model is also capable of simulating the wetting and drying processes on tide flats over a tidal cycle.

While the results from the Puget Sound model are encouraging because of good performance in simulating fundamental processes at various scales in Puget Sound, the development of the model is still a work-in-progress, considering the complexity of this fjordal estuarine system. The development of a robust, accurate, and comprehensive circulation and transport model of Puget Sound at multi-scales is an ongoing long-term effort. Future improvements of the Puget Sound model and its application to investigation of scientific problems may include (1) utilization of more comprehensive open-boundary conditions for interactions with the continental shelf, (2) inclusion of more observations for skill assessment, (3) employment of spatially variable wind forcing and heat flux, (4) sensitivity analysis to explore the system-wide response to seasonal and annual variability of forcing mechanisms (meteorological input, river runoff, open-

boundary conditions) and effects of climate change (sea-level rise and global warming).

Acknowledgements This study is primarily funded by the Laboratory Directed Research and Development grants of the Pacific Northwest National Laboratory and subsequently partially supported by various funding sources, including a grant from the U.S. Department of Energy as part of the Energy Efficiency and Renewable Energy program. The authors thank Dr. Taiping Wang for assisting with harmonic analysis and error assessment of the model results and figures.

References

Aoki K, Isobe A (2007) Application of finite volume coastal ocean model to hindcasting the wind-induced sea-level variation in Fukuoka bay. J Oceanogr 63(2):333–339

Babson AL, Kawase M, MacCready P (2006) Seasonal and interannual variability in the circulation of Puget Sound, Washington: a box model. Atmos Ocean 44(1):29–45

Barnes CA, Collias EE (1958) Some considerations of oxygen utilization rates in Puget Sound. J Mar Res 17:68–80

Barnes CA, Ebbesmeyer CC (1978) Some aspects of Puget Sound's circulation and water properties. In: Bjorn Kjerfve (ed) In Estuarine Transport Processes. University of South Carolina Press, Columbia South Carolina, pp 209–228

Blumberg AF, Mellor GL (1987) A description of a 3-D coastal ocean circulation model. In: Heaps N (ed) Three-Dimensional Coastal Ocean Models. American Geophysical Union, Washington, DC, pp 1–16

Bretschneider DE, Cannon GA, Holbrook JR, Pashinski DJ (1985) Variability of subtidal current structure in a fjord estuary: Puget Sound, Washington. J Geophys Res 90(C6):11949–11958. doi:10.1029/JC090iC06p11949

Cannon GA (1983). An overview of circulation in the Puget Sound estuarine system. NOAA Technical Memorandum ERL PMEL-48. NOAA, Washington, DC, 30 pp

Cannon AC, Ebbesmeyer CC (1978) Winter replacement of bottom water in Puget Sound. In: Kjerfve B (ed) Estuarine Transport Processes. University of South Carolina Press, Columbia, pp 229–238

Chen C, Rawson M (2005) An ecosystem management model system for the Satilla River Estuary, Georgia. Oceans, 2005. Proc MTS IEEE 1:622–632

Chen C, Liu H, Beardsley RC (2003) An unstructured, finite-volume, three-dimensional, primitive equation ocean model: application to coastal ocean and estuaries. J Atm Oceanic Tech 20:159–186

Chen C, Beardsley RC, Cowles G (2006) An unstructured grid, finite-volume coastal ocean model (FVCOM) system. Special Issue entitled "Advance in Computational Oceanography". Oceanography 19(1):78–89

Chen C, Gao G, Qi J, Proshutinsky A, Beardsley RC, Kowalik Z, Lin H, Cowles G (2009) A new high-resolution unstructured-grid finite-volume Arctic Ocean model (AO-FVCOM): an application for tidal studies. J Geophys Res. doi:10.1029/2008jc004941

Cokelet ED, Stewart RJ, Ebbesmeyer CC (1990). The annual mean transport in Puget Sound. NOAA Technical Memorandum ERL PMEL-92. NOAA, Washington, DC

Collias E (1972). The salt water system of Washington State. Washington State Department Ecology Report. Washington State Department Ecology, Olympia, 24 pp

County K (2001) Features of Puget Sound region: oceanography and physical processes. Chapter 3 in the State of the Nearshore Report. Department of Natural Resources, Seattle

Crawford WR, Cherniawsky JY, Cummins PF (1999) Surface currents in British Columbia waters: comparison of observations and model predictions. Atmos Ocean 37:255–280

Crean PB (1978) A numerical model of barotropic mixed tides between Vancouver Island and the mainland and its relation to studies of the estuarine circulation. In: Nihoul JC (ed) Hydrodynamics of estuaries and fjords. Elsevier, Amsterdam, pp 283–313

Cummins PF, Oey LY (1997) Simulation of barotropic and baroclinic tides off northern British Columbia. J Phys Oceanogr 27:762–781

Cummins PF, Masson D, Foreman MGG (2000) Stratification and mean flow effects on diurnal tidal currents off Vancouver Island. J Phys Oceanogr 30:15–30

de Brye B, de Brauwere A, Gourgue O, Kärnä T, Lambrechts J, Comblen R, Deleersnijder E (2010) A finite-element, multi-scale model of the Scheldt tributaries, river, estuary and ROFI. J Coast Eng 57:850–863

Ebbesmeyer CC, Barnes CA (1980) Control of a fjord basin's dynamics by tidal mixing in embracing sill zones. Estuarine and Coastal Mar Sci 11:311–330

Ebbesmeyer CC, Coomes CA (1989) Strong low frequency decadal environmental fluctuations during the 20th Century in the North Pacific Ocean, on the Washington Coast, and in Puget Sound. Oceans 89 Proceedings 1:242–246

Ebbesmeyer CC, Cannon GA, Nairn BJ, Kawase M, Fox B (2002) Puget Sound physical oceanography related to the Triple Junction region. Prepared for King County Department of Natural Resources and Parks. King County Department of Natural Resources and Parks, Seattle

Foreman M, Henry R, Walters R, Ballantyne V (1993) A finite element model for tides and resonance along the north coast of British Columbia. J Geophys Res 98(C2):2509–2531

Hickey BM, Banas NS (2003) Oceanography of the U.S. Pacific Northwest coastal ocean and estuaries with application to coastal ecology. Estuaries 26(4B):1010–1031

Huang HC, Chen JO, Blanton FA, Andreade (2008) Numerical study of tidal asymmetry in the Okatee Creek, South Carolina, Estuaries. Coastal and Shelf Science 78:190–202

Isobe A, Beardsley RC (2006) An estimate of the cross-frontal transport at the shelf break of the East China Sea with the Finite Volume Coastal Ocean Model. J Geophys Res 111:C03012. doi:10.1029/2005JC003290

Ji R, Davis C, Chen C, Beardsley R (2008) Influence of local and external processes on the annual nitrogen cycle and primary productivity on Georges Bank: a 3-D biological-physical modeling study. J Mar Syst 73:31–47

Ji R, Davis C, Chen C, Beardsley RC (2009) Life history traits and spatiotemporal distributional patterns of copepod populations in the Gulf of Maine-Georges Bank region. Mar Ecol Prog Ser 384:187–205. doi:10.3354/meps08032

Large WG, Pond S (1981) Open ocean momentum flux measurements in moderate to strong winds. J Phys Oceanogr 11:324–336

Lavelle JW, Mofjeld HD, Lempriere-Doggett E, Cannon GA, Pashinski DJ (1988). A multiply-connected channel model of tides and tidal currents in Puget Sound, Washington and a comparison with updated observations. NOAA Technical Memorandum ERL PMEL-84. NOAA, Washington, DC, 103 pp

Lavelle JW, Cokelet ED, Cannon GA (1991) A model study of density intrusions into and circulation within a deep, silled estuary—Puget Sound. J Geophys Res Oceans 96 (C9):16779–16800

Li C, Chen C, Guadagnoli G, Georgiou IY (2008) Geometry induced residual eddies in estuaries with curved channel-observations and modeling studies. J Geophys Res 113:C01005. doi:10.1029/2006JC004031

Logutov OG, Lermusiaux PFJ (2008) Inverse barotropic tidal estimation for regional ocean applications. Ocean Model 25:17–34. doi:10.1016/j.ocemod.2008.06.004

Masson D, Cummins PF (2004) Observations and modeling of seasonal variability in the Straits of Georgia and Juan de Fuca. J Mar Res 62:491–516

Masson D, Cummins PF (2000) Fortnightly modulation of the estuarine circulation in Juan de Fuca Strait. J Mar Res 58:439–463

Mellor GL, Blumberg A (2004) Wave breaking and ocean surface layer thermal response. J Phys Oceanogr 34:693–698

Mellor GL, Yamada T (1982) Development of a turbulence closure model for geophysical fluid problems. Rev Geophys Space Phys 20:851–875

Mofjeld HO, Larsen LR (1984). Tides and tidal currents in the inland waters of Western Washington. NOAA Technical Memorandum ERL PMEL-56. NOAA, Washington, DC, 52 pp

Moore SK, Mantua NJ, Newton JA, Kawase M, Warner MJ, Kellogg JP (2008a) A descriptive analysis of temporal and spatial patterns of variability in Puget Sound oceanographic properties. Estuar Coast Shelf Sci 80:545–554

Moore SK, Mantua NJ, Kellogg JP, Newton JA (2008b) Local and large-scale climate forcing of Puget Sound oceanographic properties on seasonal to interdecadal timescales. Limnol Oceanogr 53:1746–1758

Nairn BJ, Kawase M (2002). Comparison of observed circulation patterns and numerical model predictions in Puget Sound, WA. In: Droscher T (ed) Proceedings of the 2001 Puget Sound Research Conference, Puget Sound Water Quality Action Team, Olympia, Washington

Pond S, Pickard GL (1983) Introduction to Dynamical Oceanography, 2nd edn. Butterworth-Heinemann Ltd, Woburn

Public Utility District No. 1 of Snohomish County (2007). Tidal current Survey, Admiralty Inlet of Puget Sound, WA. Snohomish County PUD, Everett

Qi JC, Chen RC, Beardsley WP, Cowles G (2009) An unstructured-grid finite-volume surface wave model (FVCOM-SWAVE): implementation, validations and applications. Ocean Model 28:153–166. doi:10.1016/j.ocemod.2009.01.007

Smagorinsky J (1963) General circulation experiments with the primitive equations. I. The basic experiment. Mon Weather Rev 91:99–164

Sternberg RW (1968) Friction factors in tidal channels with differing bed roughness. Mar Geol 6:243–260

Tian R, Chen C, Stokesbury KDE, Rothschild BJ, Xu Q, Hu S, Cowles G, Harris BP, Marino MC II (2009) Modeling exploration of the connectivity between sea scallop populations in the Middle Atlantic Bight and over Georges Bank. Mar Ecol Prog Ser 380:147–160. doi:10.3354/meps07916

U. S. Geological Survey (1979). Feature detail report for Puget Sound. Report ID 1507653. U. S. Geological Survey, Reston

U. S. Geological Survey (2007). Nitrogen to Hood Canal and Lynch Cove, Washington. Scientific Investigations Report 2006-5106. Version 1.10, August 2007. U. S. Geological Survey, Reston, 104 pp

Wang B, Fringer OB, Giddings SN, Fong DA (2009) High-resolution simulations of a macrotidal estuary using SUNTANS. Ocean Model 26:60–85

Washington Department of Ecology (2008). South Puget Sound Dissolved Oxygen Study, Interim Data Report. Publication No. 08-03-037, December 2008. Washington Department of Ecology, Olympia, 167 pp

Weisberg RH, Zheng L (2006a) Circulation of Tampa Bay driven by buoyancy, tides, and winds, as simulated using a finite volume coastal ocean model. J Geophys Res 111:C01005. doi:10.1029/2005JC003067

Weisberg RH, Zheng L (2006b) Hurricane storm surge simulations for Tampa Bay. Estuaries Coasts 29(6A):899–913

Xue P, Chen C, Ding P, Beardsley RC, Lin H, Ge J, Kong Y (2009) Saltwater intrusion into the Changjiang River: a model-guided mechanism study. J Geophys Res 114:C02006. doi:10.1029/2008JC004831

Yang Z, Khangaonkar T (2009) Modeling tidal circulation and stratification in Skagit River estuary using an unstructured grid ocean model. Ocean Model 28:34–49. doi:10.1016/j.ocemod.2008.07.004

Yang Z, Khangaonkar T, Calvi M, Nelson K (2010a) Simulation of cumulative effects of nearshore restoration projects on estuarine hydrodynamics. Ecol Model 221:969–977. doi:10.1016/j.ecolmodel.2008.12.006

Yang Z, Sobocinski KL, Heatwole D, Khangaonkar T, Thom R, Fuller R (2010b) Hydrodynamic and ecological assessment of nearshore restoration: a modeling study. Ecol Model 221:1043–1053. doi:10.1016/j.ecolmodel.2009.07.011

Yang Z, Khangaonkar T, Wang T (2011). Use of advanced meteorological model output for coastal ocean modeling in Puget Sound. Int J of Climate and Ocean Sys (in press)

Zhao L, Chen C, Cowles G (2006) Tidal flushing and eddy formation in Mount Hope Bay and Narragansett Bay: an application of FVCOM. J Geophys Res 111:C10015. doi:10.1029/2005JC003135

Zheng L, Chen C, Liu H (2003) A modeling study of the Satilla River Estuary, Georgia. Part I: flooding/drying process and water exchange over the salt marsh-estuary-shelf complex. Estuaries 26(3):651–669