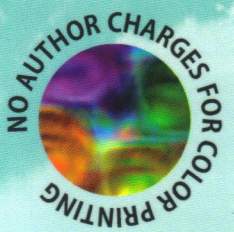
Volume 10 Nos. 1–2 April 2010 ISSN 1567-7419

Environmental Fluid Mechanics

Special Issue: Tracer and Timescale Methods for Understanding Complex Geophysical and Environmental Fluid Flows

Guest Editors: Eric Deleersnijder, Fabien Cornaton, Thomas W.N. Haine, Marnik Vanclooster and Darryn W. Waugh



Springer

Table of contents of the special issue of
Environmental Fluid Mechanics
on

Tracer and Timescale Methods for Understanding Complex Geophysical and Environmental Fluid Flows

guest-edited by

Eric Deleersnijder, Fabien Cornaton, Thomas W.N. Haine, Marnik Vanclooster and Darryn W. Waugh

Huang W., X. Liu, X. Chen and M.S. Flannery, 2010, Estimating river flow effects on water ages by hydrodynamic modeling in Little Manatee River estuary, Florida, USA, *Environmental Fluid Mechanics*, 10, 197-211

Orre S., J.N. Smith, V. Alfimov and M. Bentsen, 2010, Simulating transport of ^{129}I and idealized tracers in the northern North Atlantic Ocean, *Environmental Fluid Mechanics*, 10, 213-233

Terenzi F. and T.M. Hall, 2010, Idealized tracer transport models with time-varying transport: applications to ocean boundary currents, *Environmental Fluid Mechanics*, 10, 235-255

Dutay J.-C., J. Emile-Geay, D. Iudicone, P. Jean-Baptiste, G. Madec and C. Carouge, 2010, Helium isotopic constraints on simulated ocean circulations: implications for abyssal theories, *Environmental Fluid Mechanics*, 10, 257-273

Maltrud M., F. Bryan and S. Peacock, 2010, Boundary impulse response functions in a century-long eddying global ocean simulation, *Environmental Fluid Mechanics*, 10, 275-295

Environ Fluid Mech (2010) 10:1–5
DOI 10.1007/s10652-009-9164-1

EDITORIAL

Tracer and timescale methods for understanding complex geophysical and environmental fluid flows

Eric Deleersnijder · Fabien Cornaton · Thomas W. N. Haine · Marnik Vanclooster · Darryn W. Waugh

Received: 15 December 2009 / Accepted: 16 December 2009 / Published online: 7 January 2010

Understanding advective-diffusive transport of trace constituents in natural fluid flows is an important challenge in Earth and environmental sciences with many diverse applications, including simulating the fate of contaminants, inferring the location of their source, and model assessment (e.g., [7,21]). Eulerian and Lagrangian methods are routinely used, including novel representations of mixing processes that resort to the so-called fractional-order diffusion. Moreover, geophysical and environmental fluid-flow models routinely produce huge amounts of output, and to make sense of these results sophisticated interpretation methods are required. Among these methods, an approach that is becoming progressively more popular consists in using real, or hypothetical, tracers to tag fluid masses and estimate associated timescales, such as age, residence time, and transit time. These timescales lead to very useful diagnoses that are increasingly applied in interdisciplinary environmental studies (e.g. [2,5]).

This special issue presents a number of studies that are relevant to the above-mentioned field of research. Groundwater, soil water, riverine, estuarine, marine and ocean flows are considered, as well as the transport of sinking particles in water.

E. Deleersnijder (✉)
Centre for Systems Engineering and Applied Mechanics (CESAME),
Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium
e-mail: eric.deleersnijder@uclouvain.be

F. Cornaton
Centre of Hydrogeology, Université de Neuchâtel, Neuchâtel, Switzerland
e-mail: fabien.cornaton@unine.ch

T. W. N. Haine · D. W. Waugh
Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, USA
e-mail: thomas.haine@jhu.edu

D. W. Waugh
e-mail: waugh@jhu.edu

M. Vanclooster
Department of Environmental Sciences and Land Use Planning, Université catholique de Louvain (UCL),
Louvain-la-Neuve, Belgium
e-mail: marnik.vanclooster@uclouvain.be

The article of Hanert [12] presents a family of non-local dispersion models that do not assume Brownian motion but rather assume the random displacements of the particles follow a heavy-tail Levy distribution, which allows for large displacements. It is seen that such a hypothesis leads to fractional-order Eulerian transport models, which cannot be discretised efficiently by standard finite-difference and finite-element methods. Instead, non-local numerical methods, like the spectral-element method, are shown to be better suited to solve that equation.

Gräwe and Wolff [9] present a Lagrangian particle-tracking model to simulate short-term suspended particulate matter transport. It is well known that Lagrangian transport schemes are quite attractive for such applications, mostly because they give a straightforward physical interpretation of the processes involved and are well suited to massive parallel computers. The authors present three different tracking schemes and apply them to a realistic winter storm event. They show that particle-tracking schemes can accurately reproduce the dynamics provided that the divergence of the diffusivity tensor is handled carefully.

The paper of Ababou et al. [1] deals with the problem of identifying contaminant sources given certain solute distributions in environmental flows. The authors present a novel, stable, and convergent particle-tracking scheme (RAW: Reverse Anti-diffusive Walk) able to back-diffuse concentration plumes, that is, diffusion backwards-in-time. This study is an interesting, theoretical exploration that extends the authors' previous works on Lagrangian particle methods to the challenging task of back-tracking a diffusion process in time in order to identify pollution sources. As stated by the authors the proposed numerical schemes are limited to known isotropic diffusion and, therefore, still need to be generalized to tackle real-world multiple-source identification problems involving heterogeneous transport parameters.

Tosco and Sethi [20] present a comparison between two very different computational approaches for the delineation of well-head protection zone areas in porous media, an important practical problem in hydrogeology. Deterministic well-capture zones (purely advective) are defined using the widely used reverse particle-tracking approach. Probabilistic well-capture zones (advective and dispersive) are estimated using a backward probability model derived from the adjoint of the advection-dispersion equation. Both types of capture zone areas are compared using simplified synthetic 2-D aquifer configurations. They can display important differences in shape and extent, thus showing the importance of including the dispersion/diffusion phenomena in calculations addressing the delineation of well-head protection zones.

In Lemieux and Sudicky [14], the concept of groundwater age is applied to study the impact of the Wisconsinian glaciation on the Canadian continental groundwater flow systems. Large-scale coupled surface-subsurface water flow and glaciation solutions originating from the authors' previous works are exploited by solving for the transient evolution of groundwater mean age. The authors show that mean age in regions affected by the ice sheet advance and retreat is younger than it was at the last interglacial period. This result also holds for frozen groundwater in the permafrost area, which suggests that significant parts of this water is of glacial origin. Estimated ages (up to 42 Myr) are in good agreement with the observations made in the Great Lake region (up to 50 kyr). Meltwater penetration-depth estimates (up to 3 km) are also in good agreement with the observations made in the Illinois, Michigan and Williston basins, and in the Canadian Shield. This paper is important as a means of understanding the large-scale impact of an ice sheet on groundwater systems.

In Darracq et al. [3], the distributions of advective solute travel times through hydrological catchments is studied for two hydrological catchments in Sweden. The travel-time distributions are essential for understanding how local water flow and solute transport and attenuation processes affect the catchment-scale transport of solute, for instance, with regard to

biogeochemical cycling, contamination persistence, and contamination load. Results show that in these hydrological catchments travel times are largely determined by the groundwater component and hence the physical properties describing flow through the subsurface groundwater body such as the hydraulic gradient, porosity, and hydraulic conductivity. Different hypotheses on the spatial variability of these properties have a considerable effect on catchment travel time.

In Mattern and Vanclooster [16], the authors estimate the travel time of percolation water through a deep vadose zone of the Brusselian aquifer by combining a transfer-function model and a physically-based model, and considering time-variable percolation rates. A log-normal probability-density function is chosen for the transfer-function model, in accordance with a porous medium exhibiting a log-normal distribution of pore velocities. The Hydrus-1D model is applied to numerically solve for the 1-D Richards and advective-dispersive solute transport equations. The important discrepancies between the results from the two types of models are explained by the conceptual differences between the two modelling approaches and the uncertainty associated to the parameterization of the physically-based flow and transport models.

Gualtieri [10] examines the values of the transverse diffusion coefficient for depth-integrated simulations of tracer transport in river flows. Numerical results obtained by means of the classical k-ε closure scheme compare rather favourably with experimental data. This finding is of importance for a wide range of tracer transport simulations as well as for estimating the associated diagnostic timescales.

In de Brauwere and Deleersnijder [4], the concept of residence time is seen to be very efficient to assess depth-integrated models of the fate of decaying and sinking particles. A comprehensive sensitivity analysis is carried out with, and without, stratification.

Seeking inspiration in Mercier and Delhez [17], Gong and Shen [8] estimate the age of sediment in the water column and at the seabed in an estuary, where it helps diagnose transport of material adsorbed on sediment particles. Not surprisingly, the magnitude of the settling velocity and the critical shear stress are seen to impact critically the timescales obtained. These timescales are useful for a range of problems, including the prediction of eutrophication and the fate of contaminants.

Huang et al. [13] estimate water ages in an estuary from the travel-time median of dye released at the entrance of the estuary. At several locations in the estuary this age is seen to be well-correlated with an empirical power law of the river flow. It is underscored that water age is an important indicator of estuarine water quality and, hence, is important in ecosystem assessment.

Orre et al. [18] discuss the transport of ^{129}I in the northern North Atlantic Ocean. In simulations using a global ocean circulation model, the authors find reasonable agreement with field measurements of the isotope. They then study the transport characteristics of their model flow using transit-time distributions, and discuss propagation of other passive tracer anomalies through the sub-polar Atlantic and Nordic Seas.

In Terenzi and Hall [19], new solutions are derived for one-dimensional advection-dilution and advection-diffusion models with non-constant coefficients. These solutions are used to compare tracer fluctuations due to temporal variations in the transport with those due to temporal variations in the boundary conditions. When applied to CFC observations in the North Atlantic Deep Western Boundary Current (DWBC) it is shown that fluctuations resulting from boundary conditions dominate near the source region, while transport fluctuations acting on background tracer gradients dominate away from the source. This contrasts a common view that fluctuations in tracers, such as CFCs, temperature, and salinity, in the subtropical and tropical DWBC are propagated signals from the northern formation region.

Dutay et al. [6] gauge the impact of geothermal heating of the sea floor and abyssal vertical mixing on an ocean general circulation model, evaluating their results with ^{3}He data. They find that in order to produce realistic ^{3}He simulations, both geothermal heating and enhanced vertical mixing are necessary.

Maltrud et al. [15] present results from simulations of age-related passive tracers in a 1/10° resolution global ocean model. This is the first time age tracers have been analyzed in a global model that resolves eddying motions. A century-long simulation is performed including five global Boundary Impulse Response functions (BIRs), which are used to characterize the related Transit Time Distribution functions (TTDs), and hence the variability in water-mass ventilation timescale in the presence of eddies. The authors show that variability is largest at short timescales, but at longer timescales there is a remarkable conformity between the ensemble of BIR realisations. Furthermore, variations in the BIR first moment over transit time compared to the corresponding ensemble average are typically only a few percent. Invoking the study of Haine et al. [11], these results imply that only a few BIR realisations may be necessary for a reasonable estimate of the TTD.

All of these papers testify to the wide range of applications and the usefulness of tracer and timescale methods. Clearly, this field of research is alive and well. The field has not yet reached full maturity, however, and can be confusing to the newcomer. In particular, similar concepts and equations are sometimes given different names by different authors — even for flows in the same medium. For oceanography alone, Wunsch [22] deplored this situation back in 2002 and, unfortunately, there has been no sign of improvement since then. Attempts towards unification should be made, which, hopefully, would lead to a coherent set of definitions, equations and notations that would be applicable to most, if not all, geophysical and environmental fluid flow problems.

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). His work was performed under the auspices of the Interuniversity Attraction Pole TIMOTHY (www.climate.be/timothy), which is funded by the Belgian Science Policy office (BELSPO) under contract IAP6.13.

References

1. Ababou R, Bagtzoglou AC, Mallet A (this issue) Anti-diffusion and source identification with the 'RAW' scheme: a particle-based censored random walk. Environm Fluid Mech
2. Cornaton F, Perrochet P (2006) Groundwater age, life expectancy and transit time distributions in advective-dispersive systems: 1 Generalized reservoir theory. Adv Water Resour 29:1267–1291
3. Darracq A, Destouni G, Persson K, Prieto C, Jarsjö J (this issue) Quantification of advective solute travel times and mass transport through hydrological catchments. Environ Fluid Mech
4. de Brauwere A, Deleersnijder E (this issue) Assessing the parameterisation of the settling flux in a depth-integrated model of the fate of decaying and sinking particles, with application to fecal bacteria in the Scheldt Estuary. Environ Fluid Mech
5. Deleersnijder E, Delhez EJM (eds) (2007) Timescale- and tracer-based methods for understanding the results of complex marine models. Estuar Coast Shelf Sci (Special Issue) 74:585–776
6. Dutay J-C, Emile-Geay J, Iudicone D, Jean-Baptiste P, Madec G, Carouge C (this issue) Helium isotopic constraints on simulated ocean circulations—implications for abyssal theories. Environ Fluid Mech
7. England MH, Maier-Reimer E (2001) Using chemical tracers to assess ocean models. Rev Geophys 39:29–70
8. Gong W, Shen J (this issue) A model diagnostic study of age of river-borne sediment transport in the tidal York River Estuary. Environ Fluid Mech
9. Gräwe U, Wolff J-O (this issue) Suspended particulate matter dynamics in a particle framework. Environ Fluid Mech

10. Gualtieri C (this issue) RANS-based simulation of transverse turbulent mixing in a 2D geometry. Environ Fluid Mech
11. Haine TWN, Zhang H, Waugh DW, Holzer M (2008) On transit-time distributions in unsteady circulation models. Ocean Model 21:35–45
12. Hanert E (this issue) A comparison of three Eulerian numerical methods for fractional-order transport models. Environ Fluid Mech
13. Huang W, Liu X, Chen X, Flannery MS (this issue) Estimating river flow effects on water ages by hydrodynamic modeling in Little Manatee River estuary, Florida, USA. Environ Fluid Mech
14. Lemieux J-M, Sudicky EA (this issue) Simulation of groundwater age evolution during the Wisconsinian glaciation over the Canadian landscape. Environ Fluid Mech
15. Maltrud M, Bryan F, Peacock S (this issue) Boundary impulse response functions in a century-long eddying global ocean simulation. Environ Fluid Mech
16. Mattern S, Vanclooster M (this issue) Estimating travel time of recharge water through a deep vadose zone using a transfer function model. Environ Fluid Mech
17. Mercier C, Delhez EJM (2007) Diagnosis of sediment transport in the Belgian Coastal Zone. Estuar Coast Shelf Sci 74:670–683
18. Orre S, Smith JN, Alfimov V, Bentsen M (this issue) Simulating transport of and idealised tracers in the northern North Atlantic Ocean. Environ Fluid Mech
19. Terenzi F, Hall TM (this issue) Idealized tracer transport models with time-varying transport: applications to ocean boundary currents. Environ Fluid Mech
20. Tosco T, Sethi R (this issue) Comparison between backward probability and particle tracking methods for the delineation of well head protection areas. Environ Fluid Mech
21. Waugh DW, Hall TM (2002) Age of stratospheric air: theory, observations, and models. Rev Geophys 40(4):1010. doi:10.1029/2000RG000101
22. Wunsch C (2002) Oceanic age and transient tracers: analytical and numerical solutions. J Geophys Res 107:3048. doi:10.1029/2001JC000797

Environ Fluid Mech (2010) 10:7–20
DOI 10.1007/s10652-009-9145-4

ORIGINAL ARTICLE

A comparison of three Eulerian numerical methods for fractional-order transport models

Emmanuel Hanert

Received: 12 March 2009 / Accepted: 16 July 2009 / Published online: 1 August 2009

Abstract Tracer transport in complex systems like turbulent flows or heterogeneous porous media is now more and more regarded as a non-local process that can hardly be represented by second-order diffusion models. In this work, we consider diffusion models that assume that tracer particles follow a heavy-tail Lévy distribution, which allows for large displacements. We show that such an assumption leads to a fractional-order diffusion operator in the governing equation for tracer concentration. A comparison of three Eulerian numerical methods to discretize that equation is then performed. These consist of the finite difference, finite element and spectral element methods. We suggest that non-local methods, like the spectral element method, are better suited to transport models with fractional-order diffusion operators.

Keywords Tracer transport · Fractional calculus · Lévy distributions · Spectral element method

1 Introduction

For many environmental problems, it is important to be able to describe the transport dynamics without having recourse to a direct numerical simulation of all the processes taking place in the system, which often requires excessive computational resources. For efficient and practical predictions, simpler and more manageable models should be sought in order to capture the phenomena of interest without having to calculate all the details of the complex system. Examples include moisture, pollutants and chemical components in the atmosphere; salinity and temperature in the ocean; sediments in rivers and lakes, contaminants and nutriments in heterogeneous porous media, etc.

E. Hanert (✉)
Department of Environmental Sciences and Land Use Planning, Université Catholique de Louvain, Place Croix du Sud 2/16, 1348 Louvain-la-Neuve, Belgium
e-mail: emmanuel.hanert@uclouvain.be

Dispersion processes in complex systems are generally described by a second-order advection–diffusion equation (ADE). The assumption underlying this model is that turbulent dispersion, like molecular diffusion, can be described by Fick's law, which states that particle flux is directly proportional to the spatial concentration gradient. Fick's law therefore assumes that the spatial concentration gradient is causing particles movement in a turbulent flow. This assumption is contradicted by observations that suggest that particles in a turbulent flow do not "push" each other as it is the case for Brownian motion but are rather dispersed by the velocity fluctuations. A number of studies have shown that Fickian diffusion was unsatisfactory to simulate tracer transport in environmental flows (see for instance [1,24,27,28,31]). All these studies have highlighted the dependence of the eddy viscosity on the size of the dispersion cloud. Different Lagrangian [10,12,19] and Eulerian [4] modelling approaches have been proposed to account for the observed non-Fickian dispersion patterns. The limitations of eddy diffusivity models are further discussed in a recent paper by Cushman-Roisin [8].

Despite the apparent shortcomings of the second-order diffusion operator to represent dispersion in complex systems, it is still used in the majority of today's models of environmental flows. However, a number of recent studies advocate the use of non-local and scale-dependent dispersion models to simulate the transport of tracers in complex systems. Such models are usually based on fractional-order diffusion operators (see for instance [2,7,22,30]). They have been used to model tracer transport in systems such as saturated and unsaturated soil layers [5,16,18,25], river flows [11,20] or the atmospheric boundary layer [9]. For these applications, the use of fractional temporal models is sometimes required as well [33].

Although non-local tracer transport models are now an active field of research, they are still in their infancy and many questions remain. One of them concerns the numerical methods that should be used to efficiently discretize the fractional-order diffusion operator. If such models are to be used operationally to simulate complex environmental systems, they have to achieve good accuracy at the same computational cost as traditional models. In this work, we compare three different Eulerian numerical methods: the finite difference, finite element and spectral element methods.

2 Theory

In laminar homogeneous flows, the dispersion of tracer particles is mainly driven by the mean flow velocity and by local interactions between particles that result in regular and isotropic random displacements. These random displacements constitute a Brownian motion. Since all these random fluctuations can be assumed to be independent and identically distributed, and the variance of their sum can be assumed to be finite, then de Moivre's Central Limit Theorem (CLT) indicates that the sum of all the random displacements will follow a Gaussian probability distribution function whose center moves with the fluid mean velocity. In one dimension, this probability distribution function is a solution of the following second-order ADE:

$$\frac{\partial c}{\partial t} + v\frac{\partial c}{\partial x} = K\frac{\partial^2 c}{\partial x^2}. \tag{1}$$

The concentration c can be seen as the probability distribution function of a large number of tracer particles that experience a deterministic drift of velocity v and a normal random fluctuation with standard deviation equal to $\sqrt{2Kt}$, where K is a macroscopic diffusion coefficient. Einstein [13] first made the link between the second-order ADE and Brownian

motion. It should be noted that Eq. 1 can also be obtained by using Fick's parameterization of the concentration flux.

In turbulent heterogeneous flows, particles can be transported over large distances by the fluctuations of the flow velocity. In that case, although local interactions are still possible, longer and non-local interactions should also be taken into account. In addition to that one should also allow for non-isotropic random displacements than can arise from the heterogeneous nature of the system. The resulting ensemble of random displacements, in the limit where the ensemble size tends to infinity, does not necessarily have a finite variance. The standard version of the CLT can therefore not be applied. Instead, the generalization due to Lévy–Gnedenko [17,21] can be used. That version of the theorem does not rely on the assumption that the sum of all the fluctuations has a finite variance but instead assumes that it has a power-law tail distribution decreasing as $|x|^{-\alpha-1}$ with $0 < \alpha \leq 2$. If we further assume that the probability of particle jumps to the right is $\frac{1-\beta}{2}$ and jumps to the left is $\frac{1+\beta}{2}$ with $-1 \leq \beta \leq 1$, then the sum of all these fluctuations will tend to a stable Lévy distribution with exponent α and skewness parameter β [14]. Such a distribution does not usually have an explicit expression but can be expressed in terms of its Fourier transform. If we define the Fourier and inverse Fourier transforms as:

$$F(\omega) = \mathcal{F}(f) = \int_{-\infty}^{\infty} f(x) e^{-i\omega x}\, \mathrm{d}x,$$

$$f(x) = \mathcal{F}^{-1}(F) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega x}\, \mathrm{d}\omega,$$

we can then define the Lévy distribution as follows:

$$S_\alpha(\beta, \gamma, \delta; x) \equiv \begin{cases} \mathcal{F}^{-1}\left\{\exp\left(i\delta\omega - \gamma^\alpha |\omega|^\alpha \left(1 - i\beta \mathrm{sgn}(\omega) \tan\frac{\alpha\pi}{2}\right)\right)\right\} & \alpha \neq 1, \\ \mathcal{F}^{-1}\left\{\exp\left(i\delta\omega - \gamma^\alpha |\omega|^\alpha \left(1 + i\beta \mathrm{sgn}(\omega) \frac{2}{\pi} \log|\omega|\right)\right)\right\} & \alpha = 1, \end{cases} \tag{2}$$

where the real parameters $\gamma \in [0, \infty)$ and $\delta \in (-\infty, \infty)$ are called the scale and the location of the distribution. The former is a measure of the width of the distribution while the latter indicates the position of the center of the distribution. These parameters are similar to the mean (μ) and variance (σ^2) but the latter cannot be defined when $\alpha < 2$. Indeed, the heavy-tail behavior of stable distributions leads to an infinite variance and undefined mean for all $\alpha < 2$. Although Lévy distributions do not generally have an analytical expression, there are two noteworthy exceptions: (i) The Gaussian/normal distribution which corresponds to a Lévy stable distribution with exponent $\alpha = 2$ and variance $\sigma^2 = 2\gamma$ and (ii) the Cauchy distribution which is a Lévy distribution with exponent $\alpha = 1$. In these two cases, the skewness parameter β is irrelevant.

In Fig. 1, we show two Lévy distributions with exponents $\alpha = 2$ and $\alpha = 1.75$. For both cases, the other parameters have the following values: $\beta = 0$, $\gamma = 1$ and $\delta = 0$. We also show the dispersion pattern obtained when one particle is randomly displaced by fluctuations following these distributions without any deterministic drift. In both cases, the particle random displacement consists in 10^4 steps. When $\alpha = 2$, the particle motion is Brownian and is characterized by a succession of disordered displacements of similar length. When $\alpha = 1.75$, the motion is still mostly Brownian but the particle now has a non-negligible probability to jump over larger distances. As the value of α decreases, the thickness of the distribution's tails increases and hence the probability of large jumps increases as well.

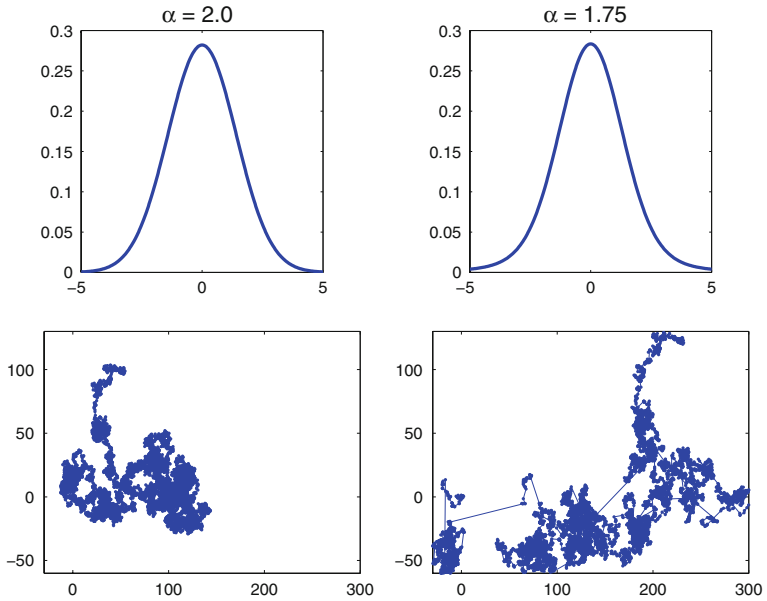


Fig. 1 *Top row*: Lévy distributions with parameters $\alpha = 2$ (*left*) and $\alpha = 1.75$ (*right*), $\beta = 0, \gamma = 1$ and $\delta = 0$. *Bottom row*: Illustration of the displacement of single particle that is randomly moved by fluctuations following these distributions. The total number of steps made by the particle is equal to 10^4. Note that the distribution's tails are thicker when $\alpha = 1.75$

Following the same approach as Benson et al. [3], we can show that the Lévy distribution is the solution of an ADE of fractional order α. Since the Fourier transform of the usual m^{th} ($m \in \mathbb{N}$) order derivative is $\mathcal{F}\{f^{(m)}\} = (i\omega)^m \mathcal{F}\{f\}$, we can define the left and right fractional-order derivatives as

$$\frac{\partial^\alpha}{\partial x^\alpha} f(x,t) = \mathcal{F}^{-1}\left\{(i\omega)^\alpha \mathcal{F}\{f\}\right\},$$
$$\frac{\partial^\alpha}{\partial(-x)^\alpha} f(x,t) = \mathcal{F}^{-1}\left\{(-i\omega)^\alpha \mathcal{F}\{f\}\right\}.$$

It can be shown (see for instance Podlubny [26] for details) that these expressions are equivalent to the Riemann–Liouville fractional derivatives defined as:

$$\frac{\partial^\alpha}{\partial x^\alpha} f(x,t) = \frac{1}{\Gamma(n-\alpha)} \frac{\partial^n}{\partial x^n} \int_{-\infty}^{x} \frac{f(y,t)}{(x-y)^{\alpha-n+1}}\, \mathrm{d}y,$$
$$\frac{\partial^\alpha}{\partial(-x)^\alpha} f(x,t) = \frac{(-1)^n}{\Gamma(n-\alpha)} \frac{\partial^n}{\partial x^n} \int_{x}^{+\infty} \frac{f(y,t)}{(y-x)^{\alpha-n+1}}\, \mathrm{d}y,$$

where $n = 1 + [\alpha]$ and $[\alpha]$ is the largest integer not greater than α, i.e., $n = 2$ for $1 < \alpha \leq 2$, and $\Gamma(.)$ is Euler's gamma function.

By using these differential operators and assuming $\alpha \neq 1$, we can define the following fractional-order ADE:

$$\frac{\partial c}{\partial t} + v\frac{\partial c}{\partial x} = K_\alpha \frac{1-\beta}{2} \frac{\partial^\alpha c}{\partial x^\alpha} + K_\alpha \frac{1+\beta}{2} \frac{\partial^\alpha c}{\partial(-x)^\alpha}. \tag{3}$$

The analytical solution of Eq. 3 can be found by taking its Fourier transform:

$$\frac{\partial C}{\partial t} + (iv\omega)C = K_\alpha \frac{1-\beta}{2}(i\omega)^\alpha C + K_\alpha \frac{1+\beta}{2}(-i\omega)^\alpha C, \tag{4}$$

where $C(\omega, t) = \mathcal{F}\{c(x,t)\}$. The solution of (4) then reads:

$$C(\omega, t) = \exp\left[-(iv\omega)t + K_\alpha \frac{1-\beta}{2}(i\omega)^\alpha t + K_\alpha \frac{1+\beta}{2}(-i\omega)^\alpha t\right]. \tag{5}$$

Using the following relations:

$$(i\omega)^\alpha = |\omega|^\alpha \left(\cos\frac{\pi\alpha}{2} + i\,\mathrm{sgn}(\omega)\sin\frac{\pi\alpha}{2}\right),$$
$$(-i\omega)^\alpha = |\omega|^\alpha \left(\cos\frac{\pi\alpha}{2} - i\,\mathrm{sgn}(\omega)\sin\frac{\pi\alpha}{2}\right),$$

we can express (5) as

$$C(\omega, t) = \exp\left[-(iv\omega)t + K_\alpha t\cos\frac{\pi\alpha}{2}|\omega|^\alpha \left(1 - i\beta\,\mathrm{sgn}(\omega)\tan\frac{\pi\alpha}{2}\right)\right],$$

and the solution of Eq. 3 thus reads

$$c(x,t) = S_\alpha\left(\beta, \left(-K_\alpha t\cos\frac{\pi\alpha}{2}\right)^{1/\alpha}, -vt; x\right).$$

It is interesting to note that the width of the distribution grows like $t^{1/\alpha}$, which clearly illustrates that fractional-order diffusion is "faster" than the usual second-order diffusion as soon as $\alpha < 2$.

3 Three Eulerian numerical methods for the fractional-order ADE

In this section, we present different numerical techniques to solve the fractional-order ADE on a finite computational domain $[0, L]$. In that case, we have to redefine the right and left fractional derivatives as follows:

$$\frac{\partial^\alpha}{\partial x^\alpha} f(x,t) = \frac{1}{\Gamma(n-\alpha)}\frac{\partial^n}{\partial x^n}\int_0^x \frac{f(y,t)}{(x-y)^{\alpha-n+1}}\,\mathrm{d}y, \tag{6}$$

$$\frac{\partial^\alpha}{\partial(-x)^\alpha} f(x,t) = \frac{(-1)^n}{\Gamma(n-\alpha)}\frac{\partial^n}{\partial x^n}\int_x^L \frac{f(y,t)}{(y-x)^{\alpha-n+1}}\,\mathrm{d}y, \tag{7}$$

to account for the restricted domain of definition of f. We consider three different Eulerian numerical methods to solve Eq. 3: the finite difference (FD), finite element (FE) and spectral element (SE) methods.

3.1 Finite difference method

A FD discretization of the fractional-order ADE can be obtained by using the so-called Grunwald fractional derivative [26] instead of the Riemann–Liouville derivatives (6) and

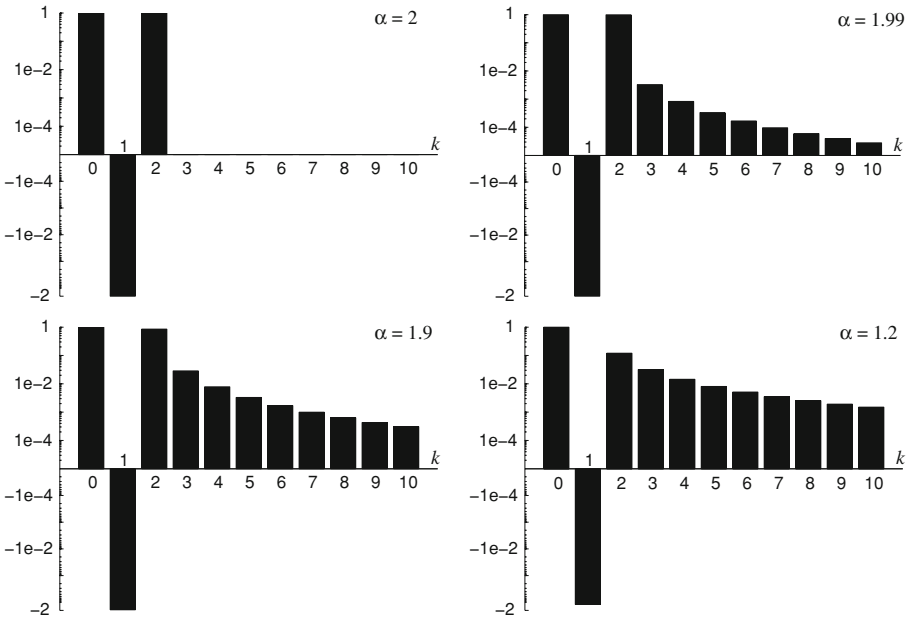


Fig. 2 First Grunwald weights w_k for different values of the exponent coefficient α and $k = 0, 1, \ldots, 10$. When $\alpha = 2$, only weights corresponding to $k = 0,\ 1$ and 2 are non-zero. As soon as $\alpha < 2$, all the other weights become positive and see their value increase as α decreases. Note that the y-axis is logarithmic and discontinuous between positive and negative values

(7). In that case, the left and right fractional derivatives are expressed as:

$$\frac{\partial^\alpha f}{\partial x^\alpha} = \lim_{M\to\infty} \frac{\sum_{k=0}^{M} w_k f(x-(k-1)\Delta x)}{\Delta x^\alpha}, \tag{8}$$

$$\frac{\partial^\alpha f}{\partial(-x)^\alpha} = \lim_{M\to\infty} \frac{\sum_{k=0}^{M} w_k f(x+(k-1)\Delta x)}{\Delta x^\alpha}, \tag{9}$$

where

$$w_k = \frac{\Gamma(k-\alpha)}{\Gamma(k+1)\Gamma(-\alpha)} \quad k = 0, 1, 2, \ldots.$$

and M is a positive integer, and $\Delta x = L/M$. The Grunwald derivative can be interpreted as a generalization to non-integer orders of the fundamental definition of derivatives in terms of quotient of differences. It can be shown that the Grunwald and Riemann-Liouville fractional-order derivatives are equivalent [26]. Note that in (8)–(9), a shifted Grunwald formula has been used for stability reasons [23].

The coefficients w_k, also called Grunwald weights, illustrate the non-locality of the fractional-order derivatives as soon as $\alpha < 2$. As shown in Fig. 2, the importance of points faraway from the point under consideration increases as α decreases. Unlike integer-order derivatives, the fractional-order differential operator is a global operator that takes into account the global behavior of the function and not just the local slope.

The expression of the Grunwald formula suggests a simple way of discretizing the fractional-order ADE in space by means of the FD method. A discrete approximation of (3) can be obtained by truncating (8)–(9) and using a finite space increment Δx. Such a FD

scheme can easily be implemented and generalized to higher dimensions. It is only first-order accurate in space but Tadjeran et al. [32] have proposed a method to improve the accuracy to second order. For instance, to solve Eq. 3 over the one-dimensional domain $[0, L]$, the following FD scheme could be used:

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} + v_i \frac{c_{i+1} - c_{i-1}}{2\Delta x} = K_\alpha \frac{1-\beta}{2} \frac{\sum_{j=0}^{i} w_j c_{i-j}}{\Delta x^\alpha} + K_\alpha \frac{1+\beta}{2} \frac{\sum_{j=0}^{N-i} w_j c_{i+j}}{\Delta x^\alpha}, \quad (10)$$

where the domain has been divided in N segments of finite size $\Delta x = \frac{L}{N}$. Equation 10 has to be combined with suitable boundary conditions and discretized in time as well.

As can be seen from (10), the FD stencil used to discretize the fractional-order derivative covers the whole domain as soon as $\alpha < 2$. This could obviously be expected as fractional-order derivatives are global operators that have more similarities with integrals than with traditional derivatives. However, since the FD method usually requires a large number of nodes to obtain a precise result, the resulting increase in the computational cost can be important. In 1D, the number of operations required to compute the fractional-order derivative at a given node will be about $N/3$ times larger than the number of operations required to compute a second-order derivative. Moreover, the use of an implicit time integration scheme seems totally prohibitive as it would require to solve a full-matrix system of equations.

3.2 Finite element method

The FE method has first been applied by Fix and Roop [15] and Roop [29] to the fractional-order ADE. Unlike the FD method, the FE method does not discretize the differential operators but rather approximates the exact solution by a linear combination of basis functions ϕ_i, which are usually piecewise polynomials defined on a partition of the computational domain. The solution of (3) is thus expressed as:

$$c(x,t) \approx c^h(x,t) = \sum_{j=1}^{N} c_j(t)\phi_j(x), \quad (11)$$

where the superscript h denotes the discrete solution and the c_j's are unknown coefficients.

Before deriving the discrete equations that will allow us to compute the expansion coefficients, we shall first rewrite the fractional derivative as follows:

$$\begin{aligned}
&\frac{1-\beta}{2}\frac{\partial^\alpha c}{\partial x^\alpha} + \frac{1+\beta}{2}\frac{\partial^\alpha c}{\partial(-x)^\alpha} \\
&= \frac{\partial^n}{\partial x^n}\left[\frac{1-\beta}{2\Gamma(n-\alpha)}\int_0^x \frac{c(y,t)}{(x-y)^{\alpha-n+1}}\,\mathrm{d}y + \frac{1+\beta}{2\Gamma(n-\alpha)}\int_x^L \frac{c(y,t)}{(y-x)^{\alpha-n+1}}\,\mathrm{d}y\right] \\
&\equiv \frac{\partial^n}{\partial x^n} G_{\alpha\beta}(x,c).
\end{aligned}$$

That allows us to express the fractional derivative of c as the integer-order derivative of a space-averaging function $G_{\alpha\beta}$, which could also be expressed in terms of the convolution of $c(x,t)$ with $x^{n-\alpha-1}$. For most environmental systems, $1 < \alpha \leq 2$ and n is thus equal to 2. In that case and provided that we are able to compute $G_{\alpha\beta}$, the discrete equations can be derived

by using exactly the same standard Galerkin formulation as for the second-order ADE:

$$\int_0^L \frac{\partial c^h}{\partial t}\phi_i \,\mathrm{d}x + \int_0^L v\frac{\partial c^h}{\partial x}\phi_i \,\mathrm{d}x = K_\alpha \int_0^L \frac{\partial^2 G_{\alpha\beta}(x,c^h)}{\partial x^2}\phi_i \,\mathrm{d}x,$$
$$= -K_\alpha \int_0^L \frac{\partial G_{\alpha\beta}(x,c^h)}{\partial x}\frac{\partial \phi_i}{\partial x}\,\mathrm{d}x + \left[K_\alpha \frac{\partial G_{\alpha\beta}(x,c^h)}{\partial x}\phi_i\right]_0^L,$$

for $i = 1, \ldots, N$. By replacing c^h by the expansion in (11), we obtain the following set of N discrete equations:

$$\sum_{j=1}^N \left(\int_0^L \phi_i\phi_j \,\mathrm{d}x\right)\frac{\mathrm{d}c_j}{\mathrm{d}t} + \sum_{j=1}^N \left(\int_0^L v\phi_i\frac{\partial\phi_j}{\partial x}\,\mathrm{d}x\right)c_j$$
$$= \sum_{j=1}^N \underbrace{\left(-K_\alpha \int_0^L \frac{\partial\phi_i}{\partial x}\frac{\partial G_{\alpha\beta}(x,\phi_j)}{\partial x}\,\mathrm{d}x\right)}_{\equiv \mathrm{D}_{ij}} c_j, \tag{12}$$

where we have assumed, for simplicity, that the non-local diffusive flux $K_\alpha \frac{\partial}{\partial x}G_{\alpha\beta}(x,c^h)$ is vanishing on the boundaries.

Since ϕ_j is usually a low-order polynomial, its fractional derivative can be computed analytically (see Appendix for more details). Numerical calculations are also possible provided that the singularity at x is handled carefully. In Fig. 3, we show a piecewise linear basis function, the corresponding space-averaging function $G_{\alpha\beta}$ and the fractional derivative of order $\alpha - 1$, i.e., $\frac{\partial G_{\alpha\beta}(x,\phi_j)}{\partial x}$, for different values of α. It can be seen that the global space-averaging effect increases as α decreases. The decentering effect of the skewness parameter β is also illustrated. In this case, $G_{\alpha\beta}$ would have been symmetric for $\beta = 0$.

Like the FD method, the FE method results in a full diffusion matrix as the fractional-order derivative of a basis function is no more a local-support function. As a result, the computational cost of the numerical model is expected to substantially increase when going from second-order to fractional-order diffusion. Moreover, the calculation of the integrals involving the fractional-order derivative of a basis function might be costly. In that respect, linear basis functions are particularly interesting as their first derivative is constant. The diffusion term can thus be expressed as follows:

$$\mathrm{D}_{ij} = -K_\alpha \frac{\partial\phi_i}{\partial x}\Big|_{[x_{i-1},x_i]} \int_{x_{i-1}}^{x_i} \frac{\partial G_{\alpha\beta}(x,\phi_j)}{\partial x}\,\mathrm{d}x - K_\alpha \frac{\partial\phi_i}{\partial x}\Big|_{[x_i,x_{i+1}]} \int_{x_i}^{x_{i+1}} \frac{\partial G_{\alpha\beta}(x,\phi_j)}{\partial x}\,\mathrm{d}x$$
$$= -\frac{K_\alpha}{x_i - x_{i-1}}\left[G_{\alpha\beta}(x,\phi_j)\right]_{x_{i-1}}^{x_i} + \frac{K_\alpha}{x_{i+1} - x_i}\left[G_{\alpha\beta}(x,\phi_j)\right]_{x_i}^{x_{i+1}},$$

where we have used the fact that ϕ_i is linear and only defined on $[x_{i-1}, x_i]$ and $[x_i, x_{i+1}]$ (See Fig. 3a). Regarding conservation, the use of a Galerkin formulation and the integration by parts of the diffusion term guarantees that mass is going to be conserved globally.

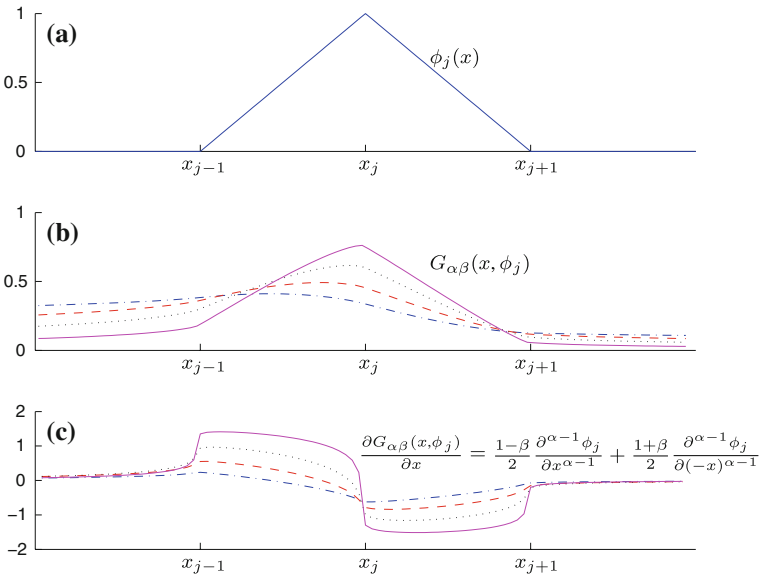


Fig. 3 Example of **(a)** a linear FE basis function ϕ_j, **(b)** the corresponding space-averaging function $G_{\alpha\beta}(x,\phi_j)$, and **(c)** the fractional derivative of ϕ_j for $\alpha = 1.2$ (.-), 1.4 (--), 1.6 (..), 1.8 (−) and $\beta = 0.5$. Note that although ϕ_j is a local-support function, its fractional derivative is not

3.3 Spectral element method

The main disadvantage of the FD and FE methods is that they both result in full diffusion matrices that could render the computational code inefficient. These methods thus require quite more operations to solve the fractional-order ADE than to solve the second-order ADE. This issue is related to the fact the both the FD and (standard low-order) FE methods are local methods that usually update one degree of freedom (dof) by using only a small number of neighboring dof's. As a result, the FD and FE methods require a large number of dof's to provide an accurate solution but are well-suited to local (i.e., integer-order) differential operators as they result in sparse system matrices. However, when applied to non-local differential operators, such numerical methods could become inefficient.

Therefore a non-local numerical method, like the SE method, might be better suited to solve Eq. 3. That method is quite similar to the FE method and also approximates the exact solution c with a discrete solution c^h defined by Eq. 4. The main difference is that the basis functions ϕ_j are now high order functions defined over the whole computational domain. Sines and cosines are typically used for periodic problems while Chebyshev or Legendre functions are used for non-periodic ones (see for instance Boyd [6] for more details). With the SE method, each dof directly depends on all the dof's defining the solution. Such methods are thus of order N, where N is the number of dof's, and the convergence rate is exponential. As a result, to obtain a given level of accuracy, the SE method requires a much smaller number of dof's than the FD and FE methods. Moreover, as the SE method always requires the calculation of integrals of complex functions over the entire computational domain, the solution of a fractional-order ADE is not expected to be substantially more expensive. To our knowledge, the SE method has never been used to solve the fractional-order ADE.

If Chebyshev basis functions and a Galerkin formulation are used, the resulting set of discrete equations is similar to (12) with the exception that (i) ϕ_i is a Chebyshev polynomial of degree i and (ii) $\mathrm{d}x$ should be multiplied by the weight $w(x) = (1-(2x/L-1)^2)^{-1/2}$ in all the integrals in order to make use of the orthogonality property of Chebyshev polynomials. In that case, the discrete equations read:

$$\sum_{j=1}^{N}\left(\int_0^L \phi_i\phi_j w(x)\,\mathrm{d}x\right)\frac{\mathrm{d}c_j}{\mathrm{d}t}+\sum_{j=1}^{N}\left(\int_0^L v\phi_i\frac{\partial\phi_j}{\partial x}w(x)\,\mathrm{d}x\right)c_j = \sum_{j=1}^{N}\underbrace{\left(K_\alpha\int_0^L \phi_i\frac{\partial^2 G_{\alpha\beta}(x,\phi_j)}{\partial x^2}w(x)\,\mathrm{d}x\right)}_{\equiv \mathrm{D}_{ij}}c_j,$$

for $i = 1,\ldots,N$. It should be noted that the diffusion term has not been integrated by parts. This is mainly because the diffusive flux $\phi_i w\frac{\partial G_{\alpha\beta}(x,c^h)}{\partial x}$ does not vanish on the boundaries of the domain since ϕ_i is a global function that is generally not equal to zero on the boundary. As a result (and unlike with the FE method), the flux term has to be computed even when Dirichlet boundary conditions are imposed, which makes it rather cumbersome.

As the SE methods requires less dof's to obtain an accurate solution, there will be less evaluations of $G_{\alpha\beta}(x,\phi_j)$ and less integrals to compute. Moreover, since Chebyshev functions are polynomials of increasing order, their fractional-order derivative can be computed analytically. Figure 4 shows a fourth order Chebyshev basis function, the corresponding space-averaged function $G_{\alpha\beta}$ and the fractional derivative for different values of α.

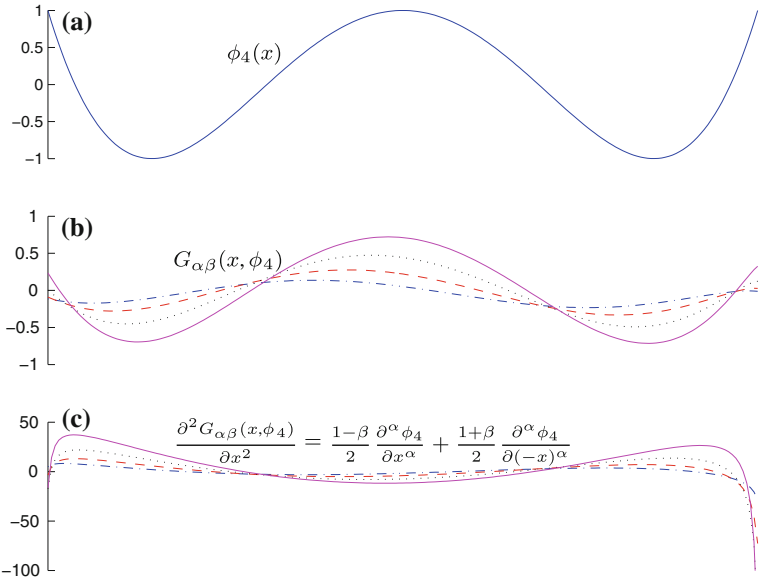


Fig. 4 Same as Fig. 3 but for a fourth-order Chebyshev polynomial. In this case, both the basis function ϕ_4 and its fractional derivative are global functions

4 Numerical example

In this section, we compare the three numerical discretizations of the fractional-order ADE introduced in the previous section. In order to assess the efficiency and accuracy of the three methods, we consider a one-dimensional benchmark problem introduced by Tadjeran et al. [32]. The problem consists in finding $c(x,t)$ such that

$$\frac{\partial c(x,t)}{\partial t} = d(x)\frac{\partial^{1.8} c(x,t)}{\partial x^{1.8}} + q(x,t) \quad \text{for } x \in [0,1] \text{ and } t > 0, \tag{13}$$

with $d(x) = \Gamma(2.2)x^{2.8}/6$, $q(x,t) = -(1+x)e^{-t}x^3$, $c(x,0) = x^3$, $c(0,t) = 0$ and $c(1,t) = e^{-t}$. In that case, the exact solution of (13) reads:

$$c(x,t) = e^{-t}x^3.$$

Equation 13 has been discretized with FD, FE and SE schemes and solved until $t = 1$. At the end of the simulation, the numerical solutions have been compared with the exact solution.

Figure 5 shows the rate of convergence of the three methods. For the FE and SE schemes, the relative error has been computed in the L_2 norm. For the FD scheme, a root mean square error has been computed. The integrals defining the FE and SE diffusion matrices have been computed numerically by means of Gauss and Gauss–Laguerre quadrature rules, respectively. As expected, the convergence rates for FD, FE and SE schemes are linear, quadratic and exponential, respectively. As mentioned previously, the linear convergence rate of the FD scheme can be increased to quadratic by using the approach proposed by Tadjeran et al. [32]. Since all the schemes result in a full diffusion matrix, the computational cost per dof is similar for the three schemes. It therefore appears that the SE scheme is the most efficient since it requires much less dof's to achieve the same accuracy. For instance, to achieve a 0.1% relative error, the FD, FE and SE schemes require about 80, 30 and 5 dof's, respectively.

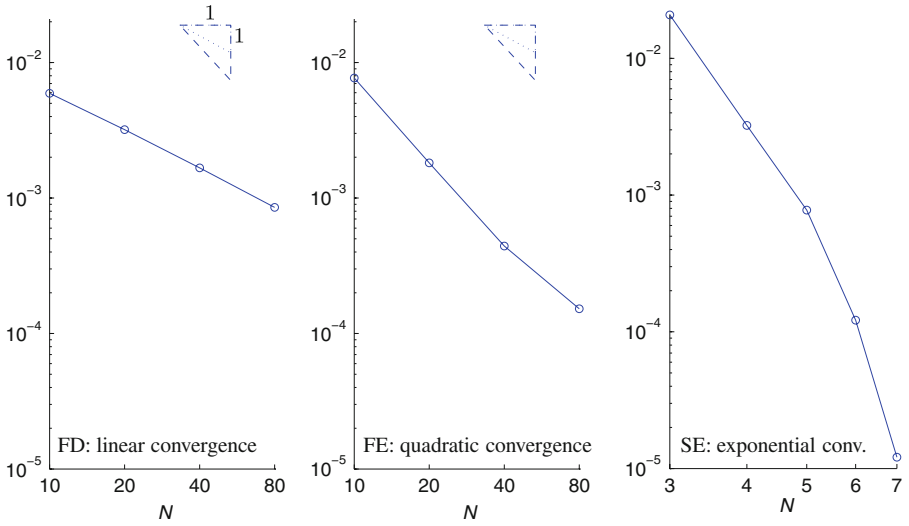


Fig. 5 Convergence analysis of the FD, FE and SE methods for Tadjeran et al.'s [32] test case. Note that approximatively 80, 30 and 5 dof's are needed for the FD, FE and SE schemes, respectively, to achieve an error level of 10^{-3}

5 Conclusion

Just as the Gaussian distribution is a special case of a Lévy distribution, second-order transport models are just a special case of fractional-order transport models. Such models allow the representation of a much larger array of processes, ranging from local molecular diffusion to non-local diffusion in complex heterogeneous systems. The non-local nature of fractional-order transport models requires a specific numerical treatment as traditional methods like the FD and FE methods are not designed to efficiently discretize non-local spatial differential operators.

Such operators are non-local in the sense that they can be expressed in terms of the integral of the solution over the entire computational domain. They are therefore better suited to non-local numerical methods, like the SE method, which naturally take the global behavior of the solution into account. The computational cost of the method does not significantly increase when moving from an integer-order to a fractional-order transport model. In addition to being more efficient than the FD and FE methods, the SE method also achieves a higher rate of convergence.

Further work will be required to design an optimal SE scheme for fractional-order transport models. For instance, the efficiency of the SE method could be further improved by deriving basis functions that would diagonalize the diffusion matrix, i.e., basis functions that are eigenfunctions of the fractional-order diffusion operator. In addition to that model parameters estimation and mathematical properties like stability, positivity, and uniqueness of the solution will have to be rigorously assessed before using such a numerical scheme for realistic environmental problems.

Appendix: analytical results

In this section, we present some analytical results useful for calculating the fractional derivatives of FE and SE basis functions. Since both sets of basis functions are polynomials, we can assume that ϕ_j has the following expression:

$$\phi_j(x) = \sum_{k=0}^{m} c_k x^k.$$

The left Riemann–Liouville derivative of ϕ_j then reads:

$$\frac{\mathrm{d}^\alpha \phi_j}{\mathrm{d}x^\alpha} = \sum_{k=0}^{m} c_k \frac{\mathrm{d}^\alpha x^k}{\mathrm{d}x^\alpha} = \frac{1}{\Gamma(n-\alpha)} \sum_{k=0}^{m} c_k \frac{\mathrm{d}^n}{\mathrm{d}x^n} \int_0^x \frac{y^k}{(x-y)^{\alpha-n+1}}\,\mathrm{d}y, \tag{14}$$

where n is such that $n-1 < \alpha \leq n$. The expression of the right Riemann–Liouville derivative is similar to (14). The integrals of the powers in (14) can be computed analytically by using the following primitives:

$$\int \frac{y^0}{(x-y)^{\alpha-n+1}}\,\mathrm{d}y = -\frac{(x-y)^{n-\alpha}}{n-\alpha}$$

$$\int \frac{y^1}{(x-y)^{\alpha-n+1}}\,\mathrm{d}y = \frac{(x-y)^{n+1-\alpha}}{n+1-\alpha} - \frac{x(x-y)^{n-\alpha}}{n-\alpha}$$

$$\int \frac{y^2}{(x-y)^{\alpha-n+1}}\,\mathrm{d}y = -\frac{(x-y)^{n+2-\alpha}}{n+2-\alpha} + \frac{2x(x-y)^{n+1-\alpha}}{n+1-\alpha} - \frac{x^2(x-y)^{n-\alpha}}{n-\alpha}$$

$$\begin{array}{ccc} \vdots & = & \vdots \end{array}$$

$$\int \frac{y^k}{(x-y)^{\alpha-n+1}}\,\mathrm{d}y = \sum_{i=0}^{k}(-1)^{i+1}\binom{k}{i}\frac{x^{k-i}(x-y)^{n+i-\alpha}}{n+i-\alpha}.$$

As an example, let us consider the FE basis function represented in Fig. 3. It can be expressed as:

$$\phi_j(x) = \begin{cases} 0 & x < x_j \text{ or } x \geq x_{j+1}, \\ \frac{x-x_{j-1}}{x_j-x_{j-1}} & x_{j-1} \leq x < x_j, \\ \frac{x_{j+1}-x}{x_{j+1}-x_j} & x_j \leq x < x_{j+1}. \end{cases}$$

For that basis function, the left Riemann–Liouville derivative reads:

$$\frac{\mathrm{d}^\alpha \phi_j(x)}{\mathrm{d}x^\alpha} = \frac{1}{\Gamma(n-\alpha)} \times \frac{\mathrm{d}^n}{\mathrm{d}x^n} \begin{cases} 0 & x < x_{j-1}, \\ \frac{1}{x_j-x_{j-1}}\frac{(x-x_{j-1})^{n-\alpha+1}}{(n-\alpha)(n-\alpha+1)} & x_{j-1} \leq x < x_j, \\ \frac{1}{x_j-x_{j-1}}\left[(x-x_j)^{n-\alpha}\left(\frac{x-x_j}{n+1-\alpha}-\frac{x-x_{j-1}}{n-\alpha}\right)+\frac{(x-x_{j-1})^{n-\alpha+1}}{(n-\alpha)(n-\alpha+1)}\right] & \\ \quad +\frac{1}{x_{j+1}-x_j}\left[(x-x_j)^{n-\alpha}\left(\frac{x-x_j}{n+1-\alpha}-\frac{x-x_{j+1}}{n-\alpha}\right)\right] & x_j \leq x < x_{j+1}, \\ \frac{1}{x_j-x_{j-1}}\left[(x-x_j)^{n-\alpha}\left(\frac{x-x_j}{n+1-\alpha}-\frac{x-x_{j-1}}{n-\alpha}\right)+\frac{(x-x_{j-1})^{n-\alpha+1}}{(n-\alpha)(n-\alpha+1)}\right] & \\ \quad +\frac{1}{x_{j+1}-x_j}\left[(x-x_j)^{n-\alpha}\left(\frac{x-x_j}{n+1-\alpha}-\frac{x-x_{j+1}}{n-\alpha}\right)+\frac{(x-x_{j+1})^{n-\alpha+1}}{(n-\alpha)(n-\alpha+1)}\right] & x \geq x_{j+1}. \end{cases}$$

It can be seen that for $1 < \alpha \leq 2, n = 2$ and the expression above diverges at the mesh nodes. However, in the FE formulation (12), only derivatives of order $\alpha - 1$ are considered. In that case, $n = 1$ and the fractional derivative of a basis function remains bounded.

References

1. Batchelor GK (1950) The application of the similarity theory of turbulence to atmospheric diffusion. Quart J R Meteorol Soc 76:133–146
2. Benson DA, Wheatcraft SW, Meerschaert MM (2000) Application of a fractional advection-dispersion equation. Water Resour Res 36:1403–1412
3. Benson DA, Wheatcraft SW, Meerschaert MM (2000) The fractional-order governing equation of levy motion. Water Resour Res 36:1413–1423
4. Berkowicz R, Prahm L (1979) Generalization of K theory for turbulent diffusion. Part 1: spectral diffusivity concept. J Appl Meteorol 18:266–272
5. Berkowitz B, Cortis A, Dentz M, Scher H (2006) Modelling non-Fickian transport in geological formations as a continuous time random walk. Rev Geophys 44(RG2003):3
6. Boyd JP (2001) Chebyshev and Fourier spectral methods, 2nd edn. Dover Publications, New York
7. Chaves AS (1998) A fractional diffusion equation to describe Lévy flights. Phys Lett A 239:13–16
8. Cushman-Roisin B (2008) Beyond eddy diffusivity: an alternative model for turbulent dispersion. Environ Fluid Mech 8:543–549
9. Cushman-Roisin B, Jenkins AD (2006) On a non-local parameterization for shear turbulence and the uniqueness of its solutions. Boundary-Layer Meteorol 118:69–82
10. Davies RE (1983) Oceanic property transport, Lagrangian particle statistics, and their prediction. J Mar Res 41:163–194
11. Deng ZQ, Bengtson L, Singh VP (2006) Parameter estimation for fractional dispersion model for rivers. Environ Fluid Mech 6:451–475

12. Durbin PA (1980) A stochastic model of two-particle dispersion and concentration fluctuations in homogeneous turbulence. J Fluid Mech 100:279–302
13. Einstein A (1905) Uber die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Annalen der Physik 17:549–560
14. Feller W (1971) An introduction to probability theory and its applications, vol II. Wiley, New York
15. Fix GJ, Roop JP (2004) Least square finite-element solution of a fractional order two-point boundary value problem. Comput Math Appl 48:1017–1033
16. Frippiat CC, Holeyman AE (2008) A comparative review of upscaling methods for solute transport in heterogenenous porous media. J Hydrol 362:150–176
17. Gnedenko B, Kolmogorov A (1954) Limit distributions for sums of independent random variables. Addison-Wesley, Cambridge, MA
18. Huang G, Huang Q, Zhan H (2006) Evidence of one-dimensional scale-dependent fractional advection-dispersion. J Contam Hydrol 85:53–71
19. Jenkins AD (1985) Simulation of turbulent dispersion using a simple random model of the flow field. Appl Math Model 9:239–245
20. Kim S, Kavvas ML (2006) Generalized Fick's law and fractional ADE for pollution transport in a river: detailed derivation. J Hydrol Eng 11(1):80–83
21. Lévy P (1954) Théorie de l'Addition des Variables Aléatoires. Gauthier-Villars, Paris
22. Meerschaert MM, Benson DA, Bäumer B (1999) Multidimensional advection and fractional dispersion. Phys Lett A 59:5026–5028
23. Meerschaert MM, Tadjeran C (2004) Finite difference approximations for fractional advection-diffusion flow equations. J Comput Appl Math 172:65–77
24. Okubo A (1971) Oceanic diffusion diagrams. Deep Sea Res 18:789–802
25. Pachepsky Y, Timlin D, Rawls W (2003) Generalized Richards' equation to simulate water transport in unsaturated soils. J Hydrol 272:3–13
26. Podlubny I (1999) Fractional differential equations: mathematics in science and engineering, vol 198. Academic Press, New York
27. Richardson LF (1926) Atmospheric diffusion shown on a distance-neighbour graph. Proc R Soc Lond 110:709–737
28. Richardson LF, Stommel H (1948) Note on eddy diffusion in the sea. J Meteorol 5:238–240
29. Roop JP (2006) Computational aspects of FEM approximation of fractional advection dispersion equations on bounded domains in $\mathbb{R}^2$. J Comput Appl Math 193:243–268
30. Schumer R, Benson DA, Meerschaert MM, Wheatcraft JW (2001) Eulerian derivation of the fractional advection-dispersion equation. J Contam Hydrol 48:69–88
31. Stommel H (1949) Horizontal diffusion due to oceanic turbulence. J Mar Res 8:199–225
32. Tadjeran C, Meerschaert MM, Scheffler H-P (2006) A second-order accurate numerical approximation for the fractional diffusion equation. J Comput Phys 213:205–213
33. Zhang Y, Benson DA, Reeves DM (2009) Time and space nonlocalities underlying fractional-derivative models: Distinction and literature review of field applications. Adv Water Resour 32(4):561–581

Environ Fluid Mech (2010) 10:21–39
DOI 10.1007/s10652-009-9141-8

ORIGINAL ARTICLE

Suspended particulate matter dynamics in a particle framework

Ulf Gräwe · Jörg-Olaf Wolff

Received: 12 March 2009 / Accepted: 25 June 2009 / Published online: 8 July 2009

Abstract Suspended particulate matter (SPM) dynamics in ocean models are usually treated with an advection–diffusion equation for one or more sediment size classes coupled to the hydrodynamical part of the model. Numerical solution of these additional partial differential equations unavoidably introduces numerical diffusion, i.e. in the case of sharp gradients the possible occurrence of artificial oscillations and non-positivity. A Lagrangian particle-tracking model has been developed to simulate short-term SPM dynamics. Modelling individual sediment particles allows a straightforward physical interpretation of the processes. The tracking of large numbers of individual and independent particles (up to 25 million in total in a single sediment class) can be achieved on high performance computer clusters, due to efficient parallelisation of particle tracking. The movement of the particles is described by a stochastic differential equation, which is consistent with the advection–diffusion equation. Here, the concentration profile is represented by a set of independent moving particles, which are advected according to the 3D velocity field, while the diffusive displacements of the particles are sampled from a random distribution, which is related to the eddy diffusivity field. To account for erosion a new parameterisation is proposed. Three numerical particle tracking schemes (EULER, MILSTEIN and HEUN) are presented and validated in idealised test cases. Finally, the particle tracking algorithms are applied to a realistic scenario, a severe winter storm in the East Frisian Wadden Sea (southern North Sea). The comparison with observations and an Eulerian SPM transport model seems to indicate a somewhat better fidelity of the Lagrangian approach.

Keywords Lagrangian particles · Random walk · Stochastic processes · SPM transport

1 Introduction

The behaviour of particles in turbulent flows has been studied for many years, ranging for meteorology [4,8] to ocean dynamics [20,35,36]. Extensive literature exists on the treatment

U. Gräwe (✉) · J.-O. Wolff
ICBM, Physical Oceanography (Theory), University of Oldenburg, Oldenburg, Germany
e-mail: graewe@icbm.de

of Lagrangian trajectories, ranging from highly idealised flows to situations as complex as the unstable convective boundary layer or frontal zones. The level of understanding of these types of models has greatly increased over the years. In the same time the need to predict the transport of particles, pollutants or biological species has resulted in a rapid rise in the use of these numerical models.

The random walk simulation model enables the observation of phenomena on scales much smaller than the grid size, as well as the tracing of the movement of individual particles, thereby describing the natural processes more accurately. Furthermore, information on integrated properties like: residence/settling time or individual tracks are easily extracted from the simulations. Concentrations of particles can be directly calculated from the spatial positions of the particles and, more importantly, when and where required. Additionally, errors due to numerical diffusion inherent in methods such as finite differences or finite elements, are avoided, particularly in areas where high concentration gradients exist, such as close to point sources or frontal zones. Although there are methods to circumvent these difficulties [10], their implementation is problematic in complex geometries, where it is difficult to control the potential sources of error.

The development of particle tracking methods (or random walk/random dispersion methods) started by tracking neutrally buoyant particles, i.e. water parcels [18,35]. Hunter [14] and Visser [35] also showed that due to the high spatial variability of turbulence, the tracking algorithms need special modification to avoid numerical artefacts. In recent years a catalogue of test cases was developed to compare the performance of tracking schemes and also to validate the models [4,11,28]. Deleersnijder et al. [11] extended the test catalogue to particles that have a finite sinking velocity. By this, particle tracking schemes dealing with sediment or buoyant particles could be validated against an analytical solution.

The random walk schemes for modelling SPM dynamics are quite attractive, because they give a straightforward physical interpretation of the processes and automatically account for suspension and bed load. Because of these advantages Lagrangian schemes became more common in the SPM modelling community [9,16,26]. Nevertheless most of these models used only small number of particles $O(10^4)$. Nowadays with easy access to high performance computer clusters, the tracking of individual particles can be parallelised with high efficiency and therefore makes huge particle numbers feasible [9]. This means to deal with particles in the order of $>10^7$. This is still negligible, by realising that a bucket of muddy water contains more individual particles, compared to the ability of state of the art Lagrangian schemes. Nonetheless increasing the number of particles leads to a better statistical description and makes the answers (a Lagrangian model can give) more reliable.

In this paper, random walk schemes associated with individual particles with a finite sinking velocity are established for the multidimensional case. These schemes are not new, but the analysis will show there limitations and advantages. In the first section the underlying theory is shortly summarised, the numerical schemes are established and the assumptions made are explained. In Sect. 2 the tracking schemes are validated against simplified analytical test cases, two 1-D and a 2-D test. In the third section the particle tracking schemes are applied to a realistic test. The model is used to simulate the short-term SPM dynamics during a severe winter storm in the East Frisian Wadden Sea. Finally, conclusions are given in Sect. 4.

2 The Lagrangian model

Dealing with concentration fields (SPM, pollutants, biology, etc.), the time evolution of these fields is usually formulated as partial differential equations (PDEs) in an Eulerian framework.

$$\partial_t C = -\nabla \cdot (\mathbf{u}\, C - \mathbf{K} \cdot \nabla C) \tag{1}$$
$$\nabla \cdot \mathbf{u} = 0$$

The first equation is an advection–diffusion equation for the concentration field C of a passive tracer, that is coupled to a 3D velocity field $\mathbf{u}$, that shall be divergence free. The diffusivity tensor $\mathbf{K}$ is symmetric and positive definite. In the following only diagonal diffusivity tensors are considered

$$\mathbf{K} = \begin{bmatrix} K_H & 0 & 0 \\ 0 & K_H & 0 \\ 0 & 0 & K_Z \end{bmatrix} \tag{2}$$

where K_H is the horizontal and K_Z the vertical diffusivity. Hence, the three spatial dimensions are decoupled. Instead of solving the PDE, one can transform the whole solution process into the solution of a system of stochastic differential equations (SDEs) also called Langevin equations. The basic idea is to interpret the concentration field $C(\mathbf{x}, t)$ as a transition density field and reinterpret Eq. 1 as a Fokker-Planck equation, i.e. a deterministic PDE with regard to transition density functions. This can be solved by the following system of SDEs defined in the Îto sense [1]:

$$d\mathbf{X}(t) = (u + \nabla \cdot \mathbf{K})dt + \sqrt{2\mathbf{K}}\, d\mathbf{W}(t) \tag{3}$$

Here $\mathbf{X}(t)$ is the position vector of the particles and $d\mathbf{W}(t)$ is a Wiener noise increment with the following properties. $\mathbf{W}(t)$ is a Gaussian process with independent increments for which holds

$$\langle \mathbf{W}(t) \rangle = 0; \;\; \mathrm{Std}\,(\mathbf{W}(t) - \mathbf{W}(s)) = \sqrt{|t-s|}\mathbf{I} \tag{4}$$

where $\mathbf{I}$ is the identity matrix. Therefore the noise process has a vanishing mean $\langle \cdot \rangle$, its standard deviation scales as $\sqrt{dt}$ and the increments are uncorrelated.

The first term on the right hand side of Eq. 3 represents the deterministic part, whereas the second term is the stochastic term. In the case of vanishing turbulent diffusivity, the system of equations reduces to a system of ordinary differential equations (ODEs). Because the ocean is a turbulent environment, turbulent diffusion has to be included. This is incorporated via the stochastic term. The particles experience a random displacement due to eddies of average size $\sqrt{2\mathbf{K}dt}$. Because the turbulent diffusivity $\mathbf{K} = \mathbf{K}(\mathbf{x}, t)$ is spatially highly variable, the term $\nabla \cdot \mathbf{K}$ needs to be added to the deterministic part. This corrects for an artificial noise induced drift [14,35].

2.1 Numerical approximation

Because the diffusivity tensor is diagonal, the three spatial directions can be treated separately in developing a numerical approximation to the 3D Langevin equation. Focussing for simplicity on the vertical dimension the following equation needs to be discretised.

$$dZ(t) = (w + \partial_z K_Z(z))\, dt + \sqrt{2\, K_Z(z)}\, dW(t) \tag{5}$$

This equation can further be simplified to

$$dZ(t) = a(z)\, dt + b(z)\, dW(t) \tag{6}$$

where $a = w + \partial_z K_Z(z)$, represents the deterministic part and $b = \sqrt{2K_Z(z)}$ is the stochastic part. Again Eqs. 6 and 5 are only valid in the Îto interpretation [1]. Instead of writing Eq. 6 in differential form, it is also common to use the integral representation

$$Z_t = Z_0 + \int_0^t a(Z_s)\, ds + \int_0^t b(Z_s)\, dW_s \tag{7}$$

A straightforward translation of Eq. 6 into a numerical scheme, is simply to replace dt by Δt and dW by ΔW. This is equivalent to assuming that $a(Z_s)$ and $b(Z_s)$ in Eq. 7 are constant and can be taken out of the integrals. Therefore, the lowest order approximation reads as

$$Z_{n+1} = Z_n + a\Delta t + b\Delta W_n \tag{8}$$

This is also known as Euler scheme. In the following, this approximation is named EULER. This scheme is commonly used [4,20,24,28,35]. Although this is a straightforward approach, some difficulties arise in the case of SDEs. To define the accuracy or order of convergence for stochastic scheme two cases have to be distinguish. For SDEs the order of convergence is separated into *weak* and *strong* [1,15]. A method is said to have *weak/strong* order of convergence of γ if there exists a constant Λ such that

$$\begin{aligned} |\langle p(Z_n)\rangle - \langle p(Z(\tau))\rangle| &\leq \Lambda\ \Delta t^{\gamma} : \text{weak} \\ \langle |Z_n - Z(\tau)|\rangle &\leq \Lambda\ \Delta t^{\gamma} : \text{strong} \end{aligned} \tag{9}$$

for any fixed $\tau = n\Delta t \in [0, T]$ and Δt sufficiently small. Z_n represents the true solution and $Z(\tau)$ is the approximation. $p(\cdot)$ is an arbitrary function (in most cases a probability density function). The *weak* criterion asks for the difference in a distribution, whereas the *strong* criterion accounts for the difference in the trajectory.

As discussed after Eq. 4 the increment ΔW scales as $\sqrt{\Delta t}$, hence the whole EULER scheme is only of order $\sqrt{\Delta t}$ in the strong convergence. Since we are interested in the time evolution of a sediment distribution rather than individual trajectories of single sand grains, the weak convergence is used. In this case the EULER schemes is or order Δt in the weak sense.

To develop numerical schemes, that have a higher accuracy in the strong definition, the assumption that $a(Z_s)$ and $b(Z_s)$ in Eq. 7 are constant is not valid any more. Using the appropriate Taylor approximation for the integrals, see e.g. [1,15], the next higher order approximation reads as

$$Z_{n+1} = Z_n + a\Delta t + b\Delta W_n + \frac{1}{2} bb' \left[(\Delta W_n)^2 - \Delta t\right] \tag{10}$$

where b' is the spatial derivative. This is also known as the MILSTEIN scheme. This scheme is of order Δt in the weak and strong convergence. Additional accuracy is gained by including information of the derivative of the noise term b. If this scheme is used in numerical algorithms, the term bb' can cause problems due to round off errors. To avoid this, a symmetry property of Eq. 6 can be used. By replacing b again by $\sqrt{2K_Z(z)}$ and computing the term bb' explicitly it turns out that it is equal to $\partial_z K_Z(z)$. After some rewriting the schemes read as follow

$$\begin{aligned} Z_{n+1} &= Z_n + w\Delta t + \partial_z K\, \Delta t + \sqrt{2K}\Delta W && : \text{EULER} \\ Z_{n+1} &= Z_n + w\Delta t + \partial_z K\, \tfrac{(\Delta W)^2 + \Delta t}{2} + \sqrt{2K}\Delta W && : \text{MILSTEIN} \end{aligned} \tag{11}$$

As stated above the EULER scheme is commonly used, but it is easily appreciated that a higher accuracy is gained here by a simple multiplication of $\Delta W \cdot \Delta W$ and one addition. There are no approximations involved and the extra computational cost is negligible.

To further improve the accuracy of the numerical schemes, more terms in the Taylor approximations have to be included [15]. Due to the slow convergence of the numerical schemes for SDEs, the extra computational costs are so far prohibitive. Therefore, a multi step scheme is proposed, similar to Runge-Kutta schemes for ODEs.

$$Z_{n+1} = Z_n + \frac{1}{2}\left(a(\tilde{Z}) + a\right)\Delta t + b\Delta W_n \tag{12}$$

with

$$\tilde{Z} = Z_n + a\Delta t + b\Delta W_n \tag{13}$$

Equation 12 is a stochastic version of the trapezoidal method also known as HEUN scheme. Note that the predictor step (13) is only applied to the deterministic part, the stochastic part is not corrected to keep the numerical approximation consistent with Eq. 6 [15]. The HEUN scheme, like the MILSTEIN scheme, is of order Δt in the strong and weak convergence.

At this stage, the question might arise, why three algorithms are presented with the same accuracy. All three algorithms should behave identical in predicting the time evolution of an initial concentration, because they have the same order of convergence (in the weak sense). Assume the limit of vanishing diffusivity, here the MILSTEIN scheme becomes identical to the EULER scheme. This is not the case for the HEUN scheme. Due to the predictor–corrector step, the accuracy is higher. Therefore, in the case of advection-dominated problems, differences will be visible. This should not be the case if diffusion dominates. However, especially Sect. 3.3 will show the limitation of the EULER scheme. Moreover, near boundaries, the proper approximations of the particle trajectories become important to resolve for instance the boundary layer.

2.2 Boundary conditions

The treatment of boundary conditions is always a critical issue in ocean modelling, especially in coastal regions. The moving sea surface, the sea bottom and lateral boundaries like islands or beaches have to be considered appropriately. For example, in the framework of PDEs, the sea surface is an impermeable boundary and a no flux condition is normally imposed (at least for suspended particulate matter). This no flux condition can be easily violated by overshooting of the trajectories of simulated particles, due to either too large time steps or the random nature of the stochastic increment ΔW. This can lead to a crossing of the boundary. To correct this, a straightforward approach would look like this: When a particle crosses the boundary (due to a too large random displacement), it is simply reflected back into the domain by the amount it penetrates into the boundary domain. It would be advantageous in general to minimise the number of particles that crosses the boundaries in the first place. The first solution that comes to mind is to reduce the time step of the particle displacement. However, this would also lead to additional computation time in 'open water'. A more expensive method is the use of a higher-order numerical scheme. This may perhaps not completely prevent the crossing from happening, but it will at least reduce the number of times that it does occur. This was also mentioned by Stijn et al. [34].

In the following all boundaries are treated as reflective boundaries (no flux condition), if not stated otherwise.

3 Idealised test cases

In the following section two simple 1-D test cases are described and a 2-D test is considered.

Because the equations which need to be solved are SDEs, special care is taken for the stochastic increment ΔW. The increments are computed by a state of the art random number generator: the *Mersenne Twister* [19]. This generator produces uniform random numbers in the interval [0,1]. Because the particle numbers are quite large, standard generators are limited by their periodicity. Moreover the *Mersenne Twister* produces uncorrelated random numbers in higher dimensions. These random numbers are transformed to Gaussian random numbers by the *Box-Müller* algorithm [23].

3.1 1-D diffusion

Firstly, the numerical algorithms are applied to a diffusion test in a bounded region. This can be visualised as a one dimensional water column that is bounded by the sea surface and the sea floor. The model is discussed in detail in Deleersnijder et al. [11] or Spivakovskaya et al. [28]. The governing PDE for this case is written as

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial z}\left(K_Z(z)\frac{\partial C}{\partial z}\right) \tag{14}$$

This describes a simple diffusion equation. The diffusivity $K_Z(z)$ has the following form

$$K_Z(z) = 6z(1-z) \tag{15}$$

This parabolic profile is a good approximation of the diffusivity profile in the upper mixed layer, but it is also a good description for a shallow, well-mixed, coastal region [6,37]. Moreover, the parabolic profile is until now the only realistic profile, for which analytical solutions exists (beside constant diffusivity). For simplicity, time dependence is not considered. The boundary (BC) and initial conditions (IC) are

$$\text{BC:} \quad \left[K_Z(z)\frac{\partial C}{\partial z}\right]_{z=0,1} = 0; \quad \text{IC:} \quad C(0,z) = \delta(z - z_0) \tag{16}$$

i.e. "no flux" boundary conditions are imposed at the boundaries of the normalised domain [0,1]. The initial condition is a delta like concentration peak.

The Langevin equation for the particle trajectories takes the following form

$$dZ(t) = \partial_z K_Z(z)dt + \sqrt{2\,K_Z(z)}\,dW(t) \tag{17}$$

Using this setup, an explicit solution for the dispersion of the initial peak is

$$C_A(t,z) = 1 + \sum_{n=1}^{\infty}(2n+1)P_n(2z-1)P_n(2z_0-1)\exp(-6n(n+1)t) \tag{18}$$

where $P_n(z)$ denotes the n-th order Legendre polynomial. Figure 1a presents the analytical solution for different time steps obtained for $z_0 = 0.5$.

To compare the performance of the three numerical algorithms (EULER, MILSTEIN, HEUN) the root mean square error (RMS) is computed

$$\text{RMS} = \sqrt{\frac{1}{4}\sum_{n=1}^{4}\frac{1}{100}\sum_{i=1}^{100}[C_A(z_i,t_n) - C_P(z_i,t_n)]^2} \tag{19}$$

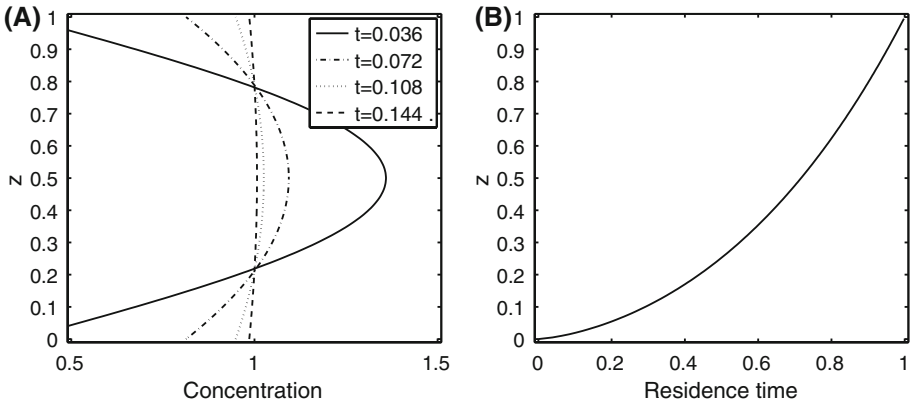


Fig. 1 Analytical solution of **a** the 1-D diffusion test for different moments of the simulation using Eq. 18 and **b** the residence time for the parabolic diffusivity profile Eq. 15 for $w_s = 5$

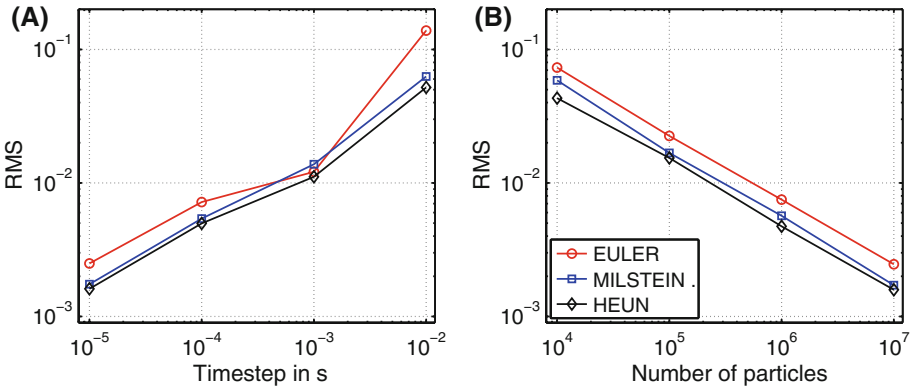


Fig. 2 RMS of the dispersion test: **a** for fixed particle number $N = 10^6$ and **b** for fixed time step $\Delta t = 10^{-4}$

the four time steps t_n are given in Fig. 1a. $C_A(t_n)$ is the analytical solution (18) and $C_P(t_n)$ is the prediction of the Lagrangian models. To estimate $C_A(t_n)$ the water column is binned into 100 equally sized boxes and the concentration is obtained by a box counting approach. The results are shown in Fig. 2. All three schemes converge to the analytical solution if either the step size is decreased or the number of released particles is increased. The HEUN scheme shows the smallest error whereas the EULER scheme has the largest deviation. All three algorithms converge to the true solution, by decreasing the step size, with the same rate, because they have the same order of convergence (see Sect. 2.1). Further by increasing the number of particles N, the approximated solution comes closer to the analytical one. This is due to the intrinsic nature of random processes, because the results include statistical errors proportional to $N^{-\frac{1}{2}}$.

3.2 1-D residence time

In the previous test, the settling velocity w_s was set to zero to obtain an analytical solution. This might be appropriate for tracking water parcels, for buoyant particles, this

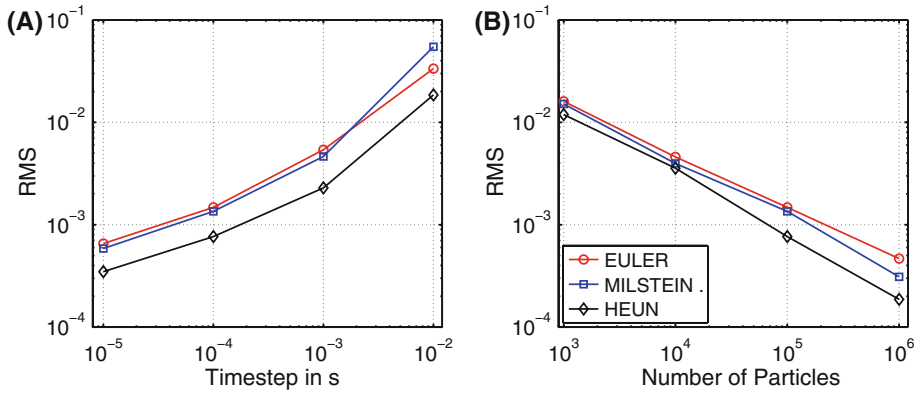


Fig. 3 RMS of the residence time test: **a** for fixed particle number $N = 10^6$ and **b** for a fixed time step $\Delta t = 10^{-4}$ s

assumption is no longer valid. However, imposing a finite sinking velocity makes it impossible to formulate an analytical solution. Nevertheless, an exact solution for the adjoint problem of finding the residence time $\theta(z)$ is known [11,12]. To obtain the residence time $\theta(z_0)$ a number of particles are released at a distance z_0 from the bottom, and then the time is tracked until all particles have crossed the bottom. The boundary and initial condition read again as

$$\text{BC:} \quad \left[K_Z(z)\frac{\partial C}{\partial z}\right]_{z=1} = 0; \quad \text{IC:} \quad C(0,z) = \delta(z - z_0) \tag{20}$$

The boundary condition at the sea floor is now modified. The sea floor is no longer a rigid boundary, but represents for instance the pycnocline. The boundary condition changes from no flux condition to an absorbing type. This means that if a particle reaches the boundary it is immediately removed from the computational domain. Thus

$$\text{BC:} \quad C(t,0) = 0 \tag{21}$$

The diffusivity profile remains the same (15).

The analytical solution can be written as

$$\theta(z) = z + \left(\frac{z}{1-z}\right)^{\mu} B_{1-z}(1+\mu, 1-\mu) \tag{22}$$

where $B_{1-z}(1+\mu, 1-\mu)$ is a generalised incomplete beta function and $\mu = w_s/6$. The analytical solution is shown in Fig. 1b.

The Langevin equation for the particle trajectories takes now the following form

$$dZ(t) = (w_s + \partial_z K_Z(z))\, dt + \sqrt{2K_Z(z)}\, dW(t) \tag{23}$$

To compare again the performance of the three algorithms a non-dimensional sinking velocity of $w_s = 5$ is applied. To approximate the residence time, the average time is computed until all particles have left the domain. The results are shown in Fig. 3. Again, the HEUN scheme performs best. The difference between the EULER and MILSTEIN scheme is nearly vanishing.

3.3 2-D correlation test

One of the most important properties a Lagrangian stochastic model must fulfil is to maintain an initially uniform distribution of particles for all time (in the absence of an advection field)—the well mixed condition (WMC). The test is based on the fact that if a WMC does not exist, then a significant correlation would exist between the perturbation field $C'(x, y, t) = C(x, y, t) - \bar{C}$ and the perturbation diffusivity field $K'_H(x, y) = K_h(x, y) - \bar{K}_H$. Here $\bar{C}$ is the initial well mixed concentration and $\bar{K}_H$ is the xy-averaged diffusivity field. By this, the approach of Brickman and Smith [4] is adopted to demonstrate the WMC. Contrary to their work, the following diffusivity field is chosen

$$K_H(x, y) = K_0 \left[\cos\left(\frac{2\pi n}{L_x}x\right) + 1\right]\left[\cos\left(\frac{2\pi n}{L_y}y\right) + 1\right] \tag{24}$$

where $K_0 = 20\,\mathrm{m}^2/\mathrm{s}$, $L_x = L_y = 40\,\mathrm{km}$ and n an integer number. The domain has a resolution of $\Delta x = \Delta y = 200\,\mathrm{m}$. The time step is set to $\Delta t = 30\,\mathrm{s}$. In the experiments 10^6 particles are released and uniformly distributed in the computational domain. Further, the diffusivity field is chosen to have a vanishing gradient at the boundaries (which are again reflective).

The SDEs are

$$\begin{aligned} dX(t) &= \partial_x K_H dt + \sqrt{2K_H} dW_x(t) \\ dY(t) &= \partial_y K_H dt + \sqrt{2K_H} dW_y(t) \end{aligned} \tag{25}$$

Although the diffusivity field has an analytical formulation, the diffusivity field is only given at the discrete grid points. The spatial derivatives are computed using a second order approximation of the gradient. The values are bilinear interpolated to the position of the particle. n is now the test parameter. By increasing n, the spatial variation in the diffusivity field is also increasing. It is expected that for higher n the numerical schemes start to fail, due to errors in the approximation of the $\partial_x K_H$ and $\partial_y K_H$ terms. To show that the schemes really fail due to discretisation errors, a fourth scheme is applied. This is called EULER-analytic. Here the diffusivity field as well as the spatial derivatives are provided by analytical functions (24).

To test now for the WMC, the normalised correlation coefficient is calculated as

$$\rho(t) = \frac{1}{\sigma_C \sigma_{K'_H}} \sum^{i} \sum^{j} C'_{ij} K'_{Hij} \tag{26}$$

where σ is the standard deviation of the perturbation fields. To compute significance levels at a given time step, the C' field was repeatedly randomised and the test statistic $\tilde{\rho}$ recalculated. A histogram of 5,000 $\tilde{\rho}$ values was constructed and $\pm\tilde{\rho}$ values corresponding to 95% of the histogram area were computed. If the actual ρ is outside the $\pm\tilde{\rho}_{95}$ range, then it is deemed significant and if this is true for all time steps and number of particles, then it is considered that the WMC is not satisfied.

In Fig. 4 the performance of the schemes is shown. Here, for the first time, clear differences are visible. The EULER scheme starts to fail the WMC test for $n = 3$, the MILSTEIN scheme follows at $n = 8$ and the HEUN scheme starts to deviate from a uniform distribution for $n = 11$. The excellent performance is due to the predictor–corrector step of the HEUN scheme. By this the ∇K_H term is better approximated. The same holds for the MILSTEIN scheme. The first order correction terms to the EULER scheme (see Eq. 11) is a modification of the ∇K term and can therefore be better estimated. The performance of the EULER-analytic scheme shows, that the failure of the EULER scheme is indeed caused by the insufficient estimation of the ∇K_H term. This algorithm starts to fail for $n = 11$ and is therefore

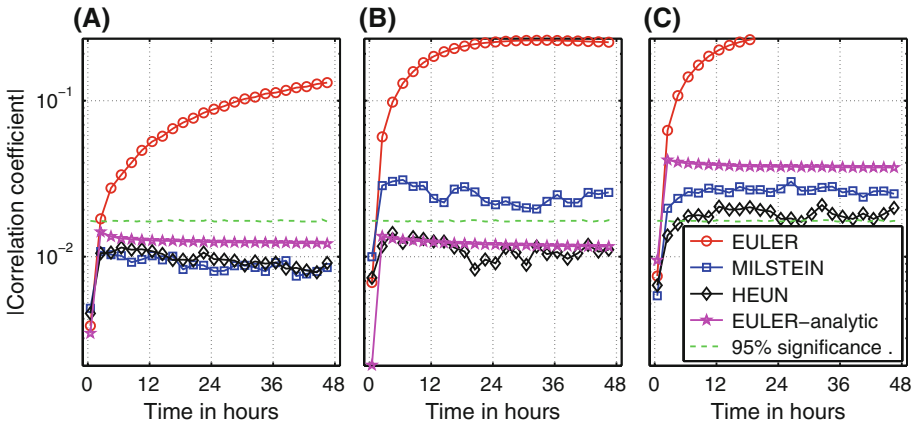


Fig. 4 Absolute value of the correlation coefficient $\rho(t)$ of the WMC test: **a** $n = 3$, **b** $n = 8$ and **c** $n = 11$ and the $\tilde{\rho}_{95}$ significance level. Plotted are the averages of five independent runs

similar to the HEUN scheme. This performance boost, by providing analytical functions to the schemes, was also expected by Brickman and Smith [4].

4 3D realistic test case

In this test, the Lagrangian model is applied to a real scenario. The study site (see Fig. 5c) is part of the East Frisian Wadden Sea, which is located in the southern part of the North Sea along the northwestern part of the German coast. It consists of several barrier islands and tidal basins with large tidal flats between those islands and the main coast. The tidal basins are connected via tidal inlets with the North Sea. The tidal range is about 3 m during spring tide and about 1 m during neap tide. Because the tidal inlets connect the North Sea with the tidal flat areas, peak velocities can reach up to 1.5 m/s. In this setup the model is used to estimate the SPM dynamics during a winter storm (*Britta* 31 October 2006 to 2 November 2006, see Fig. 6) by considering different sediment fractions. The numerical experiments are based on a coupled system of models comprising a hydrodynamic core model General Estuarine Transport Model (GETM, [5]), a wave model Simulating WAves Nearshore (SWAN, [3]) and the Lagrangian SPM module attached to GETM.

This realistic test case is performed with the identical model setup as described in Lettmann et al. [17] besides the SPM transport method outlined above. For a more detailed discussion of the model setup, the coupling between the hydrodynamic module and the wave module and for validation of the model results the interested reader is referred to Lettmann et al. [17]. In the following only model details are discussed that are essential for this study.

4.1 Hydrodynamics

GETM [5,7] is a prognostic 3-D hydrodynamical model especially suited for shallow coastal regions under the influence of tidal currents where substantial areas are prone to drying and flooding during a normal tidal cycle. The model is based on the horizontal momentum equations, the continuity equation and two prognostic equations for the turbulent kinetic energy and its dissipations rate influencing the vertical eddy viscosity coefficient. In horizontal

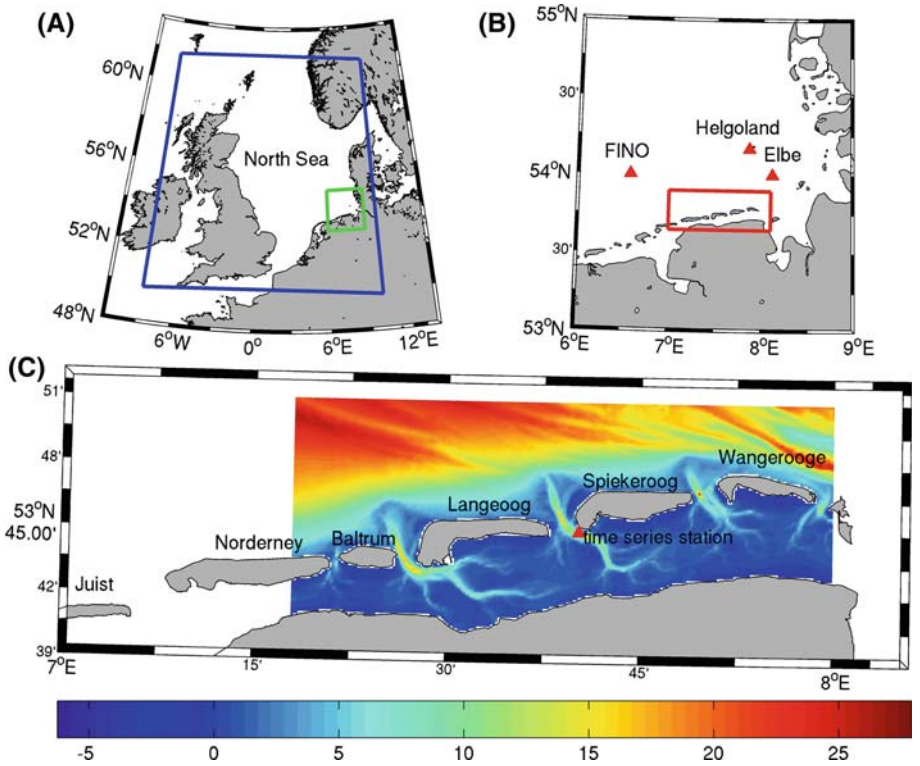


Fig. 5 Location of study site and model area. The model topography (positive downward, unit is meter) of the Wadden Sea model is shown in **c**. The *red triangle* in that panel marks the position of the time series station, from where measured SPM concentrations in the Wadden Sea area are obtained. The *blue box* in **a** denotes the extension of the model domain used for wave modelling in the North Sea area, while the *green box* marks extension of the model domain used for wave modelling in the German Bight. The *red triangles* in **b** denote the positions of the wave buoys used for validation of the model data

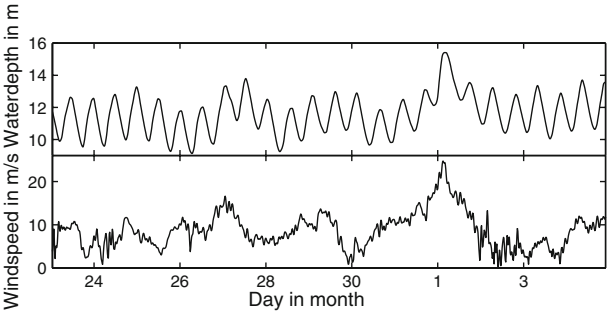


Fig. 6 Time series of water depth and wind speed measured at the pile station (see Fig. 5c) in October/November

directions, GETM uses a constant momentum viscosity/diffusivity ($K_H = 1\,\mathrm{m}^2/\mathrm{s}$). This hydrodynamic core model has been applied and validated in various recent studies (see e.g. [17,29–33]). In the current setup the model has a horizontal resolution of 200 m and uses 20 sigma-levels in the vertical.

The waves in the East Frisian Wadden Sea during the storm surge event have been modelled using a nested system of different model domains, with increasing spatial resolution. The largest area with the coarsest resolution of 20 km grid spacing covers the whole North Atlantic. This enlarged area is needed to determine also the wind generated waves in the Atlantic, which could reach the North Sea, e.g. in the form of swell. Embedded into this large domain is a model domain of the North Sea area (see blue box in Fig. 5a) with a 5 km grid resolution. The wave boundary conditions for this model domain are obtained from the Atlantic domain. Finally the third nested domain covers the area of the German Bight with a resolution of 2 km (see green box in Fig. 5a). From this third model area the wave boundary conditions are obtained for the wave simulations in the Wadden Sea with a grid resolution of 200 m.

The bottom shear stress τ_b being responsible for sediment erosion and deposition is modelled as a non-linear combination of a current induced component τ_{current} and a wave induced component τ_{wave} [27], which was used also by [13] and [22]:

$$\tau_b = \left[(\tau_m + \tau_{\text{wave}} \cos(\phi))^2 + (\tau_{\text{wave}} \sin(\phi))^2\right]^{1/2} \tag{27}$$
$$\tau_m := \tau_{\text{current}} \left[1 + 1.2 \left(\frac{\tau_{\text{wave}}}{\tau_{\text{wave}} + \tau_{\text{current}}}\right)^{3.2}\right]$$

Here, ϕ denotes the angle between the direction of wave propagation and current flow.

4.2 Lagrangian model setup

4.2.1 Particle properties

In the simulations five sediment fractions of size 40, 80, 120, 160 and 200 μm are considered with a density of 2,650 kg/m^3. The sinking velocity of the ith particle fraction is calculated after a formula from Soulsby [27], which is valid for non-cohesive irregular sand grains:

$$w_s^i = \frac{\nu}{d}\left\{\sqrt{10.36^2 + 1.049\, D_s^3} - 10.36\right\} \tag{28}$$
$$D_s = \left[\frac{g}{\nu^2}\left(\frac{\rho_S}{\rho_W} - 1\right)\right]^{1/3} d$$

Here, g denotes the acceleration due to gravity, ρ_S and ρ_W the sediment and water density in kg/m^3, ν the kinematic viscosity in m^2/s and d the sediment diameter in m.

A second important quantity that influences the SPM dynamics is the critical bottom shear stress τ_c. This is the shear stress, which must be reached by the fluid to erode sediment from the bottom into the water column. Soulsby [27] provided an empirical formula for that critical shear stress, which is given by:

$$\tau_c = g d(\rho_S - \rho_W)\psi_c \tag{29}$$
$$\psi_c := \frac{0.3}{1 + 1.2\, D_s} + 0.055\ (1 - \exp(-0.02\, D_s))$$

To determine the critical shear stress for a sediment fraction, they are modelled as separate pools, which do not interact with each other. Therefore, the different sediment fractions erode and deposit separately from each other. Each fraction uses its individual critical shear stress. In Table 1 the important properties are listed.

Table 1 Sinking velocities and critical shear stress for the different sediment fractions after formulas

d (μm)	w_s (cm/s)	τ_c (N/m^2)
40	0.09	0.099
80	0.35	0.134
120	0.77	0.154
160	1.32	0.169
200	1.95	0.180

4.2.2 Deposition/erosion

At this point, the differences in the Eulerian and Lagrangian description enter the setup. Describing SPM dynamics with the help of PDEs, deposition/erosion is commonly modelled by a Partheniades [21] type deposition/erosion flux. For the erosion this reads as

$$F_{\text{ero}} \propto \begin{cases} \left(\frac{\tau_b}{\tau_c} - 1\right) & \tau_b > \tau_c \\ 0 & \tau_b < \tau_c \end{cases} \tag{30}$$

While this formulation is well suited for PDEs, it cannot be used in a Lagrangian framework. Because the SPM dynamics is described with discrete particles, they can be either in deposition or in suspension. Therefore, the erosion formulation needs a modification. To compute a characteristic velocity at the bottom boundary layer, the bottom stress is used: $u_* = \sqrt{\tau_b}$, where u_* is defined as the friction velocity. By this, the modification of the erosion formula for a discrete particle reads as

$$\Delta z = \begin{cases} \sqrt{\tau_b - \tau_c}\Delta t \Phi(\tau_c, \tau_b) & \tau_b > \tau_c \\ 0 & \tau_b < \tau_c \end{cases} \tag{31}$$

where $\Phi(\cdot)$ is the Heavyside step function. The assumption is that due to the bottom shear, eddies are generated that can lift up a particle into the water column. These eddies lead to a vertical displacement of $\sqrt{\tau_b - \tau_c}\Delta t$. This formulation allows a smooth transition from particles in deposition to suspension. This can also be interpreted as an energy approach, where the turbulent kinetic energy in the bottom boundary layer has to reach a certain threshold, to be converted into potential energy to lift the particle against gravity. Although this is straightforward, it causes some problems. Assume that $\tau_b > \tau_c$ in a certain area. According to Eq. 31 all particles, which are in deposition, would suddenly go into suspension, which is far from being observed. Therefore $\Phi(\tau_c, \tau_b)$ was introduced

$$\Phi(\tau_c, \tau_b) = \begin{cases} 1 & U > \frac{\tau_c}{\tau_b} \\ 0 & \text{else} \end{cases} \tag{32}$$

where U is an uniform random number in the interval [0,1]. The motivation is the following: the ratio $\frac{\tau_c}{\tau_b}$ varies in the interval [0,1] for $\tau_b > \tau_c$. If the ratio is close to zero, the bottom stress is much larger than the critical shear. Therefore, the probability that a particle goes into suspension is quite high. On the other hand, if the ratio is close to one, the transition from deposited particle to suspension just started and hence the probability to be eroded is low. Thus, there is a certain probability that a particle remains in deposition; even the bottom shear stress is lager than the critical bottom shear stress.

To account for deposition no explicit formulation is used. If during a time step, a particle hits the sea floor it is marked as in deposition. If the bottom shear is below the critical shear it remains in deposition otherwise it is treated according to (31).

4.2.3 Boundary conditions

In the following, the sea surface and solid land boundaries are treated as reflective boundaries. The east/west open ocean boundaries (Fig. 5c) are being periodic. Hence particles leaving the domain at the eastern boundary, enter the domain again at the western boundary and vice versa. Further, if the particle leaves the domain in the back barrier region, it enters the domain in the same region on the other boundary. The same holds for the open ocean region. By this approach the conservation /modelling of the SPM flux in the back barrier island region and in front of the islands is incorporated. This also allows keeping the number of particles in the domain fixed. If particles cross the northern boundary, they get a new position randomly chosen in the computational domain.

4.3 Numerics

The forcing of the Lagrangian model is taken from GETM. This hydrodynamic model provides the full velocity field (u, v, w), the bottom stress τ_b, the sea surface elevation ζ and also the vertical eddy viscosity (based on a k–ϵ model), which is set equal to the sediment diffusivity. To compute the physical properties at the position of the particle, a trilinear interpolation is used. In the vertical, the velocity profiles in the bottom boundary layer are computed by logarithmic interpolation. The vertical diffusivity K_Z is locally (in the grid box of the particle) approximated by a cubic fit. This also gives an analytic expression for the local vertical gradient.

The flow field is computed by GETM in advance and the whole fields were written out every 15 min (this was a trade-off between accuracy and storage capacity). An experiment with a time step of 5 min (not shown here) did not significantly change the outcome of the Lagrangian model. The particle-tracking algorithm uses a time step of 60 s in the horizontal, with a linear interpolation of the external 15 min GETM fields. In the vertical, a time step of 3 s is used, also with linear interpolation of the external fields. The horizontal time step is chosen such that a particle does not cross two grid cells during a time step (by advection and diffusion). Visser [35] proposed a much stronger restrictions on the time step for the integration. The gradient correction in Eq. 3 is valid only if the integration time step is sufficiently short:

$$\Delta t_z \leq \min\left(\frac{1}{\partial_z^2 K_Z}\right) \tag{33}$$

Because a constant horizontal diffusivity ($K_H = 1\,\mathrm{m}^2/\mathrm{s}$) is used, this inequality only limits the vertical time step. With the chosen time step, this requirement is fulfilled.

The Lagrangian model runs in a parallel version, where the computational load is distributed over four CPUs. For each sediment class 25×10^6 particles are released in the computational domain where the depth is less than 8 m. This can be estimated to be the depth of closure. The model run starts at the 22 October 2006 to allow for a sufficient spin up.

4.4 Model validation

To validate the numerical modelling system with respect to concentrations of SPM, time series of SPM were measured/estimated (at the time series station, see Fig. 5c) via an optical method (multispectral transmissiometer), which is described in more detail in [25]. The measurements were taken at a fixed water depth, 1.5 m below low tide.

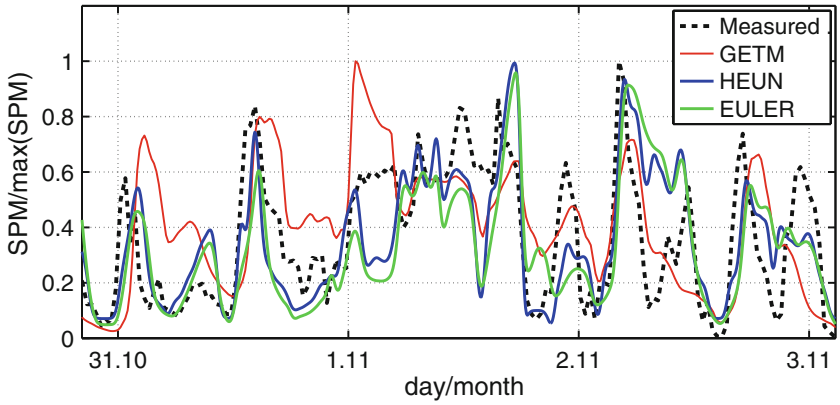


Fig. 7 Modelled and measured concentrations of suspended matter at the pile station during storm *Britta* in November 2006. The data were normalised to indicate the temporal dynamics. Measured data using a multispectral transmissiometer are also depicted. Shown are the results of the HEUN and EULER method. Additionally a time series simulated by the Eulerian SPM module of GETM is shown [17]

Table 2 Error measures for the normalised simulated time series

	EULER	MILSTEIN	HEUN	GETM SPM
ME	−0.036	−0.027	−0.012	0.032
RMS	0.223	0.210	0.187	0.229
r^2	0.271	0.320	0.437	0.293

In Fig. 7, a comparison of modelled SPM concentration and measurement at the pile station is shown. Further a time series of SPM is plotted, computed by the Eulerian SPM module of GETM [17]. Because the Lagrangian SPM module deals with a finite number or particles, the comparison with the measurements is made by computing the normalised SPM concentration, scaled in the interval [0,1]. To keep the figure clear, the time series of the MILSTEIN scheme is skipped. Because a constant horizontal diffusivity is used, the MILSTEIN scheme reduces to the EULER scheme (in the horizontal plane). Further, it is placed approx. between the HEUN and the EULER scheme and the additional curve would not give any new insight.

Figure 7 shows, that the Lagrangian model can reproduce the measurements quite well. The phase and the magnitude are matched. Comparing the performance of the EULER and HEUN scheme, the EULER scheme underestimates the peaks. Further, the scheme tends to smooth the dynamics. The differences in phase are nearly vanishing.

To quantify the performance of the different schemes, better than by visual inspection, in Table 2 different error measures are given. Because at the current state of sediment transport modelling, commonly accepted validation methods are still lacking, three simple error measures are used [38].

The *mean error* (ME) is the mean of the differences between predicted Y and observed X values. It is a measure of the overall forecast bias: a positive ME indicates that the model results on average overestimate the measured data. It is a measure of the overall transport and is independent of lag errors or amplitude errors. The ME is important to quantify budgets.

$$\mathrm{ME} = \frac{1}{N} \sum y_i - x_i \tag{34}$$

The *root mean-squared error* (RMS) is the most commonly used measure of numeric prediction quality:

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum(y_i - x_i)^2} \tag{35}$$

A common method to determine the linear relation between two series of X and Y is the linear correlation coefficient, r, given by the covariance σ_{xy} of X with Y, divided by the product of the standard deviation σ_x of X and the standard deviation σ_y of Y:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{36}$$

The statistical measure r^2 (or explained variance) is derived by squaring the correlation coefficient, which can be understood as the percentage of variance that the two series have in common.

To have a valid assessment of absolute errors, the accuracy of the field data should also be taken into account. Bartholomä et al. [2] estimated the uncertainty in their measurements to a factor of 2. They further stated that the absolute concentrations in measured suspended matter are presumably underestimated.

The ME indicates that the three particle tracking schemes underestimate the overall transport, but this might be caused by the normalisation. Further, the HEUN scheme has the smallest bias. The same holds for the RMS. The HEUN scheme performs best, whereas the EULER scheme has the largest RMS. r^2 also highlights the performance of the HEUN algorithm. The comparison of the EULER and MILSTEIN scheme shows that the later one has the better scores. Both schemes deviates only in the discretisation of the vertical particle movement, thus the impact of the higher order scheme (in the strong sense) in the treatment of boundaries is visible.

Interesting to note is the difference between the Eulerian and Lagrangian approach. The GETM SPM module overestimates the total transport (ME), but this is in agreement with the statement of Bartholomä et al. [2]. The Eulerian approach shows a phase lag of approx. 1 h, which is not visible in the Lagrangian approach. This also explains the poor performance in Table 2. Further the results indicate that the erosion of sediment (and distribution in the water column) in the Eulerian module is slower than in the Lagrangian approach. Moreover, the particle tracking schemes show sharper peaks. This is quite interesting, because both models are based on the same hydrodynamics. A possible explanation is that although GETM uses higher order schemes to solve the PDE of sediment concentration, numerical diffusion might be an issue. However, this paper is not intended to explain the difference in the outcome of an Eulerian and Lagrangian model. This might be the content of a forthcoming publication. This study intended to demonstrate that a particle-tracking scheme could reproduce short-term SPM dynamics with high fidelity.

5 Conclusion

In the first part of the article, the underlying theory to translate SPM dynamics from the Eulerian framework to the Lagrangian picture was present and summarised. The particle tracking schemes for modelling SPM dynamics are quite attractive, because they give a straightforward physical interpretation of the processes involved. Moreover, due to the easy

parallelisation of the Lagrangian schemes, this approach is well suited to run on massive parallel computers, to improve the reliability and accuracy of the results.

Various stochastic numerical schemes are presented to find the best suitable scheme for implementation in a 3-D particle-based SPM transport model. To validate the schemes and show their performance, analytical test cases with a spatially varying diffusion coefficient have been investigated. All of the three idealised test cases were performed with different numerical schemes: EULER, MILSTEIN and HEUN. Even though the MILSTEIN scheme only requires a minor additional effort compared with the EULER scheme, the changes in terms of higher accuracy and faster convergence are rewarding. A two-step scheme like the HEUN scheme leads to a further improvement of the results. The set of experiments that was carried out indicates that any improvement over the EULER scheme is welcome. Even though this is not trivial, and some care needs to be taken in the choice and application of the scheme, such a scheme does not necessarily need to be complicated or expensive. As advanced as higher order Lagrangian particle transport models for dispersion in turbulent flow sometimes are, it is surprising to see how many of these models still use the EULER scheme, while minor adaptations may greatly improve the accuracy of the model. The analysis also showed, that especially the term $\nabla \cdot \mathbf{K}$ in Eq. 3 needs special care. The effort, to improve the estimates of the gradient, results in performance gains. Further the validation of the three tracking schemes revealed, that although they have by definition the same order of convergence, differences showed up.

In the second part of this article, a more realistic scenario was studied. The particle-tracking scheme was applied to model SPM dynamics in the East Frisian Wadden Sea during the storm *Britta* (31 October 2006 to 2 November 2006). Here, the bottom boundary condition for deposition/erosion needed some modification. The classical Partheniades [21] type flux formulation was adapted to the Lagrangian framework. Based on the bottom shear stress τ_b and the critical shear stress τ_c, a characteristic eddy size in the bottom boundary layer was computed, which lead to a lift up of particles into the water column.

The comparison of the modelled SPM concentration with measured time series showed, that the particle-tracking scheme could reproduce the dynamics. By comparing the results of an Eulerian SPM model and the particle tracking schemes, different dynamics were visible, although both models are based on the same hydrodynamic flow field. These differences may caused by challenges in the numeric (numerical diffusion) or still lacking knowledge in SPM dynamics.

Acknowledgements We would like to acknowledge Karsten Lettmann, who provided the setup for the Wadden Sea test case and the two reviewers, who helped to improve this paper.

References

1. Arnold L (1974) Stochastic differential equations. Wiley, New York
2. Bartholomä A, Kubicki A, Badewien TH, Flemming BW (2008) Suspended sediment transport in the German Wadden Sea—seasonal variations and extreme events. Ocean Dyn 59(2):213–225. doi:10.1007/s10236-009-0193-6
3. Booij N, Ris RC, Holthuijsen LH (1999) A third-generation wave model for coastal regions 1. Model description and validation. J Geophys Res 104:7649–7666
4. Brickman D, Smith PC (2001) Lagrangian stochastic modeling in coastal oceanography. J Atmos Ocean Technol 19:83–99
5. Burchard H, Bolding K (2002) GETM a general estuarine transport model. Sci Doc EUR 20253 EN. Institute for Environ and Sustainability, Ispra
6. Burchard H, Petersen O, Rippeth TP (1998) Comparing the performance of the k-ϵ and the Mellor-Yamada two-equation turbulence models. J Geophys Res 103:10543–10554

7. Burchard H, Bolding K, Umlauf L (2007) GETM—General Estuarine Transport Model, source code and test case documentation. http://getmeu/indexphp
8. Cencini M, Bec J, Biferale L, Boffetta G, Celani A, Lanotte AS, Musacchio S, Toschi F (2006) Dynamics and statistics of heavy particles in turbulent flows. J Turbul 7:36
9. Charles WM, van den Berg E, Lin HX, Heemink AW, Verlaan M (2008) Parallel and distributed simulation of sediment dynamics in shallow water using particle decomposition approach. J Parallel Distrib Comput 68:717–728
10. Chung TJ (2002) Computational fluid dynamics. Cambridge University Press, Cambridge
11. Deleersnijder E, Beckers JM, Delhez EJM (2006a) The residence time of settling in the surface mixed layer. Environ Fluid Mech 6:25–42
12. Deleersnijder E, Beckers JM, Delhez EJM (2006b) On the behaviour of the residence time at bottom of the mixed layer. Environ Fluid Mech 6:541–547
13. Gayer G, Dick S, Pleskachevsky A, Rosenthal W (2006) Numerical modelling of suspended matter transport in the North Sea. Ocean Dyn 56:62–77
14. Hunter JR, Craig PD, Phillips HE (1993) On the use of random walk models with spatially variable diffusivity. J Comput Phys 106:366–376
15. Kloeden PE, Platen E (1992) Numerical solutions of stochastic differential equations. Springer, Berlin
16. Krestenitis YN, Kombiadou KD, Savvidis YG (2007) Modelling the cohesive sediment transport in the marine environment: the case of Thermaikos Gulf. Ocean Sci 3:91–104
17. Lettmann K, Wolff J-O, Badewien TH (2009) Impact of wind and waves on suspended particulate matter fluxes in the East Frisian Wadden Sea (southern North Sea). Ocean Dyn 59(2):239–262. doi:10.1007/s10236-009-0194-5
18. Maier-Reimer E, Sündermann J (1982) On tracer methods in computational hydrodynamics, engineering applications of computational hydraulics, vol 1. Pitman Advanced Publishing Program, Boston, pp 198–217
19. Matsumoto M, Nishimura T (1998) Mersenne Twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans Model Comput Simul 8(1):19
20. North EW, Hood RR, Chao S-Y, Sanford LP (2006) Using a random displacement model to simulate turbulent particle motion in a baroclinic frontal zone: a new implementation scheme and model performance tests. J Mar Syst 60:365–380
21. Partheniades E (1965) Erosion and deposition of cohesive soils. Proc ASCE. J Hydraul Div 91:105–139
22. Pleskachevsky A, Gayer G, Horstmann J, Rosenthal W (2005) Synergy of satellite remote sensing and numerical modelling for monitoring of suspended particulate matter. Ocean Dyn 55:2–9
23. Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) Numerical recipes: the art of scientific computing. Cambridge University Press, Cambridge
24. Proehl JA, Lynch DE, McGillicuddy DJ Jr, Ledwell JR (2005) Modelling turbulent dispersion of the North Flank of Georges Bank using Lagrangian methods. Cont Shelf Res 25:875–900
25. Reuter R, Badewien TH, Bartholomä A, Braun A, Lübben A, Rullkötter J (2009) A hydrographic time-series station in the Wadden Sea (southern North Sea). Ocean Dyn 59(2):195–211. doi:10.1007/s10236-009-0196-3
26. Rolinski S, Umgiesser G (2005) Modelling short-term dynamics of suspended particulate matter in Venice Lagoon, Italy. Estuar Coast Shelf Sci 63:561–576
27. Soulsby R (1997) Dynamics of marine sands, a manual for practical applications. Thomas Telford, London
28. Spivakovskaya D, Heemink AW, Deleersnijder E (2007) Lagrangian modelling of multi-dimensional advection-diffusion with space-varying diffusivities: theory and idealised test cases. Ocean Dyn 57:189–203
29. Stanev EV, Wolff J-O, Burchard H, Bolding K, Flöser G (2003) On the circulation in the East Frisian Wadden Sea: numerical modelling and data analysis. Ocean Dyn 53:27–51
30. Stanev EV, Wolff J-O, Brink-Spalink G (2006) On the sensitivity of sedimentary system in the East Frisian Wadden Sea to sea level rise and magnitude of wind waves. Ocean Dyn 56:266–283
31. Stanev EV, Flemming BW, Bartholomä A, Staneva JV, Wolff J-O (2007) Vertical circulation in tidal inlets and back-barrier basins. Cont Shelf Res 27:798–831
32. Stanev EV, Brink-Spalink G, Wolff J-O (2007) Sediment dynamics in tidally dominated environments controlled by transport and turbulence: a case study for the East Frisian Wadden Sea. J Geophys Res 112:C04018
33. Staneva J, Stanev EV, Wolff J-O, Badewien T, Reuter R, Flemming B, Bartholomä A, Bolding K (2008) Hydrodynamics 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
34. Stijnen JW, Heemink AW, Lin HX (2006) An efficient 3D particle transport model for use in stratified flows. Int J Numer Methods Fluids 51:331–350

35. Visser AW (1997) Using random walk models to simulate the vertical distribution of particles in a turbulent water column. Mar Ecol Prog Ser 158:275–281
36. Visser AW (2008) Lagrangian modelling of plankton motion: from deceptively simple random walks to Fokker-Planck and back again. J Mar Syst 70:287–299
37. Warner JC, Sherwood CR, Arango HG, Signell RP (2005) Performance of four turbulence closure models implemented using a generic length scale method. Ocean Model 8:81–113
38. Winter C (2007) On the evaluation of sediment transport models in tidal environments. Sediment Geol 202:562–571

Environ Fluid Mech (2010) 10:41–76
DOI 10.1007/s10652-009-9153-4

ORIGINAL ARTICLE

Anti-diffusion and source identification with the 'RAW' scheme: A particle-based censored random walk

R. Ababou · A. C. Bagtzoglou · A. Mallet

Received: 1 March 2009 / Accepted: 14 September 2009 / Published online: 26 September 2009

Abstract This paper presents a novel methodology for time reversal in advective-diffusive pollutant transport in groundwater systems and other environmental flow systems (specifically: time reversal of diffusive terms). The method developed in this paper extends previous particle-based approaches like the Reversed Time Particle Tracking Method of Bagtzoglou [6]. The reversal of the 'diffusive' and/or 'macrodispersive' component of pollutant migration is especially under focus here. The basis of the proposed scheme for anti-diffusion is a continuous time, censored, non-local random walk capable of tracking groundwater solute concentration profiles over time while conserving the (reverse) Fickian properties of the anti-diffusing particle cloud in terms of moments. This scheme is an alternative to the direct solution of the eulerian concentration-based diffusion PDE, which is notoriously unstable in reverse time. Our analysis leads to the conclusion that an adaptive time stepping scheme—with decreasing time step—is necessary in order to maintain a constant amount of anti-diffusion (the reverse form of Fick's law). Specifically, we study the relations between the following parameters: time step evolution vs. time (or vs. number of steps); variance evolution (decrease rate); total time (or number of steps) required to reach a fully anti-diffused solution. The proposed approach is shown to be quite efficient; typically, for every ten time steps, one to two orders of magnitude reduction of the dispersion width of the plume can be attained. Furthermore, the method is shown to be asymptotically exact for reverse *Fickian* diffusion. The method is applied with success to several situations involving the diffusive transport of a conservative solute in the following cases: (i) Single source recovery in one-dimensional space with constant diffusion parameters (this example serves as a validation test for the theory); (ii) Single source recovery in two-dimensional space with constant isotropic diffusion (this example also serves as a test for the theory); (iii) Multiple source recovery

R. Ababou (✉) · A. Mallet
Institut de Mécanique des Fluides de Toulouse, 1 Allée du Professeur Camille Soula, 31400 Toulouse, France
e-mail: ababou@imft.fr

A. C. Bagtzoglou
Department of Civil and Environmental Engineering, University of Connecticut, 261 Glenbrook Road, U 2037, Storrs, CT 06269-2037, USA

in two-dimensional space, assuming isotropic diffusion. It is expected that the methodology tested in this paper is applicable more generally to complex environmental pollution problems involving multiple sources, anisotropic hydrodynamic dispersion, and space-time variable advection-diffusion flow systems; the modeling of reverse diffusion/dispersion in such systems is currently under investigation.

Keywords Source identification · Inverse problem · Diffusion · Dispersion · Particle tracking · Random Walk · Advection-diffusion · Heterogeneous systems · Pollution · Back-tracking · Back-diffusion · Point source · Solute transport · Groundwater

1 Introduction

In the context of water resources, and in view of increasing demands for clean drinking water, it is highly desirable to identify pollution sources accurately. The objectives are to backtrack the pollution source, to recover the spatial extent the plume at different times, and ultimately, to reconstruct the contamination plume release time history. These objectives can be formulated mathematically as a PDE-driven advection-diffusion *inverse problem*. Its solution implies time reversal of both advective and diffusive processes with generalized diffusion terms representing hydrodynamic dispersion processes as well. However, backtracking diffusion/dispersion processes via PDE models is particularly problematic, as diffusive PDE's become ill-posed when time is reversed.

Direct approaches that perform time inversion in the context of contaminant source identification have enjoyed visibility lately. Reviews of time inversion methods for 'source identification inverse problems' can be found in Morrison [25–27], Atmadja and Bagtzoglou [4,5], Bagtzoglou and Atmadja [7,8]. However, the widespread use of some of these methods is still limited due to their non-conventional nature.

The difficulty of the source identification problem is compounded by the fact that geologic media and the associated flow systems are highly variable (see Ababou [1] for a recent review on heterogeneous aquifers and stochastic approaches to groundwater flow modeling). Other factors may increase the complexity of the source identification problem: the pollution source may be 'multiple' or spread out (rather than localized); it may also be continuous (rather than instantaneous), mobile (rather than fixed), etc.

Similar problems occur in other environmental contexts, e.g., marine and aquatic pollution from large commercial or military ships (backtracking of oceanic oil spills, for instance)—but also in other tightly related problems, such as the optimal identification of individual storm cells in rainfall radar measurements.

The objectives of this paper are to contribute, first, to the problem of source identification in terms of its unknown initial position and past history, e.g., the unknown time of initial release for an instantaneous source. Specifically, the purpose of this paper is to present a novel methodology for time reversal in tracer transport simulations driven by diffusive PDE's, particularly for applications to groundwater pollution.

The method developed towards this objective is a special kind of particle tracking scheme, building on the previous Reversed Time Particle Tracking Method or 'RTPTM' (Bagtzoglou [6]). The new method is dubbed 'RAW' for 'Reverse Antidiffusive Walk' or 'Random Antidiffusive Walk'. It is based on a continuous-time, non-local, censored random walk. Contrary to the RTPTM of Bagtzoglou [6], the RAW scheme does not require adaptively increasing particle resolution, which may render the RTPM scheme computationally burdensome. The

RAW scheme has other desirable properties, which are analyzed theoretically and tested numerically in the remainder of this paper.

The novel RAW scheme is developed in this paper under two guises:

(i) a 'non local' or 'gobal' RAW scheme that can back-diffuse a single gaussian-type plume at the proper rate towards its initial point source; and
(ii) a 'local' and adjustable RAW scheme that can deal with multiple plumes or non-point sources.

The study is mainly organized as follows:

- First, the non local RAW scheme is analyzed theoretically in the one-dimensional case. It is then tested numerically in 1D, and further extended and tested numerically in 2D.
- The non local theory is then used to provide some insights for designing a more flexible and more general local back-diffusion scheme.
- Finally, the local RAW scheme is finally investigated and tested numerically in 1D and 2D in the last part of the paper in order to address the problem of identifying multiple contaminant sources (e.g., from the observation of a multi-modal plume).

It should be emphasized that the reverse particle tracking scheme is treated in this paper assuming that all the transport parameters are known, i.e., velocity and diffusion are assumed known in advance: only the pollution source (unique or multiple) is unknown. However, the final objective is to combine, in future, the proposed "RAW" scheme with a sequential estimation of past transport parameters based on present-day observations of concentrations, for instance in the same manner as in the Eulerian MJBBE (Marching-Jury Backward Beam Equation) of Bagtzoglou and Atmadja [7,8].

2 Literature review

2.1 Review on diffusion, particles and random walks

In Lagrangian particle methods, solute concentration is discretized into "concentration packets" called "particles" or "pseudo-particles". These methods have been used with success, as an alternative to eulerian PDE's, for modeling advective-diffusive transport processes in environmental flows, particularly in the case of groundwater and soil pollution : Uffink [36], Kitanidis [20], Tompson [35], Ababou et al. [2], Fadili et al. [16,17], Bagtzoglou et al. [10], Spiller et al. [32–34], Delay et al. [15], La Bolle et al. [21,22] among many other valuable contributions and reviews.

In these approaches, diffusion is modeled with a gridless, discrete time random walk of particles in continuous space. Each particle follows a time-discretized version of the Wiener process in multi-dimensional space. This basic scheme has been extended with success to the case of non-spherical, anisotropic, and spatially variable diffusion tensors by various authors including those quoted above.

For completeness, we note that other particle methods have been proposed to model reaction-diffusion problems, e.g. random walks with fixed grids and/or with stochastic time. For example, isotropic diffusion in the Markov Automaton presented in Spiller et al. [32–34] is modeled with a Particle-in-Cell approach, stochastic in continuous time, but with fixed size displacement steps in space. In the present work on reverse diffusion modeling, we seek a Lagrangian, gridless, continuous space approach to diffusion, essentially based on (modifications of) the Wiener process.

One of the first authors who demonstrated the 'equivalence' between the continuous space-time Wiener process and the macroscopic advection-diffusion PDE (Fokker-Plank equation) was Chandrasekhar (see Wax [38]'s collection of papers). However, even in the simple case of isotropic diffusion in an infinite domain, this 'equivalence' holds only *approximately,* under conditions defined by Pawula's lemma. This lemma serves as a justification for truncating the Kramer-Moyal infinite expansion of the probabilistic 'Master Equation' to just the first two terms (advection and diffusion). For more mathematical details, the reader is referred to Wax [38], Van Kampen [37], Oppenheim et al. [31], and to the handbook of Gardiner [18].

See also Ababou et al. [2] & Fadili et al. [17] for a review and a study of the 'Moment Inverse Problem', addressing the calculation of the spatial concentration distribution C(x,y,z,t) or of the PDF of particle positions from particle moments $M_n(t)$ of increasing order.

The main consequence of the above-cited theorems is that, in practice, under general conditions, the PDF of particle positions in a random walk process may not be governed exactly by the diffusion PDE. In practice, the PDF/PDE equivalence could hold only for *a few* moments of particle positions, and it is usually assumed that the equivalence holds at least up to second order moments, including the total mass $M_0(t)$, the mean displacement $\underline{M}_1(t)$, and the dispersion (co)variances $\underline{\underline{M}}_2(t)$.

These theoretical remarks lead us to approach with caution the relation between a particle-based random walk (describing the 'microscopic' process) and its 'equivalent' *Fickian* diffusion PDE (describing the corresponding macroscopic process). This distinction will help put in perspective the new particle-based anti-diffusive 'RAW' scheme developed in this paper. Before doing so, we briefly summarized related approaches in the literature, as far as Lagrangian particle models are concerned.

2.2 Review of Reverse Time Particle Tracking Method (RTPTM) for source identification

We briefly review a precursor particle scheme for reverse time modeling of source diffusion (and advection). For more details, see Bagtzoglou [6], Bagtzoglou and Atmadja [7,8].

The Reversed Time Particle Tracking Method (RTPTM) modifies the classical Random Walk Particle Tracking Method (RWPTM) displacement equation (or Wiener step) by introducing a *minimization approach,* that is, a minimization of the variance or second spatial moment of particle positions.

In classical Fickian diffusion (e.g. molecular diffusion), the macroscopic diffusion process can be expressed in terms of the spatial variance, or dispersion variance of the particle cloud as

$$\frac{d(\sigma^2)}{dt} = 2D, \tag{1}$$

where σ^2 is the (scalar) variance of particle positions, or the second order moment of the concentration field. More generally, in multi-dimensional space, the dispersion variance *matrix* is computed as the centered second order moment *matrix* normalized by the center of mass. Here, it is denoted Σ^2 (by analogy with the variance), and it is defined by:

$$\begin{aligned}\Sigma^2(t) &= \frac{1}{M(t)} \int_{\Omega} C(\mathbf{X}, t)(\mathbf{X} - \overline{\mathbf{X}}(t))(\mathbf{X} - \overline{\mathbf{X}}(t))^T d\Omega \\ &= \sum_{p=1}^{N(t)} \tilde{m}_p(t)(\mathbf{X}_p(t) - \overline{\mathbf{X}}(t))(\mathbf{X}_p(t) - \overline{\mathbf{X}}(t))^T \end{aligned} \tag{2}$$

- $N(t)$ is the total number of particles present in the system;
- $M(t) = \sum_{p=1}^{N(t)} m_p$ is the total solute mass within domain Ω;
- $\overline{\mathbf{X}}(t)$ is the position vector of the plume center of mass;
- $\mathbf{X}_p(t)$ is the position vector of each individual particle 'p';
- $\tilde{m}_p(t) = m_p/M(t)$ is the normalized mass of particle 'p'.

When diffusive solute transport advances forward in time (as is normally the case), D is a definite positive matrix, implying that the 3D spread of solute about the center of mass is increasing.

It seems, from the Lagrangian particle-based equations, that negative diffusion and time reversal can be implemented 'simply' by decreasing the spatial spread of the particle cloud around its center of mass. At least, this is the view taken in the above-mentioned RTPTM scheme. Accordingly, the step equations for the Reverse Time Particle Tracking Scheme (RTPTM) are (Bagtzoglou et al., *op. cit.*):

$$\mathbf{X}^n = \overline{\mathbf{X}}^n + \mathbf{F}(\overline{\mathbf{X}}^n, t_n)\,\mathbf{Q}(t_n) \quad \overline{\mathbf{X}}^n = \overline{\mathbf{X}}^{n+1} - \mathbf{A}(\overline{\mathbf{X}}^{n+1}, t_{n+1})\Delta t \tag{3a}$$

$$\mathbf{F}(\overline{X}^n, t_n)\mathbf{F}^T(\overline{X}^n, t_n) = \mathbf{\Sigma_n^2} \quad \mathbf{\Sigma_n^2} = \mathbf{\Sigma_{n+1}^2} - 2\Delta t\mathbf{D} \tag{3b}$$

where all quantities are either vectors (**X, Q, A**) or second rank tensors (**F, D**, (2). Vector Q(tn) is a random vector defined as:

$$\mathbf{Q}(t_n) = \mathbf{Q}^n = \mathbf{R}^n\mathbf{S}^n,$$

with $\mathbf{R}^n$ a normalized zero-mean unit-variance random vector, and $\mathbf{S}^n$ a random sign ($\pm$) vector, both defined at time t_n. Vector '**A**' is the advective component, appearing with a negative sign for reverse advection. It is equal to the eulerian velocity **V** plus a drift term to account for spatially variable diffusion coefficient **D**(**x**). Thus, in models of subscale hydrodynamic dispersion, **D** is a tensorial, velocity-dependent, and by the same token, spatially variable coefficient, and **A** is given by:

$$\mathbf{A} = \mathbf{V} + div(\mathbf{D}).$$

Equation set (3) constitutes the backbone of the reverse time (RTPTM) scheme of Bagtzoglou et al. The scheme above is multidimensional (3D) and is easily reduced to a 2D or 1D form. The method can be also applied to forward time simulations of classical advection-diffusion problems, by rearranging these equations, yielding results equivalent to the classical Random Walk Particle Tracking Method (RWPTM). In fact, comparisons between the forward and reverse versions led to a set of relations that the **F** and **Q** vectors must satisfy (for more details, see Bagtzoglou et al., *op. cit.*).

2.3 Review of other methods: PDF back-propagation, adjoint equations, image deblurring

Other methods have been implemented in various areas for objectives such as the deblurring of images (image analyzis and processing), and also, the backward propagation of the probability of particle positions using adjoint or backward equations (groundwater contamination applications, and also, theory of statistical turbulence). We briefly review some of these works, pointing out the analogies, but also the differences with the matter at hand.

2.3.1 Image processing literature

First, let us focus on one example from the *image processing literature*. This type of application is particularly relevant here, since our inverse source identification method bears some

relation to deblurring or refocusing in image processing. However, note that the aims are a bit different: in image processing, it is not necessary to anti-diffuse at the correct rate, whereas in environmental pollution problems, it may be important to identify not only the position of the source, but also its history (e.g., date of release).

Let us consider briefly the method used by Li et al. [23] for deblurring gaussian images. These authors recognize that it is very difficult to solve the anti-diffusion problem because it is ill-posed. They propose instead an indirect method based on a pyramidal decomposition of the image into spatial scales, and they implement a time-limited anti-diffusion process on each scale (the allowable duration of their antidiffusion scheme is limited, precisely, because of the ill-posedness of the problem).

In comparison, as will be seen, our particle-based antidiffusion scheme is not time-limited; it is properly labeled in time, and it operates directly on the entire plume (or image). Based on the observed plume, our algorithm is able to obtain successfully the (single or multiple) unknown source locations after a finite backward time, approximating closely the true date of release of the source.

2.3.2 Statistical turbulence literature

In the area of statistical fluid mechanics (turbulence), an elegant and often quoted paper by Corrsin [14] adresses the problem of back-propagating the uncertainty or variability of particle locations. This very short (2 page) paper is interesting, but not directly relevant for our anti-diffusion problem. We explain this in more detail in order to point out possible confusions with the various objectives of 'back-propagation' in the literature.

Corrsin's 1972 paper focuses on a re-interpretation of hydrodynamic dispersion in a purely advective flow field V(x,t), e.g. in *stationary isotropic turbulence* (Taylor, Batchelor). Instead of studying the dispersion of arrival positions X(xo,t) from a fixed starting position (xo), he studies the so-called "backward dispersion" (this is "possibly a better phrase than reverse diffusion" in his own statement—but in our view both terms are misleading).

The "backward dispersion" in Corrsin's paper is, in fact, the dispersion of those departure points Xo(x,t) which all lead to the same arrival point (x) in a fixed time (t). The author obtains an equivalence between this dispersion of departure points, and the classical dispersion of arrival points. He interprets the results in the framework of statistical turbulence in terms of velocity covariance functions. Particle trajectories undergo advective transport without diffusion (there is no Wiener process involved). The aim of his paper is to re-interpret turbulent dispersion, not to identify a source.

2.3.3 Environmental and groundwater literature

There are a number of potential applications of antidiffusive schemes in environmental fluid dynamics, e.g., to name a few: source identification and inverse problems in groundwater pollution; backtracking of oceanic oil spills; discrimination of storm cells in rainfall radar measurements; and transit time distributions in atmospheric and ocean circulation models.

For instance, regarding the latter, Haine et al. [19] study the backward propagation of a two-time PDF to build a boundary response function that serves for recovering the space-time history of pollution C($\mathbf{x}$, t) from an observed time history of localized sea surface pollution ($C_{\Omega SEA}(t)$). However, from now on, let us focus specifically on the groundwater literature.

The 'back propagation' of observed concentrations is an important topic in groundwater pollution problems: this question arises in various inverse problems, depending on the

assumed observations, and it also intervenes in sensitivity analyses (sensitivity of contamination with respect to source properties, aquifer properties, etc.).

But first, a few technical points may require clarifications (cf. Fadili and Ababou [2] for details):

- Concentrations are often converted into discrete particles or markers with specified masses. Each 'particle' can be interpreted as a concentration packet (infinitesimal or not). An infinitesimal particle located at $\mathbf{x} = \mathbf{X}(t)$ at time 't' has infinitesimal mass $dm(X(t)) = C(\mathbf{x},t)\, d\omega$ where C is concentration (kg/m^3), $d\omega$ the elementary volume (m^3).
- There is a profusion of terminology concerning "PDF's" (Probability Density Functions). The so-called 'location PDF' is the PDF of particle positions $\mathbf{X}(t)$ at any given time 't'. It should be denoted $f_X(\mathbf{x}, t)$. It is equal to concentration $C(\mathbf{x},t)$ divided by total mass M(t).
- Also, PDF's are abusively assimilated to 'probabilities' in some works. For simplicity, let us consider the 1D case. The PDF $f_X(\mathbf{x}, t)$ can be defined in terms of probability as follows:

$$\Pr\{x \le X(t) \le x + dx\} = f_X(x,t)dx.$$

In Neupauer and Wilson [28], an adjoint operator method is used to back-propagate the PDF ($f_x(X, \tau)$) of past positions X of a groundwater contaminant (see also Neupauer and Wilson [29] and [30]). This so-called 'backward probability' represents the probability (density) of a particle being located at position X at backward time $\tau = t_{Td}$, given that it was detected with certainty at backward time $\tau = 0 (t = Td)$ at a fixed observation point (or pumping well) located at $x = Xd = 0$.

Remarkably, the backward PDF of Neupauer and Wilson (*op. cit.*) is governed by an adjoint advection-diffusion equation, with (t) replaced by (τ), (V) replaced by ($-$V), but the diffusion term remaining unchanged. The equation is solved with initial condition specified at backward time $\tau = 0$. As a result, the variance of this 'backward PDF' increases with backward time (τ). This can be explained as follows: during back-propagation, the uncertainty of particle positions (X) increases continually with backward time $\tau = t - Td$, starting with a zero variance at detection time ($\tau = t - Td = 0$).

In contrast, the antidiffusion scheme presented in this paper attempts to reconstitute a 'crisp' initial position of the pollutant source, given an observation of the entire spatially distributed plume at a given observation time $t = Td$. To achieve this, it is necessary to tackle directly the anti-diffusion problem with (D) changed to ($-D$): thus, our objective is different. It bears some similarities with other applications such as image deblurring (see earlier comments concerning Li et al. [23]).

Finally, Milnes and Perrochet [24] tackle the problem of source identification by using an adjoint backward time PDE with velocity reversed, but keeping the diffusion operator positive (much as Neupauer and Wilson, *op. cit.*). Furthermore, to achieve their purpose, they also use a result established earlier by Cornaton and Perrochet [12,13] in the context of modeling of groundwater age, transit times, and life expectancy (Cornaton [11]). The latter result can be summarized as follows (in our notations):

$$f_{\mathbf{X}}(\mathbf{x}_1, t) = f^*_{\mathbf{X}}(\mathbf{x}_0, \tau)$$

where $f_X(\mathbf{x}_1, t)$ is the forward PDF evaluated at downstream location x_1 in response to a point source release at $\mathbf{x}_0$, and $f^*_X(\mathbf{x}_1, t)$ is the backward PDF evaluated at upstream location $\mathbf{x}_0$ in response to a point source release at $\mathbf{x} = \mathbf{x}_1$ (f* is f with $\mathbf{V}$ changed into $-$V). Note that this identity is also related to Corrsin's 1972 analyzis (reviewed earlier). Now, assuming that all

advection-diffusion parameters are known, and assuming a single ideal point source, Milnes and Perrochet used the above identity to identify the source by shrinking the concentration contour of the plume during backward simulation.

In comparison with the present Reverse Antidiffusive Walk, their scheme is based on back-propagating concentration iso-values, and therefore, as the authors recognize, it cannot be considered as a direct physical shrinkage of the pollution plume back to its source. For these reasons, their scheme cannot be straightforwardly generalized to multiple sources, as the authors also recognize. In contrast, we will show in the sequel that the RAW scheme can handle unknown multiple sources.

For completeness, let us mention that Milnes and Perrochet's scheme is readily applicable to cases with (known) spatially variable transport coefficients, as their test case demonstrates, with the only *proviso* that a single ideal point source must be assumed.

3 Reverse Anti-diffusive Walk (RAW) Theory

The novel scheme developed in the remainder of this paper uses a "reverse", anti-diffusive particle random walk in a Lagrangian framework (as in the RWTPM scheme of Bagtzoglou [6]). The novel scheme is named "Reverse Anti-diffusive Walk" or "RAW" for short. Its main features were presented only summarily in earlier works (Bagtzoglou and Ababou [9]; Ababou et al. [3]).

The advective part of the transport problem will not be treated explicitly here. However, it is clear that to obtain a reverse advective-diffusive particle scheme, the reverse random walk method advocated in this paper should be implemented together with a reversed velocity field (−V instead of V), which does not cause any problem, contrary to reverse diffusion.

The main point is that the novel particle-based scheme (RAW) achieves *source identification* without having to deal with ill-posed or poorly conditioned inverse PDE problems. Moreover (contrary to the above described RTPTM of Bagtzoglou [6]) it does not require an adaptively increasing number of particles. As a consequence the computational requirements are less stringent and this enhances the ability to decrease the dispersion variance and properly identify the sources.

3.1 Preliminary formulation of anti-diffusive particle walk

The main idea underlying the proposed "RAW" scheme is that anti-diffusion could be modeled with a censored, nonlocal random walk, as defined below theoretically in continuous time. For instance, assuming for simplicity that the observed particle cloud emanates from a single diffusion source without advective transport, the following preliminary is proposed (to be revised later towards an appropriate, operational discrete time scheme).

The basis of the proposed scheme for anti-diffusion is a continuous time, censored, non local random walk, which could be formulated in multidimensional space as:

$$\mathbf{dX}(t) = \begin{cases} \sqrt{2D_0}\,\mathbf{dU}(t) = \sqrt{2D_0}\,\boldsymbol{\xi}(t)dt & \text{if} \quad S_{\mathrm{LOCAL}} = -S_{\mathrm{GLOBAL}} \\ \mathbf{0} \quad \text{otherwise} \end{cases} \tag{4}$$

where the local and global 'S' quantities of Eq. 4 appear as a type of censoring mechanism. In the case of 1D space, we propose the following censoring mechanism :

$$S_{\mathrm{LOCAL}} = Sign(dU(t)) \tag{5}$$

$$S_{\mathrm{GLOBAL}} = Sign(X(t) - \overline{X}(t)) \tag{6}$$

where dU(t) is the infinitesimal increment of a unit Wiener process U(t). 'Anti-diffusive' behavior is forced by the "minus" sign in '$S_{LOCAL} = -S_{GLOBAL}$'. This 1D scheme is non local due to the fact that local displacement depends on mean displacement of the center of mass, in Eq. 6. We will develop later on a 2D version of this scheme (see further below).

We now analyse the continuous time formulation of this scheme although we will see later that it is physically meaningful only if implemented in discrete time. First, the previous equations (4)-(6) can be reformulated as follows using the equilibrated Heaviside function:

$$H(x) = \begin{vmatrix} 0 & if\ x < 0 \\ 1/2 & if\ x = 0 \\ 1 & if\ x > 0 \end{vmatrix} . \tag{7}$$

Using also the identity $\mathbf{dX}(t) = \mathbf{X}(t + dt) - \mathbf{X}(t)$, leads to:

$$X(t + dt) = X(t) + \sqrt{2D_0}\, dU\ \{H(dU)(1 - H(\delta X)) + (1 - H(dU))H(\delta X))\} \tag{8}$$

where dU stands for dU(t), and δX is the deviation from the center of mass, given by:

$$\delta X(t) = X(t) - \overline{X}(t) \tag{9}$$

Recall that dU(t) is a random Gaussian variable (random increment) with zero mean and with a variance proportional to the time increment, that is:

$$Var(dU(t)) = dt \tag{10}$$

This can be expressed compactly as: *dU(t)* is *N(0, dt)*.

3.2 Reformulation of the general anti-diffusion algorithm in discrete time

We can now write the discrete-time form of the previously proposed algorithm as follows, taking into account the special properties of the Wiener increment. A simple derivation of the discrete time Wiener process is outlined in Appendix B. However, the reader is referred to textbooks on *Ito calculus* for more mathematical insights.

First, let us define:

$$\Delta W(t) = W(t + \Delta t) - W(t) \tag{11}$$

Note that the Wiener increment *dW(t)* differs by a factor $(2D_0)^{1/2}$ from the unit increment *dU(t)*. Therefore, we have in pseudo-code form:

$$\Delta W(t) = \sqrt{2D_0 \Delta t}\ \xi \tag{12}$$

where ξ is a unit white noise process—or an uncorrelated random N(0,1) sequence in discrete time.

The distance δX(t) to the center of mass of the plume (Eq. 9) can now be written in the form:

$$\delta X(t + \Delta t) = \delta X(t) + \Delta W(t)\ \Gamma\{\Delta W(t); \delta X(t)\} \tag{13}$$

where Γ is the censoring operator which allows only for 'reverse' displacements:

$$\begin{aligned} \Gamma\{\Delta W(t); \delta X(t)\} = {} & H(\Delta W(t))(1 - H(\delta X(t))) \\ & + (1 - H(\Delta W(t)))H(\delta X(t)) \end{aligned} \tag{14}$$

Note that Eq. 13 can also be written in terms of particle displacements and mean displacement. Together with a generalization to multidimensional space, the scheme can be finally expressed as:

$$\begin{aligned} &\mathbf{\Delta X}(t) = \mathbf{\Delta \overline{X}}(t) + \ \mathbf{\Delta W}(t)\ \Gamma\{\mathbf{\Delta W}(t); \mathbf{\delta X}(t)\} \\ &\text{where : } \mathbf{\Delta X}(t) = \mathbf{X}(t+\Delta t) - \mathbf{X}(t) \quad \text{and} \quad \mathbf{\Delta \overline{X}}(t) = \mathbf{\overline{X}}(t+\Delta t) - \mathbf{\overline{X}}(t). \end{aligned} \tag{13''}$$

Equation 14 describes the 'censoring' or 'reverse' operator Γ. The time-step Δt may be constant (Δt_0) or adaptive (Δt_n) in Eq. 13. Both constant time-step and adaptive time-step procedures will be described shortly. Note: if the operator Γ is replaced by the unit constant "1", and time step is maintained constant, the process modeled by Eq. 13 becomes the classical Wiener diffusion process.

3.3 Probabilistic analysis of the 1D anti-diffusion scheme

3.3.1 First order moment(s)

The first step of probabilistic analysis is to calculate the 1st order moments of particles. As stated earlier, the current approach can be later revised to account for reversed advection (*we continue to focus entirely on diffusion here*). Now, using *ensemble averaging*, and under certain conditions, the ensemble mean displacement remains constant:

$$\overline{X}(t+dt) = \overline{X}(t) \tag{15}$$

This equation is obtained by averaging over the *ensemble of all new positions*, assuming that the old positions are symmetrically distributed about the center of mass, and similarly assuming that the initial distribution is symmetric. In addition, if the old particle positions are symmetric, the new particle positions are also symmetrically distributed. Since these properties hold for all times (or time steps), we obtain theoretically the following result:

$$\overline{X}(t) = \overline{X}(0), \forall t \rightarrow \overline{X(t) - X(t) = \delta X(t)} = 0 \quad \overline{X(t) - \overline{X}(t)} = \overline{\delta X(t)} = 0, \quad \forall t \tag{16}$$

Namely, if the initial positions $X(0)$ are symmetrically distributed, then they remain so at all times. We can, therefore, rewrite the scheme as follows, in terms of the zero-mean variable $\Delta X(t)$:

$$\begin{aligned} &\delta X(t+dt) = \delta X(t) + \sqrt{2D_0 dt}\ dU(t) \times \\ &\{H(dU(t))(1 - H(\delta X(t))) + (1 - H(dU(t)))H(\delta X(t))\} \end{aligned} \tag{17}$$

3.3.2 Second order moment(s)

The second step is to compute the variance of particle displacements at time $t + dt$, as follows:

$$Var(X(t+dt) - \overline{X}(0)) = \sigma_X^2(t+dt) = \langle (X(t+dt) - \overline{X}(t+dt))^2 \rangle \approx \langle \delta X(t+dt)^2 \rangle \tag{18}$$

Inserting Eq. 17 for $\delta X(t+dt)$, we obtain after some derivations (Appendix A):

$$\sigma_X^2(t+dt) = \sigma_X^2(t) - \frac{4\sqrt{2D_0}}{\pi}\sigma_X(t)\ \sqrt{dt} \ \ + 2D_0\ dt \tag{19}$$

This can also be expressed by:

$$\frac{d(\sigma_X^2)}{dt} = -\sigma_X\sqrt{\frac{32D_0}{\pi^2 dt}} + 2D_0 \Rightarrow \frac{\Delta(\sigma_X^2)}{\Delta t} = -\sigma_X\sqrt{\frac{32D_0}{\pi^2 \Delta t}} + 2D_0 \tag{20}$$

The left-hand side of (20) is not a proper differential equation (it is a 'pseudo-differential' equation). However, it can be re-interpreted as the (improper) limit of the right-hand side finite difference equation as $\Delta t \to 0$. Remarkably, the rate of change of variance is negative (as desired) but blows up as $\Delta t \to 0$. In other words, the scheme is anti-diffusive but diverges in continuous time. However, the scheme seems applicable in discrete time, with *dt* replaced by a finite time step Δt.

Moreover, it appears that adaptive time stepping can be used to control the anti-diffusive rate. For instance, to maintain anti-diffusion at rate (-D), that is $\Delta\sigma^2/\Delta t = 2D_0$, we shall require a *variable* time step, dependent on the current value of variance, as follows:

$$\Delta t = \frac{2\sigma_X^2(t)}{\pi^2 D_0} \tag{21}$$

This yields the desired result, as can be seen by inserting the variable time step in previous relations.

3.4 Can anti-diffusion be achieved with constant time step?

The previous analysis led to the conclusion that an adaptive (decreasing) time stepping scheme is necessary in order to maintain a constant amount of anti-diffusion. Here, we analyse more precisely the interrelations between the following parameters:

(i) time-step evolution versus time, or versus number of steps;
(ii) variance evolution (decrease rate);
(iii) total time to reach a satisfactory anti-diffuse solution; and
(iv) total number of steps required to reach a "satisfactory" anti-diffuse solution

where the term "satisfactory" means that the concentration has focused with a high degree of spatial resolution, and at the correct rate, on the point source that originated the observed, initially diffuse concentration plume.

3.5 Analysis of constant time-stepping scheme

For constant time step, $\Delta t = \Delta t_0$, we obtain asymptotically (for Δt_0 small enough) an equivalent continuous-time differential equation governing $\sigma_X(t)$:

$$\frac{d\sigma_X(t)}{dt} \approx -\sqrt{\frac{8D_0}{\pi^2 \Delta t_0}} + \frac{D_0}{\sigma_X(t)} \tag{22}$$

This differential equation was solved explicitly in terms of $\sigma(t)$, whence the variance $\sigma^2(t)$ for all times. The resulting variance is non-increasing, and its asymptotic value (or infinite time limit) is:

$$\sigma_{MIN}^2 = \left(\frac{\pi}{4}\right)^2 2D_0\Delta t_0 \tag{23}$$

In fact, the explicit solution at all times can be expressed as follows in terms of the minimum attainable variance or the associated standard deviation $\sigma(t)$:

$$\sigma(t) \approx \sigma_{MIN} + (\sigma(0) - \sigma_{MIN}) \exp\left(-\frac{1}{2}\left(\frac{4}{\pi}\right)^2 \frac{t}{\Delta t_0}\right) \tag{24}$$

Indeed we see here that $\sigma(t)$ will be a non-increasing function *if and only if* the time step is less than a critical time step, as follows:

$$\Delta t_0 < \left(\frac{4}{\pi}\right)^2 \frac{\sigma_0^2}{2D_0}. \tag{25}$$

This severe constraint on the time step drastically limits the applicability of the constant time-stepping scheme. In addition, *Fickian* anti-diffusion can never be obtained, regardless of Δt_0.

3.6 Adaptive time stepping

Because of the limitations of the constant time step approach, a new reverse random walk scheme is proposed based on adaptive, variable time-stepping.

The continuous-time equivalent equation governing the variance of the censored random walk scheme, was considered with a time-varying time step $\Delta t(t)$, which leads to the desired result, i.e., anti-diffusion of the plume at a (negative) *Fickian* rate. Alternatively, this can be accomplished in a discrete time framework with time varying Δt_n for numerical implementation, as presented below.

Let us start again with the variable time step version of equation (20) governing the dispersion variance, in either continuous or discrete time:

$$\frac{d(\sigma_X^2)}{dt} \approx -\sigma_X \sqrt{\frac{32D_0}{\pi^2 dt(t)}} + 2D_0 \quad \text{or} \quad \frac{\Delta(\sigma_X^2)}{\Delta t_n} \approx -\sigma_X \sqrt{\frac{32D_0}{\pi^2 \Delta t_n}} + 2D_0 \tag{26}$$

We can now choose the variable time step $dt(t)$ or Δt_n in order to achieve a prescribed, constant anti-diffusive rate (D_1) such that $d\sigma^2/dt = -2D_1$. With this requirement, we obtain:

$$dt(t) = \eta \frac{\sigma_X^2(t)}{2D_0} \text{ or } \Delta t_{n+1} = \eta \frac{\sigma_X^2(t_n)}{2D_0} \quad \text{where } \eta = \left(\frac{4}{\pi}\right)^2 \frac{1}{(1 + D_1/D_0)^2} \tag{27}$$

It should be emphasized that the pseudo-diffusion coefficient D_0 is used to adjust the intensity of the jumps in the censored random walk, while D_1 is the anti-diffusion rate required to achieve source identification (and $D_0 \neq D_1$ in general). Observe also that the initial time step (Δt_1) depends on the initial spread variance of the plume and on the ratio (D_1/D_0) as follows:

$$\Delta t_1 = \left(\frac{4}{\pi}\right)^2 \frac{1}{(1 + D_1/D_0)^2} \quad \frac{\sigma_X^2(0)}{2D_0} \tag{28}$$

In the proposed scheme, either the initial time step (Δt_1) is selected as a free parameter, or else, D_0 is selected and Δt_1 is deduced from D_0.

Using recursive arguments, and using Eqs. 26–27, it can be shown that:

$$\begin{cases} \Delta t_n = \Delta t_1 \beta^{n-1}; \\ \sigma^2(t_n) = \sigma^2(t_{n-1}) - 2D_1\ \Delta t_n = \sigma^2(0) - 2D_1\ t_n = \sigma^2(0) - 2D_1\ \Delta t_1\ \frac{1-\beta^n}{1-\beta} \end{cases} \quad (29)$$

$$\beta = 1 - \eta \frac{D_1}{D_0} = 1 - \left(\frac{4}{\pi}\right)^2 \frac{D_1/D_0}{(1 + D_1/D_0)^2} \quad (30)$$

These analytical results demonstrate that Fickian anti-diffusion can be achieved with variable time steps. Other remarks can be summarized as follows.

i. First, we note that $1/2 < \beta < 1 \quad \forall\ (D_1/D_0)$ (the lower bound may not be tight)
ii. Choosing a large $D_1 >> D_0$ is equivalent to choosing a small time step, and a β close to unity (implying a slow rate of decay of the time step);
iii. Choosing a small $D_1 << D_0$ is equivalent to choosing a rather large time step, and also, a β very close to unity (implying a slow rate of decay of the time step); and
iv. Choosing $D_1 = D_0$ yields the smallest value of β ($\beta_{MIN} \approx 0.5947$), and the fastest rate of decay of the time step (as a power law).
v. The scheme leads to an upper limit time, t_{SUP} (in the direction of past) that cannot be exceeded; this asymptotic backward time t_{SUP} corresponds to the forward time required for the plume to spread to its present observed state.

The limit time, t_{SUP}, is given by:

$$\lim_{n\to\infty} t_n = t_{SUP} = \Delta t_1 \left(\frac{\pi}{4}\right)^2 \left(\frac{(1 + D_1/D_0)^2}{D_1/D_0}\right) = \frac{\sigma_X^2(0)}{2D_1} \quad (31)$$

where Eq. 28 was used for the last equality. It is seen that the maximum anti-diffusion time t_{SUP} depends only on the initial variance of the plume, and on the required antidiffusion rate D_1.

The limit of $\sigma^2(t_n)$ as $n \to \infty$ is zero. This anti-diffusion scheme is therefore asymptotically perfect, i.e., 100% efficient in the limit as $n \to \infty$ and $\Delta t_n \to 0$. However, t_{sup} is never attained numerically, and the final variance is non zero, to a degree that depends on initial time step selection (more on this later).

The practical computational performance of the scheme can be evaluated, in part, by the number of time steps needed to reduce the initial variance by a given factor. Considering here the special choice $D_0 = D_1$, a good performance should be obtained for $n \approx 10$ time steps:

$$n \approx 10 \rightarrow \sigma^2(t_n) \approx \sigma_0^2\ (0.594715)^{10} \rightarrow \sigma^2(t_n)/\sigma_0^2 \approx 0.0055 \approx 1/1000.$$

In other words, the variance reduction after 10 time steps is better than 1/1000 of the original variance, and the variance continues to be reduced indefinitely as *n* increases. Thus, after 25 steps the variance reduction ratio is on the order of 10^{-6}. The corresponding spread of the plume (square root of dispersion variance) has decreased by a factor 1/1000 in just 25 steps.

This concludes the theoretical section on the numerical implementation of anti-diffusion based on a *non local* RAW scheme in 1D space using an *adaptive time-stepping* scheme. In the next section, some tests are presented and compared to the theory using synthetic 1D gaussian plumes as a surrogate for 'observations' of the concentration distribution.

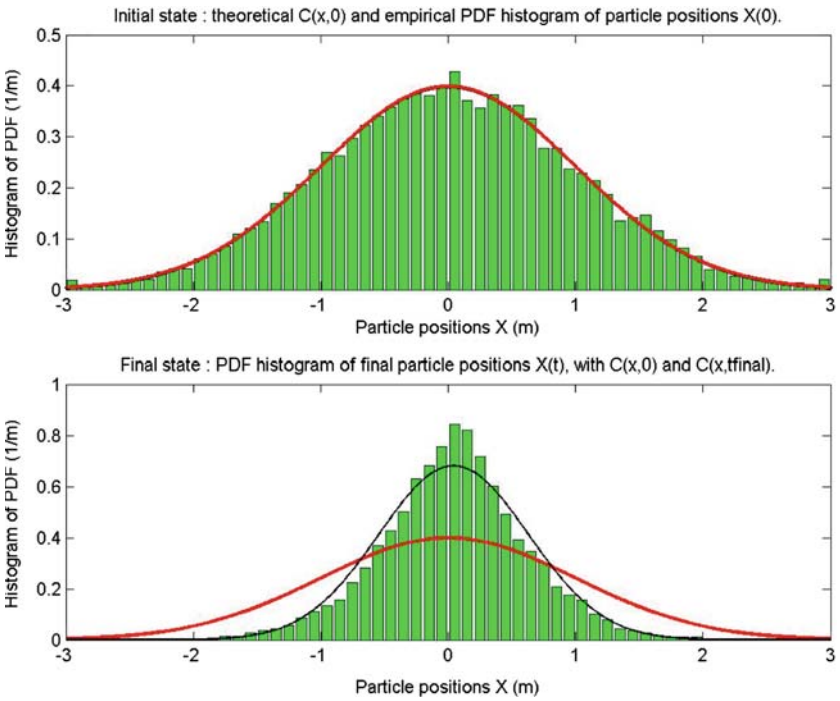


Fig. 1 The *red* curve is the initial Gaussian; the *black* curve is the fitted gaussian at the tenth time step (time t = t(10))

4 Numerical tests in 1D with the non local 'RAW' scheme

The following figures illustrate 1D anti-diffusion results obtained numerically with the 'non local' censored random walk scheme, based on adaptive time stepping, and satisfying all the *theoretical criteria* developed in the previous section. Recall that the purpose of the adaptive time-stepping procedure was to achieve a constant anti-diffusion rate such that $d\sigma^2/dt = -2D1$.

Two slightly different implementations were tested for adaptive time stepping : (i) "analytical", based on assuming a theoretical Gaussian plume; and (ii) "numerical", based on using the actual calculated moments of the plume, including not only the dispersion variance σ^2, but also the center of mass.

Another issue being investigated is the choice of the diffusion ratio D_1/D_0, where D_1 is the "target" diffusion (the backward-diffusion rate to be achieved), while D_0 is a "technical" pseudo-diffusion rate used in the censored random walk algorithm. The choice of D_0 determines the initial value and the rate of decay of the time step (and vice-versa).

Figure 1 shows some results obtained with a synthetic 1D gaussian plume discretized into 10 000 particles, over a span of 50 variable time steps, with a relatively coarse initial $\Delta t(D_1/D_0 = 1/10)$. The anti-diffusion process is depicted via PDF histograms of particle positions plotted at two different times: t=0 (above) and at the tenth time steps ($\mathbf{t}_{10}$). The concentration histograms at each time are compared to gaussian curves fitted by the method of moments. It can be seen that the plume remains approximately Gaussian while it becomes narrower due to anti-diffusion.

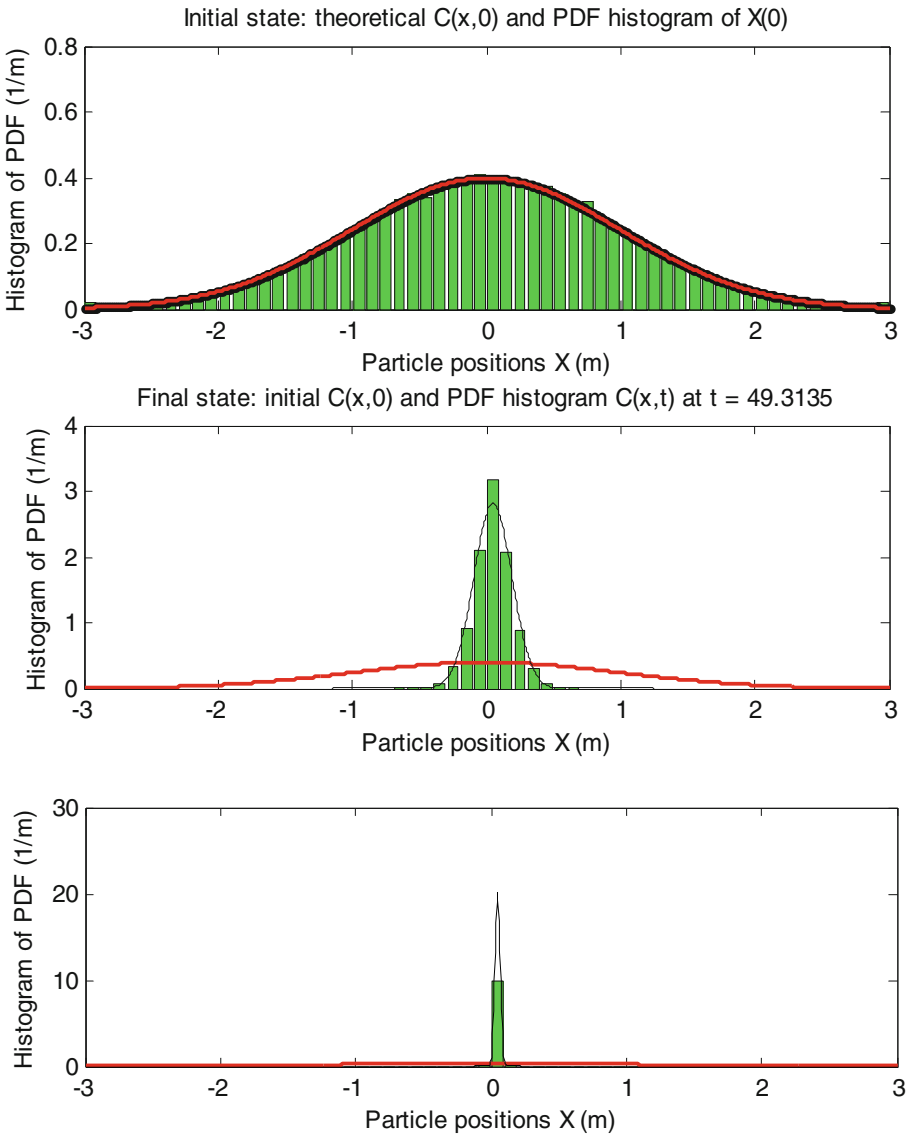


Fig. 2 Evolution of concentration from step 0 (*top*) to step 250 (*middle*) to step 500 (*bottom*); the time span for the last 250 steps is only 0.985 time units, vs. 49.3 time units for the first 250 steps

Figures 2, 3, 4 show the results obtained for a gaussian plume with 25000 particles, 500 variable time steps, and a relatively coarse initial time step corresponding to D_1/D_0=1/100. The time reached after 250 steps is t = 49.3 time units, and after 500 steps, t = 50.285 time units. Figure 2 shows the histograms of concentrations (PDF) at time steps n = 0 (initial), n = 250 and n = 500 steps. The plume narrows down to a Dirac while remaining approximately gaussian. Figure 3 shows a space-time plot of the particle cloud, where the ordinate axis represents time. And finally, Fig. 4 shows *the evolution of dispersion variance versus time*

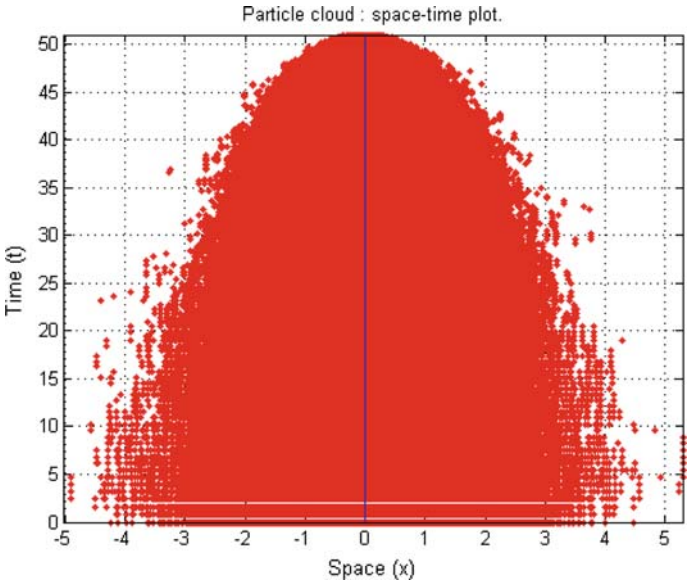


Fig. 3 Space-time plot of the particle cloud during 500 time steps of anti-diffusion with 25000 particles. The ordinate axis represents time; time evolves upwards with anti-diffusion

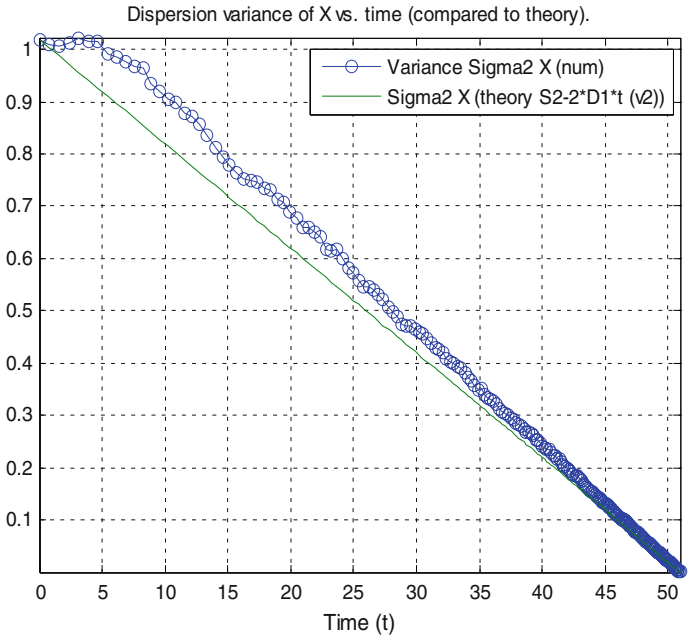


Fig. 4 Time evolution of the *dispersion variance of the plume during anti-diffusion wiuth the RAW scheme* (500 time steps, and slightly over 50 time units)

over the entire time span (500 time steps, slightly over 50 time units); it can be seen here that the simulation conducted with the RAW scheme matches quite closely the required 'reverse fickian' behavior.

In summary, the 1D numerical tests (Figs. 1, 2, 3, 4) indicate that:

i. for small initial time step (case $D_1/D_0 \geq 1$), back-diffusion reduces the dispersion variance but does not attain a 100% reduction of the spread of the plume; the intermediate concentration distributions do not seem to tend to a Gaussian, except perhaps at the end of the back-diffusion process when the plume is reduced to a single spike at the origin;
ii. for large initial time step (case $D_1/D_0 << 1$), on the contrary, the back-diffusion process appears to attain a 100% reduction of the spread of the plume in finite time; moreover, the resulting concentration distributions remain close to Gaussian during the entire process, and tends to a Dirac or single spike in finite time.

Since the best results are obtained with larger initial time steps, it may be concluded that there is a trade off between time resolution and efficiency. The price to pay for finer time resolution is incomplete convergence of the back-diffusing plume towards its original Dirac source. For this reason, the results obtained with larger time steps (say with $D_1/D_0 = 1/10$ or less) are more satisfactory.

5 Numerical results in 2D with the non local RAW scheme

5.1 The RAW algorithm for 2D isotropic anti-diffusion

We adapted the 1D theory to the case of 2D isotropic back-diffusion, starting with initially diffuse isotropic gaussian plumes. We used the same basic idea of a censored random walk in the 2D plane, similar to the 1D case. The plane direction vector of each particle is used to implement the censorship.

Briefly, for each particle, the 2D isotropic censored displacement algorithm in the (x,y) plane is:

i. Define the relative distance of particles to their global center of mass
$\delta R = \sqrt{\delta X^2 + \delta Y^2}$, where $\delta X = X - \overline{X}$ and $\delta Y = Y - \overline{Y}$
ii. Define $\delta R_+ = \sqrt{\delta X_+^2 + \delta Y_+^2}$ where:
   a. $\delta X_+(t) = \delta X(t) + dX_W(t)$ and $\delta Y_+(t) = \delta Y(t) + dY_W(t)$
   b. $dX_W(t) = \sqrt{2D_0 dt}.G_X$ and $dY_W(t) = \sqrt{2D_0 dt}.G_Y$
iii. Define the isotropic censoring operator Γ :
if $\delta R_+(t) > \delta R(t)$ *then* $\Gamma = 0$ *else* $\Gamma = 1$.
iv. Implement the censored walk displacement in terms of relative positions:
$\delta X(t + dt) = \delta X(t) + \Gamma(t) dX_W(t)$ and $\delta Y(t + dt) = \delta Y(t) + \Gamma(t) dY_W(t)$
from which one deduces the new absolute positions:
$X(t + dt) \approx \overline{X}(t) + \delta X(t + dt)$ and $Y(t + dt) \approx \overline{Y}(t) + \delta Y(t + dt)$.

Note that this algorithm is summarized here in the form of an equivalent continuous-time algorithm for convenience, but the discrete-time version can be easily deduced. With discrete time steps, G_X and G_Y are defined as two purely random uncorrelated sequences of normalized N(0,1) gaussian numbers.

In addition, a discrete time solution with adaptive time-stepping scheme is used, based on the same algorithm as in the 1D case with σ_X^2 replaced by σ^2 where $\sigma^2 = (\sigma_X^2 + \sigma_Y^2)/2$.

As can be seen further below, numerical tests demonstrate that both the new 2D censoring and adaptive time-stepping procedures produce the correct results for isotropic plumes, with one proviso, namely that the diffusion ratio $\omega = D_1/D_0$ (which controls time-stepping) must be held close to $\omega = 1/10$.

5.2 Numerical simulation tests with anti-diffusion on 2D isotropic plumes

Starting with a 2D isotropic Gaussian plume C(x,y), we choose to represent the plume at all times as a histogram of particle clouds versus radial distance (r) to its center of mass. Indeed (see any probability textbook) that, for a bivariate Gaussian, the radial function C(r) follows a Rayleigh distribution.

We implemented our 2D isotropic anti-diffusive walk, and we fitted (by the method of moments) a Rayleigh distribution to our simulated concentration profiles C(r): Figs. 5, 6

It turns out that the Rayleigh distribution fits quite well the numerical plumes obtained at all times during anti-diffusion. This indicates that the anti-diffused plumes remain close to bivariate isotropic Gaussian, even as they become narrower due to anti-diffusion (a satisfactory result).

In addition, we compared the analytical (theoretical) and computed (numerical) dispersion variance during the 2D isotropic anti-diffusion process: Fig. 7. The computed dispersion variance was rendered "isotropic", i.e., it was re-defined as:

$$\sigma^2 = (\sigma_X^2 + \sigma_Y^2)/2.$$

It can be clearly seen from the results in Fig. 7 that the 2D variance $\sigma^2(t)$ decreases almost linearly with time, at the rate $-2D_1 t$, that is at an 'anti-fickian' rate. . .as required.

However, two issues are worth noting concerning the RAW scheme for 2D isotropic plumes:

i. Satisfactory results (gaussian distribution and anti-fickian rate) were achieved only for a diffusion ratio ω=D_1/D_0 in the vicinity of 1/10. This ratio controls the initial time step. This suggests that consistant anti-fickian diffusion can only be obtained within a narrow range of initial time steps in 2D (it is not known whether this constraint becomes more relaxed in 3D).
ii. The previous results, analyzed in terms of the Rayleigh distribution, do not suffice to demonstrate that the isotropy of the plume is indeed preserved. However, isotropy can be verified directly by plotting space-time particle clouds or, better, the concentration distributions in the (x,y) plane at various times (Fig. 8). It is seen that the concentration plume remains indeed close to the isotropic bivariate gaussian distribution at all times.

6 Localized 'RAW' schemes : back-diffusion on local neighborhoods

The 'global' anti-diffusion scheme of the previous sections was designed to refocus an existing 'unimodal' plume towards its center of mass. The encouraging results of 1D/2D tests confirm that it is possible to achieve this objective with a particle-based random walk scheme using a simple 'non local' censoring algorithm. However, for more practical applications, it is necessary to implement more flexible schemes adapted to complex situations such as: non point sources; non isotropic macro-dispersion; heterogeneous advection velocity; multiple or multimodal plumes, unknown sources, etc. In this section, a 'localized' version of the censoring algorithm will be tested.

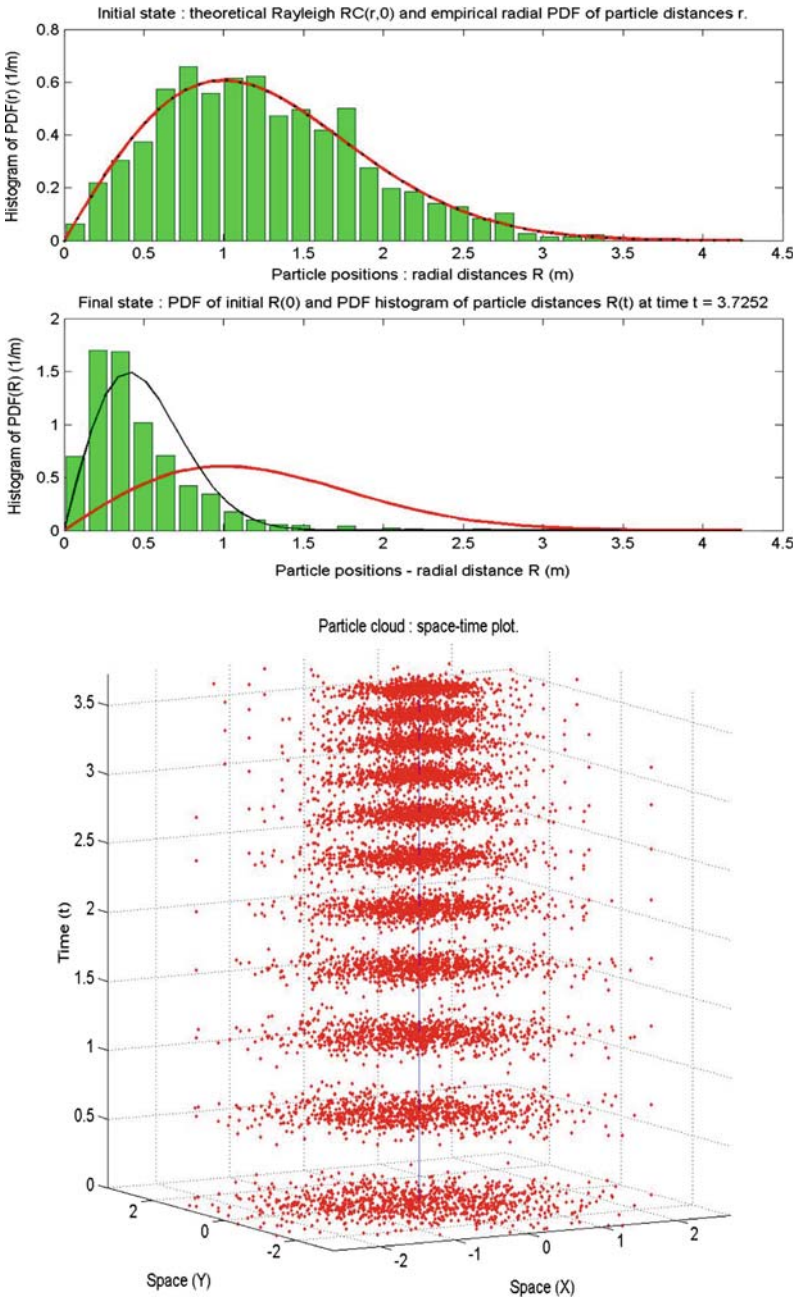


Fig. 5 *above* Initial gaussian isotropic plume with unit variance, shown in terms of its radial distribution C(r) (Rayleigh); and the radial distribution C(r) obtained after 10 steps of anti-diffusion (t = 3.72 time units). There are only 1000 particles in this simulation. *below* Space-time plot of the particle 2D cloud during 10 time steps of isotropic anti-diffusion. The ordinate axis represents time (evolving upwards during anti-diffusion). There are only 1000 particles in this simulation

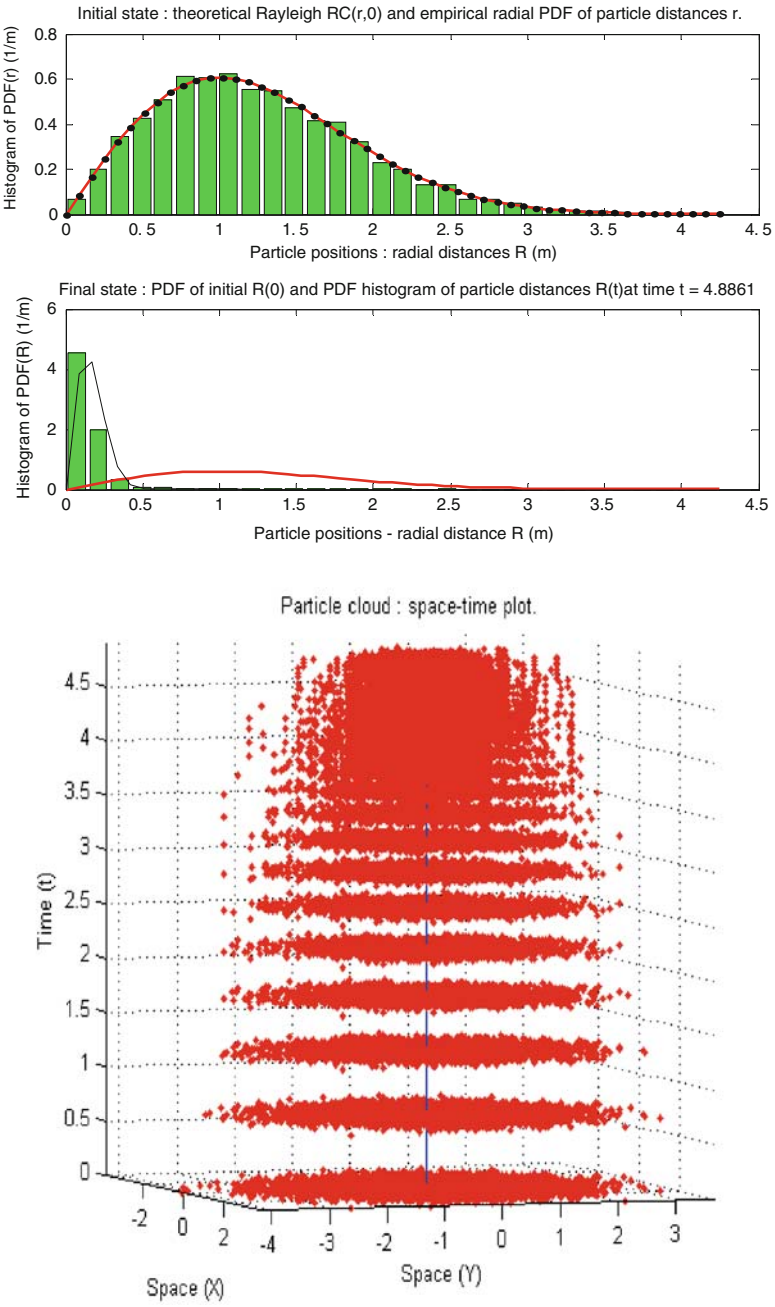


Fig. 6 *above* Initial gaussian isotropic plume with unit variance, shown in terms of its radial distribution C(r) (Rayleigh); and the radial distribution C(r) after 25 steps of anti-diffusion (t = 4.87 time units). There are 10000 particles. The diffusion ratio is: $\omega = D_1/D_0 = 1/10$. *below* Space-time plot of 2D particle cloud during 25 time steps of isotropic anti-diffusion. The ordinate axis represents time, evolving upwards during anti-diffusion. There are 10000 particles. The diffusion ratio is: $\omega = D_1/D_0 = 1/10$

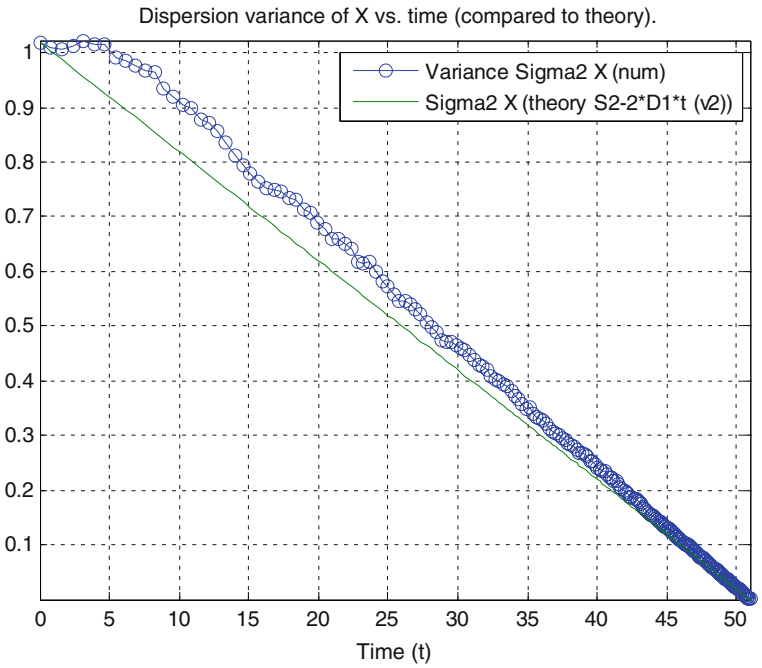


Fig. 7 The circles represents the simulated evolution of the dispersion variance $(\sigma_X^2 + \sigma_Y^2)/2$, to be compared to the theoretically required one (*green line*), over 1000 anti-diffusion steps (10 000 particles; diffusion ratio: $\omega = \mathrm{D}_1/\mathrm{D}_0 = 1/10$)

Fickian diffusion is a *local* phenomenon in the following sense: it can be modeled (macroscopically) as a *local flux* from high concentration zones to low concentration zones ('local' being understood here at the mesoscopic scale of many particles, such that a local *mass flux density* and a local *mass concentration* can be defined: see for instance Fadili et al. [17]).

It is therefore necessary to account for the preferential directions of diffusion, from high to low concentration zones, in order to back-track and back-diffuse the detailed space-time history of a diffusive plume. Clearly, the disadvantage of the previous *non local* scheme is that it does not distinguish between single unimodal plumes and multiple or multimodal plumes. Indeed, with the non local RAW scheme, the plume will always be reduced to a single Dirac located at a single point, regardless of the complexity of the observed concentration pattern.

The new *localized* class of anti-diffusion schemes studied in this section stem from that remark.

The purpose is now to develop reverse particle walks (RAW schemes) that implement back-diffusion locally and directionally (pro-gradient), using information in a local neighborhood of each particle. Again, the main idea is to "censor" the random walk, but this time, based on local information.

We distinguish several methods for defining the relevant type of neighborhood:

- *Cartesian neighborhood.* The neighborhood of a particle is defined as a rectangular box (parallelepiped) centered around the particle. The *local center of mass* is then computed inside this rectangular box. This method is easily implementable, but one drawback is that it defines a non isotropic neighborhood, with preferential directions along the (x,y,z) axes.

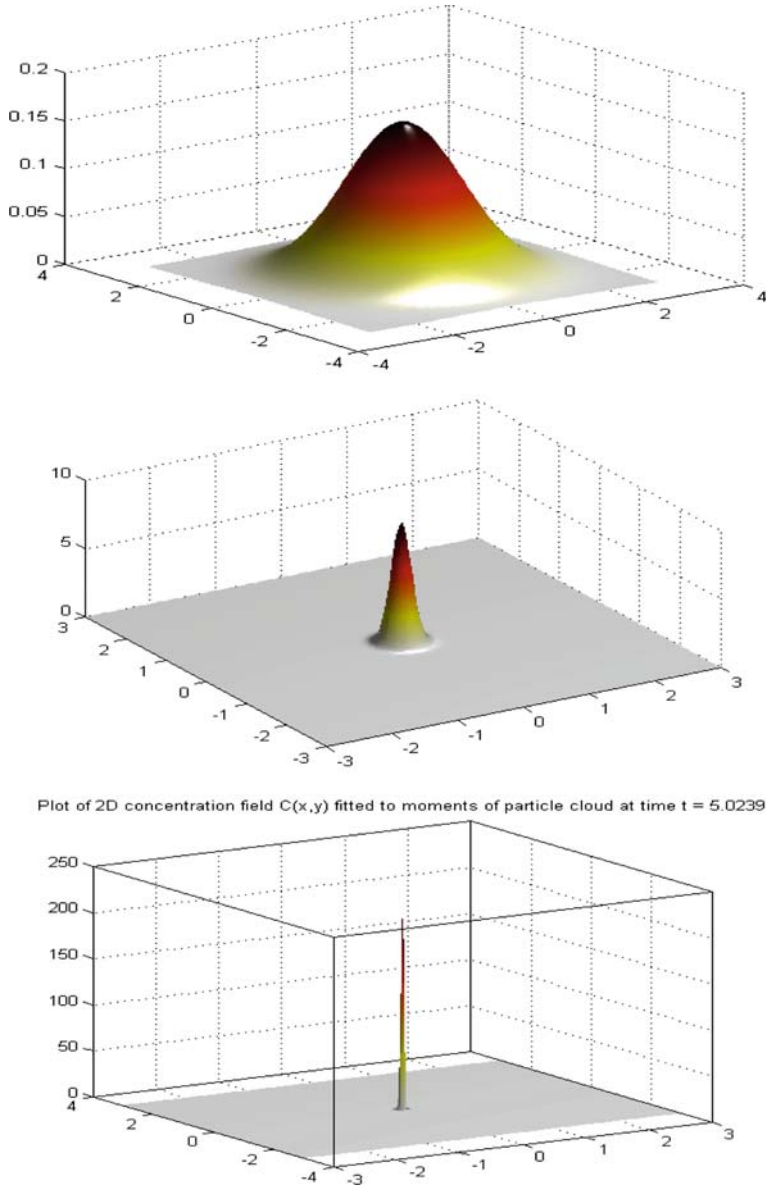


Fig. 8 Simulated 2D concentration distribution in the (x,y) plane at three different times during the anti-diffusive particle walk (10000 particles; diffusion ratio: $\omega = D_1/D_0 = 1/10$): initially "observed" isotropic gaussian plume with unit variance (*above*); anti-diffuse plume after 25 steps (*middle*); anti-diffuse plume after 1000 steps (*below*)

- *Spherical or ellipsoidal neighborhood*. The particle neighborhood is defined as either a sphere or possibly (depending on a priori information) an ellipsoid centered around the particle. The *local center of mass* is then computed inside this spheroidal neighborhood.

Once the neighborhood method has been defined, different schemes may be used to compute the local gradient or the local direction of anti-diffusive flux, for each particle, based on

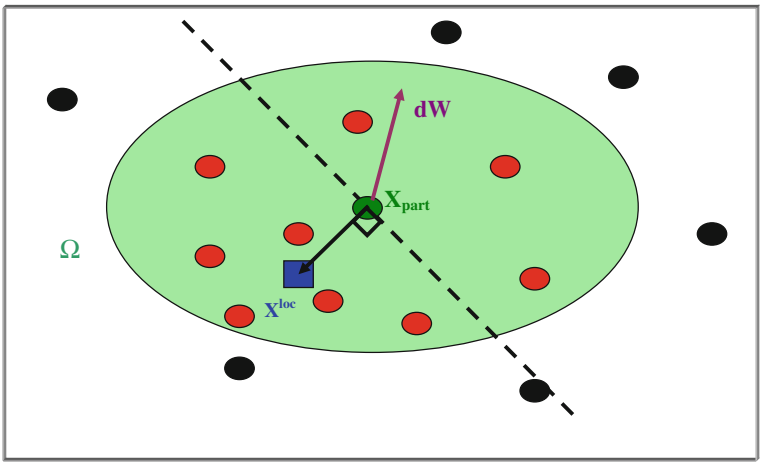


Fig. 9 Schematic illustration of 1 step of the localized anti-diffusive scheme for 1 particle (the *green* particle at the center). The random walk is censored based on the local center of mass (*blue square*) computed in a local neighborhood of the particle (*elliptical shape*). The displacement to be implemented a priori is the Wiener step **dW** (*purple arrow*), but for the configuration shown here, the particle will not be displaced

information collected inside the neighborhood. This is now detailed in the next section—the methods will be tested numerically on some multiple/composite plumes (further below).

6.1 Localized neighborhood scheme based on the local center of mass

In this variant, one computes for each particle to be displaced, the local center of mass in the neighborhood Ω of the particle, as illustrated below in the 2D case (although valid in 3D).

By analogy with the censoring operator Γ defined in previous sections, a condition is tested to decide whether or not to displace particles. For each particle "p", the test is based on the sign of the scalar product between the Wiener displacement vector $\mathbf{dW_p}$ and the separation vector $\boldsymbol{\delta}\mathbf{X_p}$ between the particle (p) and its associated center of mass (based on the local neighborhood Ω_P). That is, we let:

$$\delta\mathbf{X_p} = \mathbf{X}_p - \overline{\mathbf{X}}\,|_{\Omega p}\ ; \quad \overline{\mathbf{X}}\,|_{\Omega p} = \sum_{k\in\Omega} m_k \mathbf{X_k} / \sum_{k\in\Omega} m_k .$$

The sign test is then implemented on the scalar product $d\mathbf{W}_p \bullet \delta\mathbf{X_p}$ as follows (see also Fig. 9):

if $d\mathbf{W}_p \bullet \delta\mathbf{X_p} > 0$: implement the displacement step (censor state: $\Gamma = 1$);
if $d\mathbf{W}_p \bullet \delta\mathbf{X_p} < 0$: do not implement the displacement step (censor state: $\Gamma = 0$).

6.2 Localized neighborhood scheme based on local gradient direction

This variant implements a localized censored walk of particles by forcing them to go "pro-gradient", using a local estimation of the concentration gradient direction based on counting particles along *at least* M directions in M-Dimensional space. We tested in particular a method utilizing only 2 principal directions, in 2D space (see Fig. 10). This method can give a satisfactory estimation of the gradient, provided sufficiently large number of particles in

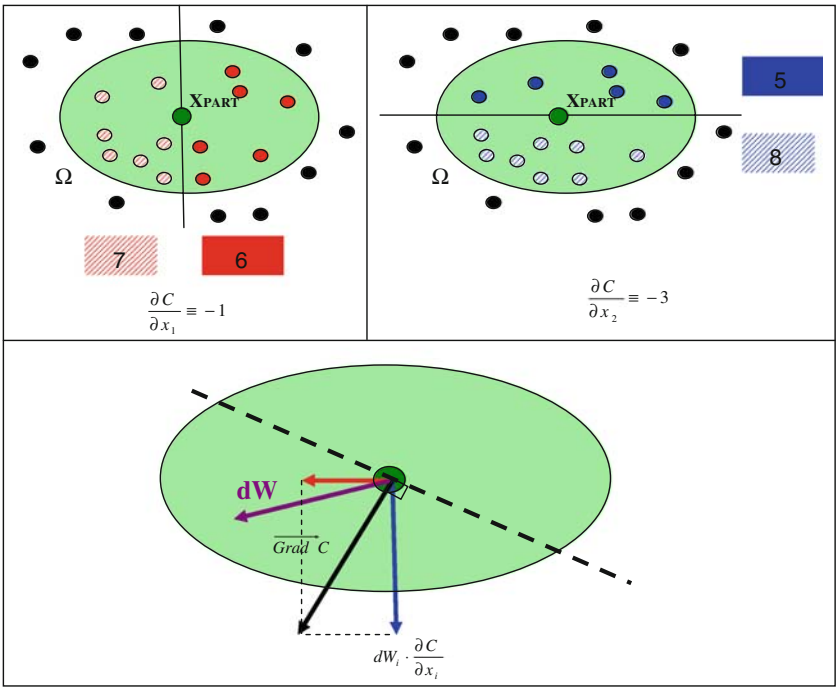


Fig. 10 Schematic representation of the localized scheme based on concentration gradient. Each component of the gradient is computed separately as shown. The components are then assembled into a vector gradient, which is used to split the particle neighborhood in two half-domains

the neighborhood (more precisely, in each subdomain or quadrant). However, is some less favorable cases (not shown here), this method appears to generate spurious anisotropy due to the choice of only 2 directions to evaluate the 2D gradient.

A more satisfactory version consists in using many directions, rather than just two (Fig. 11).

The *multi-directional gradient estimation method* is based on the fact that, for any scalar field C, the gradient vector **grad** C is aligned with the direction that maximizes the directional derivative dC/dn = **n grad**C along the unit vector **n**. Also, it is known that dC/dn = **ngrad**C, this identity being sometimes viewed as defining the directional derivative dC/dn. With these remarks in mind, the proposed method can be formalized as follows:

- We seek to maximise the quantity dC/dn = $\mathbf{n}(\theta)$ **Grad** C in 2D space, where $\mathbf{n}(\theta)$ is a unit vector defined by its polar coordinate angle (θ), such that:

$$\theta = (\mathbf{n}(\theta), \mathbf{x_1}) \in [0, 2\pi], \text{ i.e. } \mathbf{n}(\theta) = [\cos(\theta), \sin(\theta)]^T.$$

- The directional derivative of concentration is then evaluated for different values of θ, within a local neigborhood Ω, simply by partitioning Ω orthogonally to vector $\mathbf{n}(\theta)$ and counting the particles in each of the two half-domains (see Fig. 11). With this, one obtains:

$$\left(\overrightarrow{Grad}(C) \cdot \vec{n}(\theta)\right)_{p_0} \equiv \alpha \equiv \sum_{p \in \Omega_{p_0}} m_p \left\{ H\left(\overrightarrow{X_p X_{p_0}} \cdot \vec{n}(\theta)\right) - H\left(-\overrightarrow{X_p X_{p_0}} \cdot \vec{n}(\theta)\right) \right\},$$

where H is the Heaviside step function defined by H(x) = 1 if x > 0, H(x) = 0 otherwise.

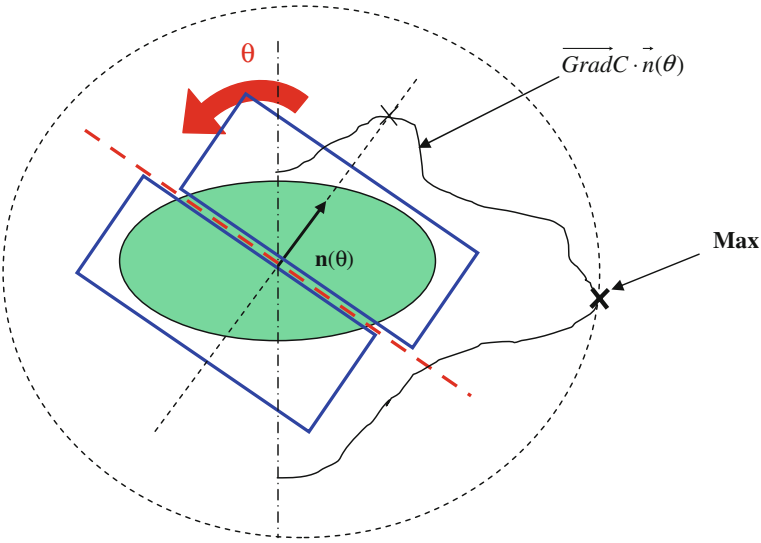


Fig. 11 Schematic representation of *a multidirectional version of* the localized scheme: it is based on the maximal value of the directional derivative. The scheme involves computing the directional derivative "dC/dn" over many directions (θ), and finding the direction that yields the maximal value. This direction of maximal gradient is then chosen to split the neighbourhood into two half-domains

The localized RAW scheme was tested with different 'localization methods', and different sizes of the neighbourhood, for both single and multiple sources, of which only the 'final' concentration distribution was 'observed'.

Overall, the tests indicate that an overly large neighbourhood makes it difficult to discriminate between different sources; indeed, the limit of infinite neighbourhood yields the 'non local' RAW scheme (which can only converge to a single source regardless of the observed concentration distribution). Convertly, if the neighbourhood is chosen too small, it leads to 'noise' and the appearance of many spurious sources as the anti-diffusion process progresses.

We now present these tests in more detail.

6.3 Simulation tests with localized back-diffusion: two point sources in 1D space

The following results were obtained for the case of two point sources in 1D: Figs. 12a, b, 13.

6.4 A simulation test with localized back-diffusion: two point sources in 2D space

The following 2D result was obtained for the case of two intertwined plumes emanating from two (a priori unknown) point sources: see Fig. 14.

6.5 A simulation test with localized back-diffusion: four point sources in 2D space

Finally, we present in more detail the results obtained with the localized back-diffusion 'RAW' scheme for the case of four point sources in 2D space (Fig. 15a, b, c, d).

In this test, the initial (observed) particle plume due to 4 instantaneous point sources is fairly mixed up. This multimodal plume was generated by forward diffusion with 40,000

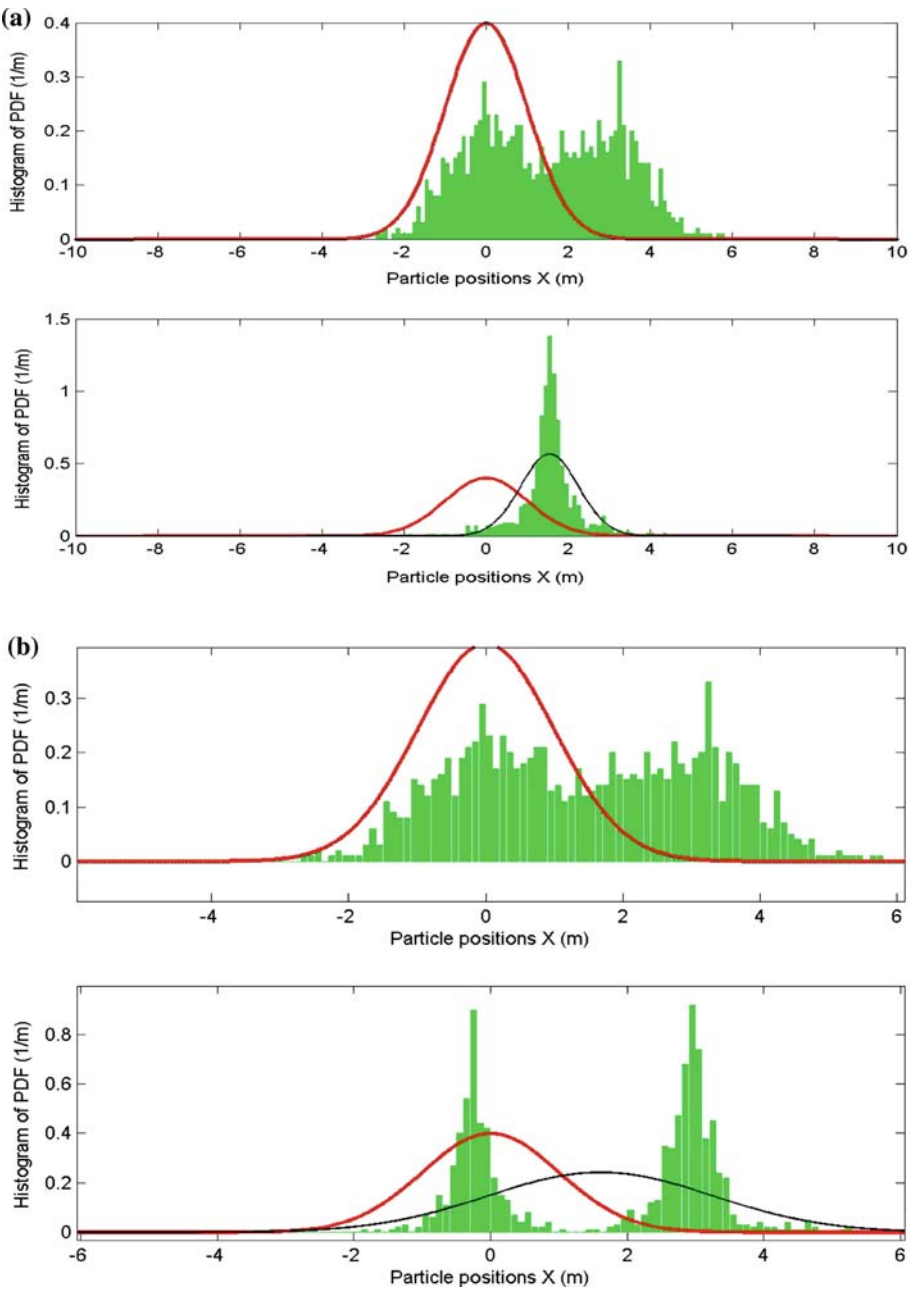


Fig. 12 **a** Unsuccessful attempt to back-diffuse the concentration field emanating from two point sources (localized RAW scheme—method 3): the process leads to a single source instead of two sources, due to an overly large search neighbourhood. **b** A more successful attempt to back-diffuse the concentration field emanating from two point sources (localized RAW scheme—method 3): the process leads to the two original point sources at their approximately correct positions

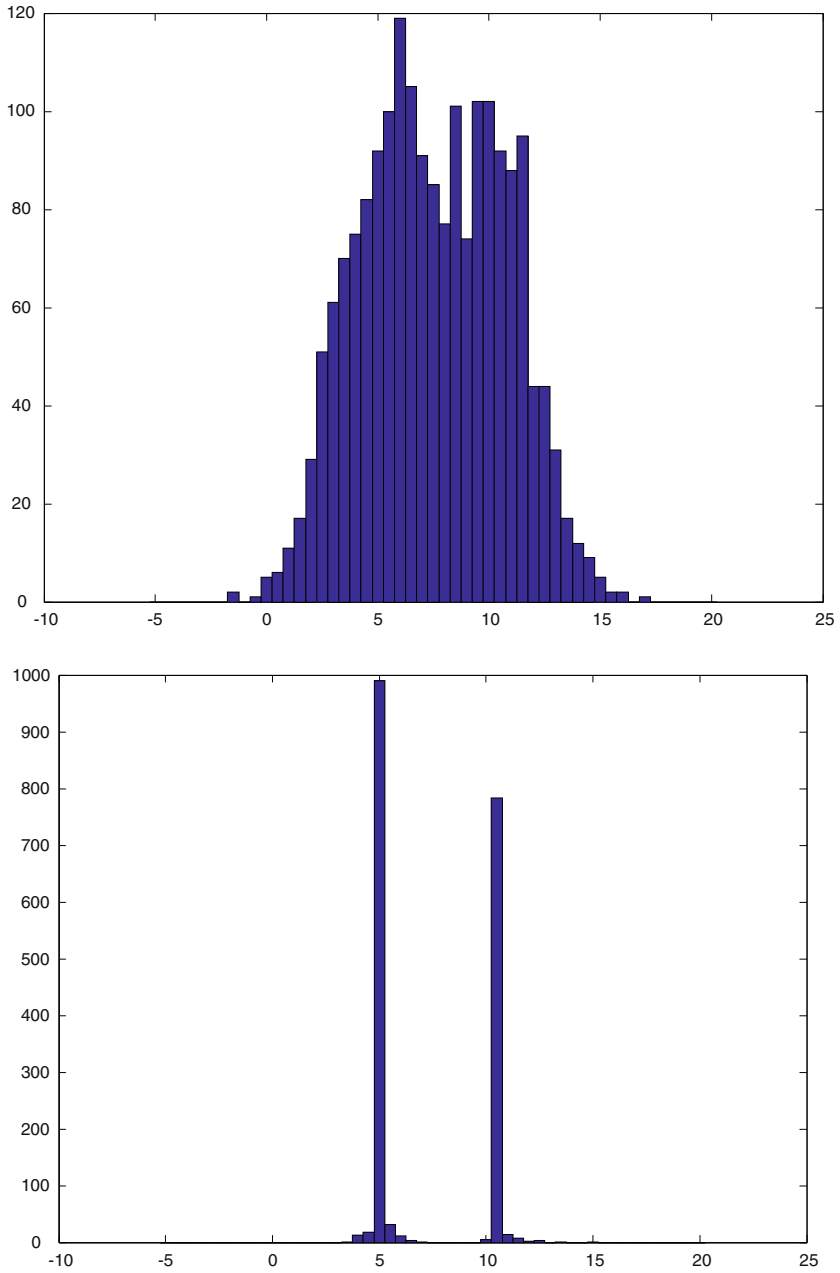


Fig. 13 A fully successful back-diffusion of a concentration field emanating from two point sources, using a non-local version of the 'RAW' scheme: initial plume (*top*) and final plume (*bottom*)

(fourty thousand) particles. The 4 point sources were located at points (x,y) = (5,5), (5,10), (10,5) and (10,10). The forward diffusion rate was $D = 0.1 m^2/s$, and forward time is T = 5 s (that is, the time of observation of the multimodal plume is T = 5 s). Note that the space-time units used here are arbitrary and can be modified. For example, length units may be changed

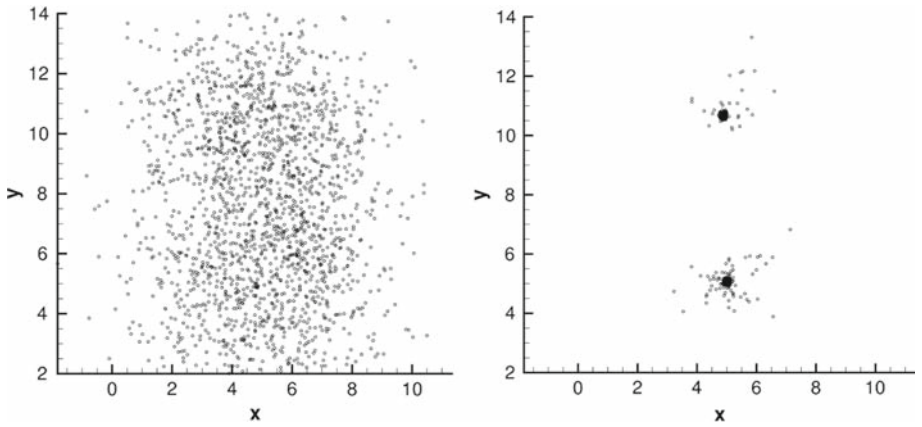


Fig. 14 A successful 2D back-diffusion test on a concentration field emanating from two point sources, using a non-local version of the 'RAW' scheme: initial plume (*left*) and final plume (*right*)

into kilometers and time units into millions of seconds; in that case, the time of observation of the multimodal plume is T = 5 × 1.0E6s ≈ 11.6 days.

Figures 15a, b, c, d display the back-diffused concentration field, respectively, at backward diffusion times $t_{backwards} = 0, 1, 2, 3$ units of time.

At early stages (Fig. 15b) it can be seen that the localized RAW algorithm does a good job at *re-focusing* the particles towards their respective point sources. However, at the same time, the algorithm cannot perfectly 'unmerge' the four different plumes in the interaction regions ("*no man's lands*" or "*passes*"). Mathematically, these regions are located around *saddle points* of the concentration field. Clearly, these are the regions where most of the spurious interferences occur during back-diffusion.

At a later stage (Fig. 15c), the efficiency of the reverse diffusion scheme becomes more obvious. It is now clearer that the initial observed plume corresponds to 4 distinct sources, which can be localized efficiently. Again, some small particle clouds appear in the intermediate merging zones where concentration gradient is small (the so-called *"no man's lands"*) but the amount of particles in these areas is negligible, as will be shown further below.

At terminal stages (Fig. 15d), a very sharp re-focalization of the multimodal plume into its four original point sources has been achieved. Recall that the total diffusion time was T = 5 units. Therefore, the remaining time necessary for fickian back-diffusion into perfectly sharp point sources would be 5 − 3 = 2 time units. However, since focalization is already quite good, the back-diffusion simulation is stopped at this point.

Finally, the previous conclusions are confirmed by the final histograms of particle mass along the X and Y directions, shown in Fig. 16a and b, respectively. The histograms were computed in order to evaluate the centroïd of the plumes, and the amount of particles in the "no man's lands". The histograms are shown here on a logarithmic scale (decimal log of the number of particles). The important fact to be noted is that the total amount of particles is 40 000 (or 10 000 particles for each of the 4 sources), while the spurious central cloud contains only a few hundreds of particles.

7 Summary, conclusions, and outlook

To sum up, a novel scheme (RAW: Reverse Anti-diffusive Walk) was developed in order to back-diffuse concentration plumes and identify pollution sources. The scheme is based on a

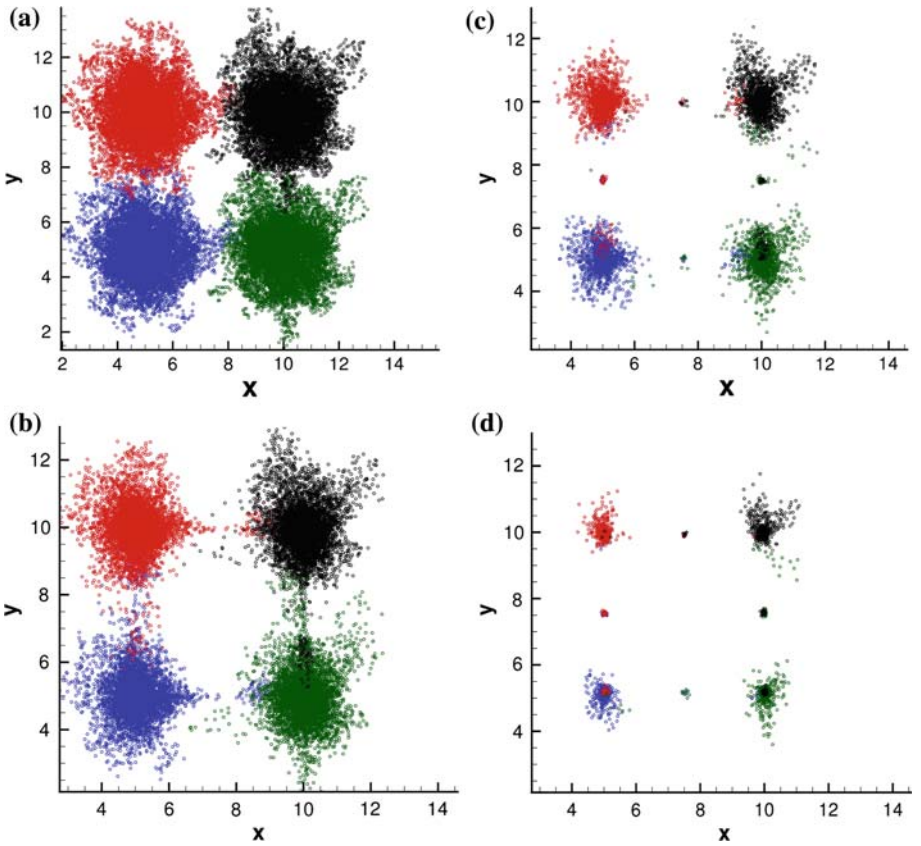


Fig. 15 **a** Time $t_{backwards} = 0$: initial observation of a particle plume due to four instantaneous point sources located at points (x,y) = (5,5), (5,10), (10,5) and (10,10). The sources have diffused for T = 5.5 units of time. The back-diffusion process will now be started (see next figures). **b** Time $t_{backwards} = 1$s: backward diffusion from the initial plume **(a)**. The 4 colours (*red, black, blue, green*) represent the original point sources, but the algorithm did not use this information. **c** Time $t_{backwards} = 2$ s: backward diffusion from the initial plume **(a)**. The 4 colours (*red, black, blue, green*) represent the original point sources, but the algorithm did not use this information. **d** Time $t_{backwards} = 3$s: backward diffusion from the initial plume **(a)**. The 4 colours (*red, black, blue, green*) represent the original point sources, but the algorithm did not use this information

censored random walk of Lagrangian particles, with adaptive time stepping and two versions with respect to the search region (global/local).

- A basic version of the RAW scheme was presented first: the non local (global) antidiffusion scheme. It is based on a "censored" random walk with adaptive time step, whose properties are tailored to produce just the right amount of "Fickian anti-diffusion" in order to force the current (or "final") plume to evolve backwards in time, at the correct rate, towards its "initial" point source.
- This scheme was formulated and then studied analytically in 1D space, first in continuous time, then in discrete time. Variable time stepping is essential to the procedure. The discrete time theory was then confirmed by numerical tests with 1D gaussian plumes. The scheme was also extended to 2D space, and tested for isotropic gaussian plumes. Overall, it was shown that the "RAW" scheme is able to back-diffuse the 1D and 2D plumes correctly to their original point source.

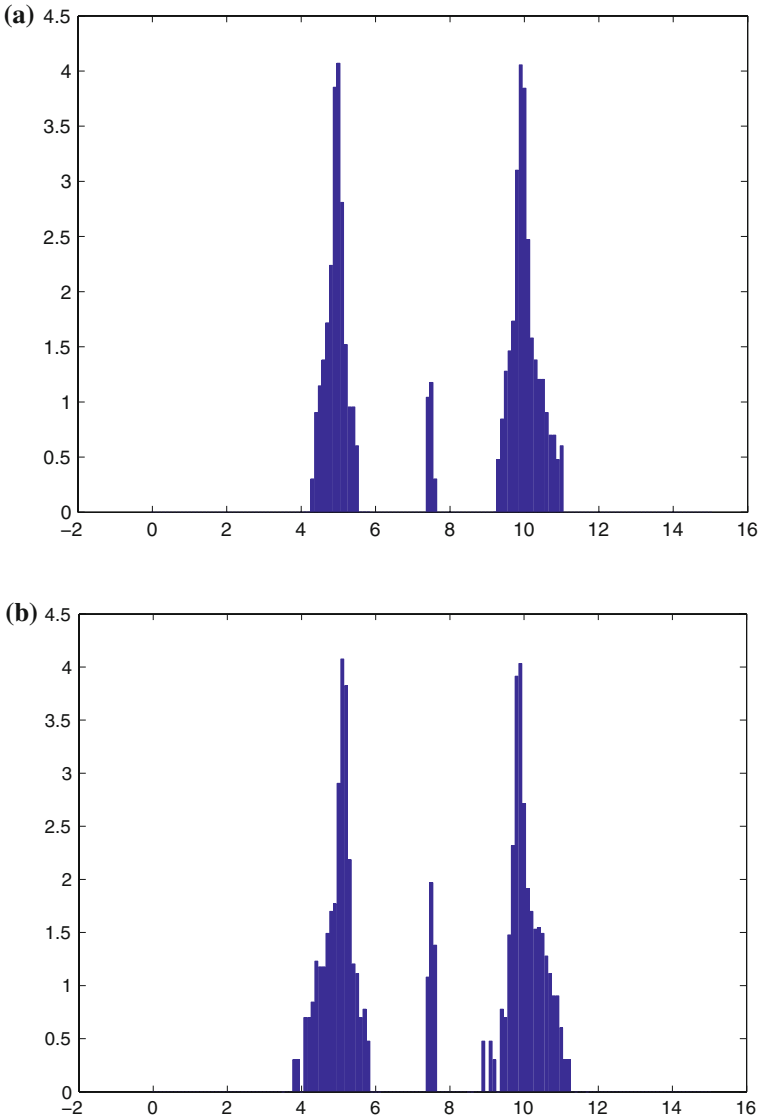


Fig. 16 **a** Final histogram of particle mass along the X direction (4 source problem). **b** Final histogram of particle mass along the Y direction (4 source problem)

- Finally, a localized version of the RAW scheme was developed to deal with complex multimodal plumes, multiple sources, and/or distributed sources. Some tests were presented to illustrate, among other examples, the scheme's ability to discriminate between several point sources, given an observation of a 'mixed up' multimodal plume.

The nonlocal version of the scheme, although somewhat limited in scope, was analyzed theoretically and tested numerically. The scheme appears to be quite efficient for recovering the original concentrated source by anti-diffusing at a nearly constant rate, and it is also asymptotically exact in the case of a Gaussian plume. "*Exact*" means here that the proposed

scheme makes the plume back-diffuse at a constant anti-fickian rate until the plume has almost completely re-focused on its initial Dirac source. In other words, the solution of the numerical scheme *converges* asymptotically to the Dirac source, in terms or particle histograms or concentrations.

Furthermore, the scheme is also asymptotically *stable* as the number of time steps grows to infinity the term *asymptotic* refers to the fact that an *infinite* number of time steps is required to achieve total anti-diffusion (albeit in a finite time span). Note that the time step decreases monotonically (for a Gaussian plume) and the scheme converges in a finite time ($T_{sup} < \infty$). For these reasons, the property of *asympotic time-stability* is defined here as follows : $\Delta t(n)$ varies according to the proposed scheme, and $n \rightarrow \infty$. This is compatible with the usual definition of time-stability for a constant finite time step (note that we are referring to time-stability, not Lax-Richtmeyer stability).

The ability of the RAW scheme to generate an *"exact", stable, convergent* anti-diffusion process with constant rate has important implications. The time taken to reverse back to the original instantaneous point source from an "observed" (Gaussian) plume, coïncides with the time required to obtain the same Gaussian plume by forward diffusion from an initial Dirac source. Now, a question arises:

> Why is it possible to implement a direct anti-diffusion scheme in spite of the ill-posedness and irreversibility of the diffusion process ?

In reply to this question, we first note that, while information on individual particles is necessarily lost in the diffusion process (increasing entropy), collective information embodied in 'moments' or 'PDF's' is a different matter. Secondly, it is important to note that our anti-diffusion scheme controls the first and second order moments of the particle cloud, but not the full PDF of particle positions. This is probably one of the reasons of the reasonable behavior of our anti-diffusion scheme.

Indeed, while it is well known that the diffusion PDE is ill-posed when solved backwards (with final conditions)—our method is driven, not by an anti-diffusive PDE, but by an anti-diffusive stochastic differential system of particles. Furthermore, our scheme is designed to drive backwards only a limited number of moments. It does not attempt to recover the full PDF. In general, an infinite number of moments is required for recovering the full concentration distribution or the PDF of particle positions (for an analyzis of the related 'Moment Inverse Problem', see Ababou et al. [2] or Fadili et al. [17]).

In conclusion, the transient back-diffused concentration field C(x,t) produced by our censored random walk is not necessarily the complete exact solution of the anti-diffusion problem, and the detailed properties of the antidiffusing concentration C(x, t) remain to be analyzed more completely. However, the numerical 1D/2D tests presented here suggest, empirically, that the anti-diffusing concentration C(x, t) remains close to gaussian at all times if the starting plume is itself gaussian.

Finally, let us comment on future applications of the more flexible *localized* RAW scheme. This scheme has a natural potential for adaptivity. It uses an adjustable (finite) neighbourhood search region for the censoring operator. The results obtained so far on various test problems appear convincing. Thus, it was shown in this paper that it is possible to discriminate among intertwined contaminant plumes from several distinct point sources (with successful recovery of the individual sources at their correct locations and at the correct date). In more complex applications, the results will certainly depend on a careful (adaptive) choice of the search region. The objective, in future, is to generalize the localized RAW scheme for identifying sources in more complex cases involving heterogeneous field pollution, with advective transport and velocity-dependent, tensorial hydrodynamic dispersion.

Appendix A

A. Analytical probabilistic displacement & time stepping formulae

(i) Truncated gaussian PDF

The PDF of a positively truncated Gaussian $\mathrm{N}^+(0, 1)$ is: $f^+(x) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) H(x)$
The PDF of a negatively truncated Gaussian $\mathrm{N}^-(0, 1)$ is: $f^-(x) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) (1 - H(x))$

(ii) Mean displacements

In the text, we defined ΔX(t), the particle position(s) relative to the center of mass $\overline{X}(t)$, $\Delta X(t) = X(t) - \overline{X}(t)$. The evolution of the center of mass must be known before examining second order moments. We claim that, under certain conditions, we have simply $\overline{X}(t + dt) = \overline{X}(t)$, where we use (for short), the notation $\overline{X}(t) \equiv E(X(t)) \equiv \langle X(t)\rangle$.

Let us show that the center of mass remains constant. Let us show that this is true for X(t + dt) (or X(t + Δt)), provided that the old distribution X(t) is symmetric about its own center of mass. Indeed, using ensemble averages, we have:

$$\begin{aligned} X(t + dt) &= X(t) + \sqrt{2D}\, dU(t)\, \{H(dU)(1 - H(\Delta X)) + (1 - H(dU))H(\Delta X)))\} \\ \langle X(t + dt)\rangle &= \langle X(t)\rangle + \sqrt{2D}\ \{\langle dU(t)H(dU)(1 - H(\Delta X))\rangle \\ &\quad + \langle dU(t)(1 - H(dU))H(\Delta X)))\rangle\} \end{aligned}$$

But, since U(t) is independent of X(t), we get:

$$\begin{aligned} \langle X(t + dt)\rangle = \langle X(t)\rangle &+ \sqrt{2D} \\ &\times\{\langle dU(t)H(dU)\rangle\langle(1 - H(\Delta X))\rangle + \langle dU(t)(1 - H(dU))\rangle\langle H(\Delta X)\rangle\} \end{aligned}$$

Since we assume that the old distribution X(t) is symmetric about its own center of mass, then ΔX(t) is symmetrically distributed about zero, and as a consequence, it easily shown that:

$$\langle H(\Delta X(t))\rangle = \langle(1 - H(\Delta X(t)))\rangle = 0,$$

whence the announced result:

$$\langle X(t + dt)\rangle = \langle X(t)\rangle, \quad \forall t$$

(iii) Averages of truncated Gaussians (censored random steps)

The probabilistic calculations (variance of censored random walk) involve the mathematical expectation (average) integral $I^+ = \int_0^\infty f^+(x)xdx$, where f^+(x) is a positively truncated $\mathrm{N}^+(0, \sigma^2)$ Gaussian density, with zero mean and variance σ^2. In the text, we need to compute this average in four cases: (i) when $\sigma^2 = \sigma_X^2$(t), (ii) when σ^2 is equal to *dt* (or discrete Δt), and similarly for the negatively truncated Gaussian. For the positively truncated $\mathrm{N}^+(0, \sigma^2)$ PDF, we obtain:

$$I^+ = \int_0^\infty f^+(x)xdx = \frac{2}{\sqrt{2\pi dt}} \int_0^\infty x \exp\left(-\frac{x^2}{2dt}\right) dx$$

$$I^+ = 2\sqrt{\frac{\sigma^2}{2\pi}} \int_0^\infty e^{-v} dv \quad \text{with} \quad v = \frac{x^2}{2\sigma^2}$$

$$I^+ = 2\sqrt{\frac{\sigma^2}{2\pi}} [-0+1] = 2\sqrt{\frac{\sigma^2}{2\pi}}.$$

Whence, in the text, the following results were obtained by replacing σ^2 by σ_X^2 or by dt, and replacing "+" by "−" as the case may be (note that the first equation gives the average of the truncated random increment $dW^+ = dW \times H(dW)$, and similarly for the other equations):

$$\langle dW(t)H(dW(t))\rangle = I^+(dt) = \frac{+2\sqrt{dt}}{\sqrt{2\pi}}$$

$$\langle \Delta X(t)(1 - H(\Delta X(t)))\rangle = I^-(\sigma_X^2) = \frac{-2\sigma_X(t)}{\sqrt{2\pi}}$$

$$\langle dW(t)(1 - H(dW(t)))\rangle = I^-(dt) = \frac{-2\sqrt{dt}}{\sqrt{2\pi}}$$

$$\langle \Delta X(t)H(\Delta X(t))\rangle = I^+(\sigma_X^2) = \frac{+2\sigma_X(t)}{\sqrt{2\pi}}$$

(iv) Second order moments of truncated Gaussians (censored random steps)

In the text we need to evaluate the following second order moments of the positively (+) and negatively (−) truncated unit Wiener process dU(t) :

$$M_2^+ \equiv \langle dU(t)^2 H(dU(t))^2\rangle \quad \text{and/or} \quad M_2^- \equiv \langle dU(t)^2(1 - H(dU(t)))^2\rangle$$

First, it is clear that these two moments are identical by symmetry, since dU is symmetric about its zero mean. So we need only compute M_2^+. Secondly, we know that dU(t) is a zero-mean random Gaussian increment with variance *dt*. The random variable $dU^+(t) = dU \times H(dU(t))$ is the corresponding truncated Gaussian increment. Now, let us replace formally the symbol dU(t) by the symbol "x". Then, "x" is Gaussian N(0,σ^2) with variance $\sigma^2 = dt$. Denoting $f^+(x)$ the corresponding truncated Gaussian PDF, our task boils down to evaluating the second order moment:

$$M_2^+ = \int_0^\infty x^2 f^+(x)\, dx \quad \text{where } f^+(x) = \begin{cases} \frac{2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x}{2\sigma^2}\right) & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$M_2^+ = \frac{2}{\sqrt{2\pi\sigma^2}} \int_0^\infty x^2 \exp(-x^2/2\sigma^2)\, dx$$

Letting $u = x^2/2\sigma^2$, we get $M_2^+ = \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty \sqrt{u} \exp(-u)\, du = \frac{2\sigma^2}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2} = \sigma^2$.

But, since we have here formally $\sigma^2 = dt$, the final result is $M_2^+ = dt$ and similarly $M_2^- = dt$.

Appendix B

B. Analytical Derivation of Wiener Process Statistics in Discrete Time

It was claimed in the text that, in discrete time, a consistent explicit finite difference approximation of the Wiener process is given by :

$$\Delta W(t_n) = \sqrt{2D\Delta t_n}\ \xi_n \quad \text{(cf. Eq. 9 in the text)},$$

where ξn is an uncorrelated random sequence of N(0;1) gaussian variables, such that:

$$\langle \xi_n \xi_m \rangle = \delta_{nm}.$$

We show here that this time-stepping scheme is indeed a consistant approximation of the continuous time Wiener process W(t). That is, the Wn's have the same statistical properties as the theoretical continuous-time Wiener process W(t) taken at times {t = t0, t = t1, t = t2, t = t3, …}.

The continuous time Wiener process

The Wiener process X(t), in continuous time, is solution of a stochastic SDE forced by white noise (and it is also a special case of the Langevin equation):

$$dX(t) = \sqrt{2D}\,\xi(t)dt,\ \text{x} \in \text{IR},\ \text{t} \in \text{IR ot IR+} \tag{A1}$$

where ξ(t) represents a unit white noise, with zero mean and a unit Dirac autocovariance function given by :

$$\text{C}_{\xi\xi}(\tau) = \delta(\tau). \tag{A2}$$

Note that $\delta(\tau)$, the so-called Dirac delta function, is not a proper function but rather a 'distribution'. The physical unit of $\delta(\tau)$ is inverse time, so that D is indeed in m^2/s.

It is then easily shown by ensemble averaging that, for the continuous Wiener process:

$$Var(X(t+\tau) - X(t)) = 2D\tau \quad (\forall\tau) \tag{A3a}$$

$$C_{XX}(t', t'') = 2D\ Min(t', t'') \quad (\forall(t',\ t'')). \tag{A3b}$$

The discrete-time Wiener process

The classical "forward Euler" explicit finite difference scheme $\text{X}^*(\text{tn} + \Delta\text{t}) = \text{X}^*(\text{tn}) + (2\text{D})^{1/2}\ \xi(\text{tn})\Delta\text{t}$ is totally useless and inconsistant because of the white noise ξ(t) in there. Instead, the correct scheme to approximate X(t) is:

$$\hat{X}(t_n + \Delta t) = \hat{X}(t_n) + \sqrt{2D\Delta t}\,Z_n \tag{A4}$$

where Z_n is a purely random (uncorrelated) gaussian sequence:

$$\text{E}(\text{Z}_\text{n}) = 0 \quad \forall\text{n} \quad \text{et E}(\text{Z}_\text{n}.\text{Z}_\text{m}) = \delta_{\text{mn}}(\forall\text{m}, \forall\text{n}) \tag{A5}$$

Z_n is, therefore, the " nth " replicate of a normalized gaussian random variable. The product $\text{Z}_n\ \sqrt{\Delta\text{t}}$ is the correct approximation of the white noise increment with resolution

Δt. In Eq. A4, the "hat" sign serves to distinguish between the continuous process and its discrete-time approximation.

Building on these discrete-time equations, it is relatively easy to show, by ensemble averaging:

$$Var(\hat{X}(t_n + \Delta t) - \hat{X}(t_n)) = 2D\Delta t, \quad \forall t_n. \tag{A6}$$

$$Cov(\hat{X}(t_n), \hat{X}(t_m)) = 2D\ Min(t_n, t_m), \ \forall (t_n, t_m). \tag{A7}$$

As announced earlier, these are exactly the variance/covariance properties of the continuous Wiener process taken at discrete times $\{t = t_0, t = t_1, t = t_2, \ldots\}$.

Acknowledgements The authors gratefully acknowledge several helpful comments and references indicated by Marco Dentz during the revision of this paper towards its final version.

References

1. Ababou R (2008) Quantitative stochastic hydrogeology: the heterogeneous environment. Chapter 8. In: Darnault CJG (ed) overexploitation and contamination of shared groundwater resources. *NATO-ASI: Advanced Studies Institute Series*, Springer Sci. & Business Media BV, pp 119–182
2. Ababou R, Fadili A (1999) Transport in random media : particles, fluxes, concentrations, and the 'Moment Inverse Problem'. Proceedings AUM99, Association Universitaire de Mécanique, 14ème Congrès, Toulouse, Ref. No. 1048, 6 pp, August 1999
3. Ababou R, Bagtzoglou AC, Mallet A (2006) Contaminant source identification with the "RAW" scheme: particle tracking with reverse anti-diffusion walk. Pre-proceedings IAHR-GW2006: International IAHR Symposium on groundWater flow in complex environments, 12–14 June 2006, Toulouse, France. (Extended abstract, 2 pp)—http://www.iahr-gw2006.cict.fr/
4. Atmadja J, Bagtzoglou AC (2001) State of the art report on mathematical methods for groundwater pollution source identification. Environ Forensics 2(3):205–214
5. Atmadja J, Bagtzoglou AC (2001) Pollution source identification in heterogeneous porous media. Water Resour Res 37(8):2113–2125
6. Bagtzoglou AC (2003) On the non-locality of reversed time particle tracking methods. Environ Forensics 4(3):215–225
7. Bagtzoglou AC, Atmadja J (2003) The marching-jury backward beam equation and quasi-reversibility methods for hydrologic inversion: application to contaminant plume spatial distribution recovery. Water Resour Res 39(2):1
8. Bagtzoglou AC, Atmadja J (2005) Mathematical methods for hydrologic inversion: the case of pollution source identification. Chapter in environmental impact assessment of recycled wastes on surface & ground waters. Vol. 3: Engg. Modeling & Sustainability. In: Kassim TA (ed) The handbook of environmental chemistry, Water Pollution Series, Vol 5, Part F, ISBN 3-540-00268-5, Springer-Verlag, Heidelberg-New York, pp 65–96
9. Bagtzoglou AC, Ababou R (2006) Anti-diffusion modeling using a non-local censored random walk scheme. Proceedings CMWR XVI: 16th international conference on computational methods in water resources, Copenhagen, Denmark (19–22 June 2006), 8 pp
10. Bagtzoglou AC, Tompson AFB, Dougherty DE (1992) Projection functions for particle-grid methods. Numer Methods Partial Diff Equ 8:325–340
11. Cornaton F (2004) Deterministic models of groundwater age, life expectancy and transit time distributions in advective-dispersive systems. Ph.D. thesis, Univ. of Neuchatel, Switzerland, Jan. 2004
12. Cornaton F, Perrochet P (2006) Groundwater age, life expectancy and transit time distributions in advective-dispersive systems: 1. Generalized reservoir theory. Adv Water Res 29(9):1267–1291
13. Cornaton F, Perrochet P (2006) Groundwater age, life expectancy and transit time distributions in advective-dispersive systems: 2. Reservoir theory for sub-drainage basins. Adv Water Res 29(9):1292–1305
14. Corrsin (1972) Backward diffusion in turbulent flow. Phys fluid 15(6):986–987
15. Delay F, Ackerer P, Danquigny C (2005) Simulating solute transport in porous or fractured formations using random walk particle tracking: a review. Vadose Zone J 4:360–379
16. Fadili A, Ababou R, Lenormand R (1997) Dispersive particle transport identification of macroscale behavior in heterogeneous stratified groundwater flows. Proc. IAMG'97 (International Association of

Mathematical Geology), V. Pawlowsky-Glahn ed., CIMNE, Barcelona, Spain. Part 2: Subsurface Flow and Transport, pp 825–829
17. Fadili A, Ababou R, Lenormand R (1999) Dispersive particle transport and flux-concentration moments: identification of macroscale behavior in stratified groundwater flows. Math Geol 31(7):793–840
18. Gardiner CW (1985) Handbook of stochastic methods for physics, chemistry and natural sciences. Springer-Verlag, New York
19. Haine TWN, Zhang H, Waugh D, Holzer M (2008) On transit-time distributions in unsteady circulation models. Ocean Model 21:35–45. doi:10.1016/j.ocemod.2007.11.004
20. Kitanidis PK (1994) Particle-tracking equations for solution of the advection-dispersion equation with variable coefficients. Water Resour Res 30(11):3225–3227
21. Labolle EM, Fogg GE, Tompson AFB (1996) Random-Walk simulation of transport in heterogeneous porous media : Local-mass conservation problem and implementation methods. Water Resour Res 32(3):583–593
22. Labolle EM, Quastel J, Fogg GE, Gravner J (2000) Diffusion processes in composite porous media and their numerical integration by random walks : Generalized stochastic differential equations with discontinuities coefficients. Water Resour Res 36(3):651–662
23. Li Qiang, Yoshida Yasua, Nakamori Nobuyuki (1997) A multiscale antidiffusion approach for Gaussian blurred images, 0-8186-8183-7/97, 1997 IEEE, pp 238–241
24. Milnes E, Perrochet P (2007) Simultaneous identification of a single pollution point-source location and contamination time under known flow field conditions. Adv Water Resour 30:2439–2446
25. Morrison RD (2000) Application of forensic techniques for age dating and source identification in environmental litigation. Environ Forensics 1(3):131–153
26. Morrison RD (2000) Critical review of environmental forensics techniques: part I. Environ Forensics 1(4):157–173
27. Morrison RD (2000) Critical review of environmental forensics techniques: part II. Environ Forensics 1(4):173–195
28. Neupauer RM, Wilson JL (1999) Adjoint method for obtaining backward-in-time location and travel time probabilities of a conservative groundwater contaminant. Water Resour Res 35(11):3389–3398
29. Neupauer RM, Wilson JL (2004) Numerical implementation of a backward probabilistic model of ground water contamination. Ground Water 42(2):175–189
30. Neupauer RM, Wilson JL (2005) Backward probability model using multiple observations of contamination to identify groundwater contamination sources at the Massachusetts Military Reservation. Water Resour Res Vol 41, W02015, 2005, 14 pp
31. Oppenheim E (ed) (1977) Stochastic processes in chemical physics: the master equation(s). MIT Press, Cambridge, MA (Collection of papers)
32. Spiller M, Ababou R, Köngeter J (1999) Particle modeling of active transport processes based on stochastic simulation of master equation. Proceedings, 14ème Congrès de l'Association Universitaire de Mécanique, Toulouse, 1999, Ref. No.1047, 6 pp
33. Spiller M, Ababou R, Köngeter J (2000) An alternative approach to simulate transport based on the master Equation, International conference on tracers and modeling in hydrology, Internat Assoc Hydrol Sci: IAHS Publication No.262 (ISSN 0144-7815) Lièges, Belgium, pp 121–126,
34. Spiller M, Ababou R, Becker T, Fadili A, Köngeter J (2002) Mass transport with heterogeneous diffusion: interpolation schemes for random walks. IAMG 2002, 8th annual Conference, Internat Assoc Math. Geol., Berlin, Germany, 15–20 Sept. 2002. Berlin: Selbstverlag der Alfred-Wegner-Stiftung (Terra Nostra: Schriften der Alfred-Wegener-Stiftung; 4,2), Berlin, pp 305–310,
35. Tompson AFB (1998) On the use of particle tracking methods for solute transport in porous media. Comput Methods Water Res 2
36. Uffink GJM (1988) Modelling of solute transport with the Random Walk method. Groundw Flow Qual Model 224:247–265
37. Van Kampen NG (1961) A power series expansion of the master equation. Can J Phys 39
38. Wax N (1954) Noise and stochastic processes, Dover, 1954 (collection of papers, including papers by Chandrasekhar, Einstein, Rice, and others)

Environ Fluid Mech (2010) 10:77–90
DOI 10.1007/s10652-009-9139-2

ORIGINAL ARTICLE

Comparison between backward probability and particle tracking methods for the delineation of well head protection areas

Tiziana Tosco · Rajandrea Sethi

Received: 30 December 2008 / Accepted: 2 June 2009 / Published online: 17 June 2009

Abstract In this work, a deterministic and a probabilistic method for the delineation of well head protection areas are applied and compared. The deterministic method was implemented using the automatic backward particle tracking algorithm (APA, Tosco et al., Water Resour Res, 44(7):W07419, 2008). The backward probability model rests upon the backward adjoint-based model developed by Neupauer and Wilson, and allows the inclusion of dispersion in the definition of capture zones. The two methods are evaluated comparing the "advective front" of the probability protection area and the perimeter given by the particle tracking method. Furthermore, a semi-quantitative study was performed over probability protection areas, in order to evaluate the influence of dispersivity on the extent and growth rate of capture zones identified by fixed probability isolines.

Keywords Capture zones · Backward modelling · Well head protection area (WHPA) · Adjoint equation

1 Introduction

The delineation of Well Head Protection Areas (WHPAs) is usually performed with methods based on the concept of time of travel (TOT): a time is fixed, and the corresponding region of the aquifer is identified, in which the water particles reach the pumping well. For TOT-based WHPAs, both deterministic and probability methods can be used. A deterministic capture area (D-WHPA) is identified by a perimeter, i.e. the boundary line separating the region inside which the water will reach the pumping well within the fixed TOT, and the region in which a higher time of travel would be required to reach the pumping well. On the other hand, the probability methods result in capture areas expressed in terms of probability maps (P-WHPAs).

T. Tosco · R. Sethi (✉)
DITAG – Dipartimento del Territorio, dell'Ambiente e delle Geotecnologie, Politecnico di Torino,
Corso Duca degli Abruzzi 24, 10129, Torino, Italy
e-mail: rajandrea.sethi@polito.it
URL: http://www2.polito.it/ricerca/groundwater/

The deterministic methods were developed first. In the past, analytical solutions, such as the Bear and Jacob equation [1], and analytical-graphic methods like the one proposed by Javandel and Tsang [2], and others, were commonly used because of their simplicity. However, they are effective only under very restrictive assumptions. At present, semi-analytical and numerical methods are generally used for WHPA delineation. They can be easily implemented in case of anisotropy, non-homogeneity, and complex geometry [3]. They are also easily coupled with any flow simulation code. In particular, the backward particle tracking model developed by Pollock [4] is the most common methodology. Being a backward method, the particle tracking time-related capture zones can be computed with only one simulation, and the perimeter of the capture area is then delineated by the final positions reached by the particles after a simulation time equal to the fixed TOT. In this way, the backward particle tracking model defines WHPAs in deterministic terms. An Automatic Protection Area (APA) delineation procedure in 2D geometry was recently developed and presented by the authors of this paper [5], and it is here employed for backward particle tracking. Another automatic delineation procedure was proposed by Shafer [6]. However, the delineation of a closed capture area with the particle tracking methodology is much more complicated for a 3D geometry, and complex algorithms are to be implemented. Although less used and more complex, the probability methods allow us to take into account the uncertainty related to our knowledge of the distribution of aquifer parameters. They also make it possible to include this uncertainty into the WHPA delineation, resulting in capture zones expressed as capture probability maps (or capture zone perimeters associated with their confidence limits).

In this work, the deterministic backward particle tracking model is compared to a backward probability model, based on the adjoint of the transport equation which was derived by Neupauer and Wilson [7]. A similar approach was also used by Cornaton and Perrochet, in order to evaluate the groundwater age and travel time in aquifer systems [8,9]. The backward probability model describes the spreading of a capture probability, generated in the pumping well, which moves backwards along the flow direction, according to advection-dispersion phenomena [10–12]. A value of capture probability, similarly to a concentration distribution, is associated to every point of the domain, and a time-related capture probability map can be obtained with only one backward simulation. However, probability capture zones can also be (and must be, in practical applications) reduced to a closed area, by identifying a "limit" probability and the corresponding isoline.

The backward probability model was first developed by Neupauer and Wilson [7] for the identification of the most likely position of a contamination source. Suggestions on how to adapt the backward probability equation to WHPAs are also given by Neupauer and Wilson [10,13]. Frind et al. [11] showed how to manage the modified equation for WHPA delineation using existing transport codes, and in particular for 3D complex geometries. The complete equation characterizing the backward probability transport is here directly derived for the application to capture zones, and subsequently briefly discussed. Furthermore, no quantitative comparison between the backward particle tracking method and the backward probability model has been reported in the literature. For this reason, in the second part of this work, the analysis focuses on the differences, in shape and extent, between the deterministic capture zone and the area included by the 0.5 capture probability isoline, that can be taken as a first approximation of the advective front of the probability plume. A few case studies are presented in order to understand when the two methodologies can be considered equivalent, and when they provide different results. The importance of spatial discretization and the influence of boundary conditions are underlined and a qualitative analysis of the influence of the hydrodynamic dispersion and of the limit probability is presented.

2 Probability model: the backward probability method

Neupauer and Wilson [7] developed the backward probability model working from the advection-dispersion equation (ADE). In particular, if the purpose is the delineation of WHPAs, the appropriate equation is the ADE for a conservative solute. The ADE describing the transport in a porous saturated medium of a contaminant neither degraded nor delayed by sorption in a generic domain Ω can be written in the divergence form as:

$$-\vartheta\frac{\partial C}{\partial t}+\frac{\partial}{\partial x_i}\left(\vartheta D_{ij}\frac{\partial C}{\partial x_j}\right)-\frac{\partial(\vartheta v_i C)}{\partial x_i}+q_I C_I-q_O C=0 \tag{1}$$

$$C(x_i,0)=C_0(x_i)\quad\text{on }\Omega \tag{1a}$$

$$C(x_i,t)=C_1(x_i,t)\quad\text{on }\Gamma_1\subset\partial\Omega \tag{1b}$$

$$\left[D_{ij}\frac{\partial C}{\partial x_j}\right]\cdot\vec{n}_i=\overline{q_{n,2}}(x_i,t)\quad\text{on }\Gamma_2\subset\partial\Omega \tag{1c}$$

$$\left[v_i C-D_{ij}\frac{\partial C}{\partial x_j}\right]\cdot\vec{n}_i=\overline{q_{n,3}}(x_i,t)\quad\text{on }\Gamma_3\subset\partial\Omega \tag{1d}$$

where $i, j = x, y, z$, D_{ij} is the i,j-th element of the dispersion tensor, ϑ is the porosity, $q_I C_I$ and $q_O C$ represent, respectively, the sources and sinks of contamination, v_i is the effective flow velocity along the x_i direction, C_0 is the initial concentration distribution, Γ_1, Γ_2 and Γ_3 are the portions of the domain boundary $\partial\Omega$ at which the fist, second and third type boundary conditions are respectively applied, C_1 is the given concentration at the first type boundaries Γ_1, n_i is the normal unit vector in the i-th direction, and $\overline{q_{n,2}}$ and $\overline{q_{n,3}}$ are the mass flow rates per unit volume at the second and third type boundaries Γ_2 and Γ_3, respectively.

Equation 1 is a forward model: the simulation is carried out forwards in space and time to define the contaminant position at a fixed time. The equation can also be easily re-written in terms of forward capture probability: for this purpose the concentration C is substituted by a probability p_t, describing the probability that a solute particle, released at a known contamination source at $t = 0$, would reach a point X of the model domain after a simulation time t^* [13]. The corresponding cumulative distribution function, P_t, can be used in the layout of capture zones (the particle reaches X in a time $t \le t^*$). The probabilities p_t or P_t are easily calculated using Eq. 1 if the contamination source (the starting point X) is known, but this is not the case of capture zones: it would be necessary to run the transport model for each cell of the domain, and then combining the results. For reasons similar to the particle tracking model, a backward model would be more appropriate. Starting from the ADE and using a variational analysis, Neupauer and Wilson [7,13] derived the adjoint of the ADE, which is structured as a backward model, i.e. it describes the advective-dispersive transport of a certain variable in the opposite direction of the flow field and backwards in time, in a saturated porous medium. When applied to capture zone delineation, the variable that moves and spreads in the aquifer is the probability P_t: the pumping well is considered a continuous probability source which generates a probability plume (as in forward models, where a source of contamination generates a contaminant plume) [14,15]. The advective-dispersive backward transport of the probability P_t is modelled by

$$-\vartheta\frac{\partial P_t}{\partial\tau}+\frac{\partial}{\partial x_i}\left(\vartheta D_{ij}\frac{\partial P_t}{\partial x_j}\right)+\frac{\partial}{\partial x_i}(\vartheta v_i P_t)-q_I P_t=-Q(x_{i,w})S_w \tag{2}$$

$$P_t(x_i, 0) = 0 \quad \text{on } \Omega \tag{2a}$$

$$P_t(x_i, \tau) = 0 \quad \text{on } \Gamma_1 \subset \partial\Omega \tag{2b}$$

$$\left[v_i P_t + D_{ij} \frac{\partial P_t}{\partial x_j} \right] \cdot \vec{n}_i = 0 \quad \text{on } \Gamma_2 \subset \partial\Omega \tag{2c}$$

$$\left[D_{ij} \frac{\partial P_t}{\partial x_j} \right] \cdot \vec{n}_i = 0 \quad \text{on } \Gamma_3 \subset \partial\Omega \tag{2d}$$

where the flow velocity v of the forward model of Eq. 1 is substituted by $-v$ [11], $\tau = t_{fin} - t$ is the backward time, which runs in the opposite direction of the forward time t (i.e. from the present to the past), and $Q(x_{i,w})$ is the pumping rate of the well w. The term on the right side of the equation is the so called load term, which describes the source of probability (for P-WHPAs, the pumping well). In particular, for capture zone delineation, the term is equal to [10]:

$$S_w = \delta(x_1 - x_{1,w})\delta(x_2 - x_{2,w}) \quad \text{in a 2D geometry} \tag{3a}$$

$$S_w = \delta(x_1 - x_{1,w})\delta(x_2 - x_{2,w}) \frac{B_{x_3}(x_{3w,b}, x_{3w,t})}{x_{3w,t} - x_{3w,b}} \quad \text{in a 3D geometry} \tag{3b}$$

where the $\delta(\cdot)$ functions are Dirac delta functions applied in space (in the x_1 and x_2 directions, respectively). The B_{x3} is a boxcar function, which distributes the capture probability along the x_3 dimension of the pumping well (between the top, $w_{3,t}$, and the bottom, $x_{3,b}$). The load term S_w is related to the probability source (the pumping well) and thus it is not equal to zero only at the pumping well. The term locates a cumulated distribution function of 1 at the pumping well during the entire simulation. Formally, the right side of the equation can be transferred into a first type boundary condition in the backward model. As for the initial and boundary conditions (2a)–(2d), they are slightly different from those of the forward transport model, and are applied, respectively, at the same boundaries Γ_1, Γ_2 and Γ_3 where the conditions (1a)–(1d) are applied. The second type boundaries of the forward model become third type boundaries in the backward model, and vice versa (see Eqs. 1, 2).

Provided that the chosen parameters are correct and representative, the application of the backward probability model is quite simple. If the flow and transport problems are solved using a code that couples the two equations, a modification of the source code for the transport equation is required. Otherwise, if the two equations are solved separately (eg. with MODFLOW [16] and MT3DMS [17]) the implementation is simplified: it is necessary to reverse the flow field and to modify some boundary conditions of the transport model, which can be directly used to solve the backward transport problem. The following procedure was conducted for this study:

- solution of the forward flow problem using MODFLOW 2000;
- inversion of the flow field using a Matlab script, implemented for this purpose;
- solution of the transport problem with the new boundary conditions and the reversed flow field, using MT3DMS;
- interpretation and post-processing of the results in terms of probabilities.

The appropriate boundary conditions for the implementation of the backward transport problem and for the discretization of the load term are chosen according to Neupauer and Wilson [10]. In particular, the load term is approximated with a first type boundary condition at the well cell. In a 2D geometry, a constant probability of 1 is applied at the well cell, while in a 3D geometry the boxcar function of Eq. 3b is approximated distributing the unit probability along the layers containing the well, making sure that its integral along the vertical direction is equal to 1.

3 Deterministic model: backward particle tracking

In general, when the particle tracking method [4] is applied, a circle of equally spaced particles is located around the pumping well and traced backwards. The capture zone encirclement at the desired travel time is then performed manually, connecting the end points reached by the particles at the fixed time. Consequently, equally spaced particles can be inefficient in case of strong heterogeneities in the aquifer parameters distribution, thus requiring a very high number of particles. In addition, if a limited number of traces is employed, the perimeter can be inaccurately defined near the stagnation points. The APA algorithm presented in [5] overcomes these problems. This algorithm involves a hybrid forward-backward two-step procedure, based on a numerical calculation of the stagnation point of each well, and further refinements. Forward and backward particle tracking algorithms are coupled, and a post-processing algorithm interpolates the pathlines and delineates closed capture zone perimeters for fixed TOTs in a completely automatic way. The algorithm is structured in three steps:

- APA-I: the position of the stagnation point of each well is identified, and the initial position and radius of two circles of particles is defined;
- APA-II: two particle tracking simulations are performed. For each well, a circle of equally spaced particles around the pumping well is traced backwards, while a circle of particles around the stagnation point is traced forwards. The resulting pathlines are processed and new starting positions for backward (not equally spaced) particles around the pumping well are defined;
- APA-III: the results of a second (fully) backward simulation are elaborated by a post-processing interpolation algorithm.

4 Comparison between D-WHPAs and P-WHPAs

As previously explained, for practical applications of a probability method in the delineation of WHPAs, a limit probability (or more than one) has to be adopted to identify the capture zone as a closed area: probability capture zone perimeters also have to be defined identifying a "limit" probability and the corresponding isoline. The extent of a D-WHPA depends on the value of the aquifer parameters and on the travel time, while the extent of a P-WHPA also strongly depends on the limit probability, which results from the desired level of protection. It is therefore to be determined by the decision maker, rather than by the modeler.

If compared to the particle tracking approach, the P-WHPAs are characterized by a dependence on hydrodynamic dispersion and on probability levels. It follows that the higher the dispersivity value, the higher will be the uncertainty in the aquifer characterization, and the larger will be the capture area, for fixed travel times and capture probabilities. Based on these considerations, the influence of the mentioned parameters (travel time, dispersivity and capture probability) on the extent of the capture area is here investigated.

For a given dispersivity value α, the capture area A identified by the isoline corresponding to a probability level P for a travel time TOT, can be written as the sum of two contributions, an "advective" and a "dispersive" one:

$$A(P, TOT; \alpha) = A_{Adv}(TOT) + A_{Disp}(P, TOT; \alpha) \tag{4}$$

The advective contribution is always positive, it is equal to the area of the D-WHPA and it depends on the travel time. For a homogeneous, isotropic aquifer with a constant thickness and a fully-penetrating well, it depends on the pumping rate Q, the aquifer thickness b, the effective porosity n_e and the time t according to the following expression:

$$A_{Adv}(t) = \frac{Q}{bn_e}t \tag{5}$$

where $Qt = Abn_e$ is the volume of water extracted during a time t.

The dispersive contribution, on the contrary, depends on all three mentioned parameters, and can be either positive or negative.

The effects of the variation of capture probability and dispersivity on the extent of P-WHPAs can be easily investigated if the "growing rate" of the capture area is considered, rather than the area itself:

$$\frac{dA}{dt} = \frac{dA_{Adv}}{dt} + \frac{dA_{Disp}}{dt} \tag{6}$$

The advective contribution can be directly obtained by differentiating Eq. 5. The dispersive contribution can be similarly modelled, leading to the following expression:

$$\frac{dA}{dt} = \frac{dA_{Adv}}{dt} + \frac{dA_{Disp}}{dt} = \frac{Q}{bn_e}\left(1 + C_{Disp}\right) \tag{7}$$

and therefore:

$$\frac{\frac{dA_{Disp}}{dt}}{\frac{dA_{Adv}}{dt}} = C_{Disp}(TOT, P; \alpha) \tag{8}$$

where C_{Disp} represents the growth rate of the dispersive contribution normalized to the advective one, and depends on the geometry of the problem, the parameter values (in particular, the dispersivity), the capture probability and the travel time. Also in this case, the dispersive contribution is negative for capture probabilities $P > 0.5$, as will be shown in the applications below. A high absolute value of C_{Disp} then means that the dispersion phenomenon is acting on the shape and extent of the capture area, and thus its effects could be taken into account when defining WHPAs. In the following paragraphs, D-WHPAs and P-WHPAs calculated for three case studies are presented and compared, using the relation of Eq. 7.

5 Applications

The particle tracking and the backward probability models were applied to the three case studies presented in [5], i.e. three confined 2D aquifers. The dispersive contribution to the extent of the capture areas was determined using Eq. 7. The three test cases are:

- *case 1*: one pumping well in a confined, homogeneous, isotropic aquifer, with the main flow direction along the y axis, hydraulic conductivity $K = 7.5 \times 10^{-5}$ m/s.
- *case 2*: four pumping wells in a confined, homogeneous, isotropic aquifer with the main flow direction along the y axis, hydraulic conductivity $K = 7.5 \times 10^{-5}$ m/s;
- *case 3*: one pumping well in a confined, heterogeneous, isotropic aquifer, with the main flow direction oriented along the diagonal of the domain, hydraulic conductivities $K_1 = 2.0 \times 10^{-4}$ m/s and $K_2 = 5.0 \times 10^{-5}$ m/s. The conductivities are constant within square sub-domains of 1,200 m × 1,200 m (Fig. 3a).

In all cases, the aquifer was represented by one layer, 10 m thick, and divided into square 25 m × 25 m cells. The model domain used in [5], designed as a 3,600 m × 3,600 m square domain (origin of the axes in the lower left corner), is here extended to evaluate the behaviour of the probability capture areas over long travel times (up to over 20 years). The domain for cases 1 and 2 was then extended 700 m east and west, and 6,400 m north, reaching the dimensions of

5,000 m along the x axis and 10,000 m along the y axis. For case 3, the domain was extended 3,600 m west and 3,600 m north. In case 1 and 2, the main flow direction is in the negative y direction, from north to south. The applied flow boundary conditions are two Dirichlet conditions, at the north and south boundaries, resulting in a regional gradient equal to 1.67×10^{-2}. The left and the right boundaries are no flow boundaries. In case 3, four Dirichlet conditions are applied, and a linearly-changing-in-space and constant-in-time head is applied along them, resulting in a mean flow direction along the diagonal of the cells. In case 1 and 3, the pumping well is located at $x = 1{,}812.5$ m, $y = 1{,}437.5$ m, with a pumping rate of 2.0×10^{-2} m^3/s. In case 2, the four pumping wells extract 2.5×10^{-2} m^3/s altogether ($Q = 5.0 \times 10^{-3}$ m^3/s for pumping wells W_1, W_2, W_3; $Q = 1.0 \times 10^{-2}$m^3/s for W_4). The pumping well W_1 is located at $x = 1{,}812.5$ m, $y = 1{,}437.5$ m; W_2 at $x = 1{,}412.5$ m, $y = 1{,}437.5$ m; W_3 in $x = 2{,}212.5$m, $y = 1{,}437.5$ m; W_4 at $x = 1{,}812.5$ m, $y = 1{,}037.5$ m Values of longitudinal dispersivity ranging from 5 to 125 m were considered. The transverse dispersivity was set equal to 1/10 of the longitudinal dispersivity.

The steady state flow field was solved with MODFLOW 2000 using the WHS solver [16], the backward particle tracking using the APA algorithm on MODPATH [4] simulations, the backward probability model with a Matlab script for the inversion of the flow field and MT3DMS [17] for the transport problem, using the third-order Total Variation Diminishing (TVD) method. A fictitious conservative contaminant, which represents the backward probability, was employed. The load term of Eq. 3 was discretized using a first type boundary condition of constant capture probability: a capture probability of 1 was used at the pumping well during the transport simulation. A second type boundary condition of zero probability gradient was applied at all the external boundaries of the model domain (the condition is automatically obtained by the cell-centred finite difference models, as explained above).

The capture areas for several travel times were determined using both the particle tracking and the backward probability method. The isoline for $P = 0.5$ in the capture probability plume has been compared with the deterministic perimeter given by the particle tracking simulation.

5.1 Comparison between D-WHPAs and P-WHPAs

The results for the three test cases are reported in Figs. 1, 2, 3 and 4. Initially, a longitudinal dispersivity of 50 m was considered. The differences in shape and extent between the D-WHPA and the P-WHPA for a capture probability of 0.5 were analyzed. First, for each case study, a comparison of the shape of the areas obtained with the three methods is reported (Figs. 1, 2, 3). Secondly, a more quantitative comparison is presented: a normalized difference ΔA between the extent of the deterministic area A_{Adv} (see Eq. 5) and the surface area included in the 0.5 capture probability isoline $A_{P=0.5}$ was considered (Fig. 4):

$$\Delta A = \frac{A_{Adv} - A_{P=0.5}}{A_{Adv}} \tag{9}$$

In case 1 (one pumping well, homogeneous aquifer) there is no significant difference in shape between the D-WHPA and the area included by the 0.5 probability isoline of the P-WHPA (see Fig. 1). The differences in extent are always <10% (Fig. 4a). They are more evident at low travel times, when the influence of the transport boundary condition at the pumping well and the effects of the spatial discretization (irregular shape of the curves) play an important role. Furthermore, they slightly increase at very long travel times, when the capture probability plume approaches the boundaries of the model domain. In this case, the second type homogeneous boundary conditions, used as an approximation for boundary conditions (2c),

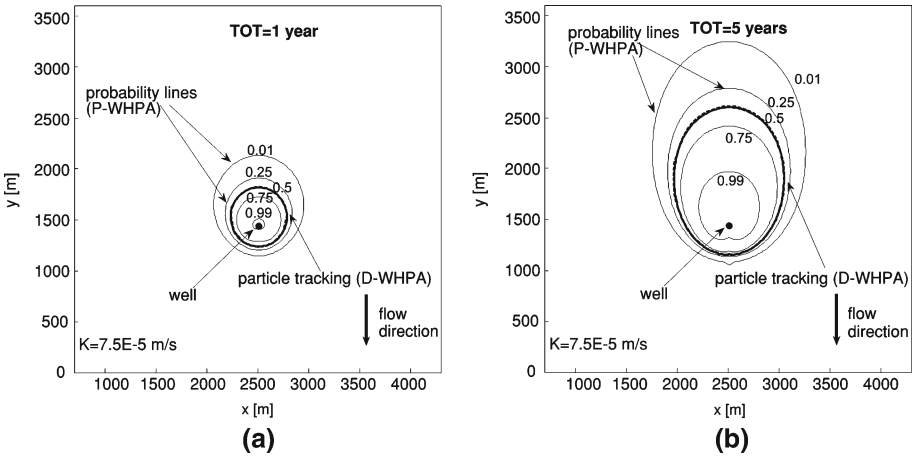


Fig. 1 *Case 1*: comparison between backward probability capture area (P-WHPA) and particle tracking capture area (D-WHPA). P-WHPAs are plotted for several capture probabilities and their shape is compared to the D-WHPA perimeter, at TOT = 1 year (**a**), and TOT = 5 year (**b**). A good agreement between the D-WHPA (*dashed line*) and the 0.5 isoline of the P-WHPA (*thicker continuous line*) can be observed

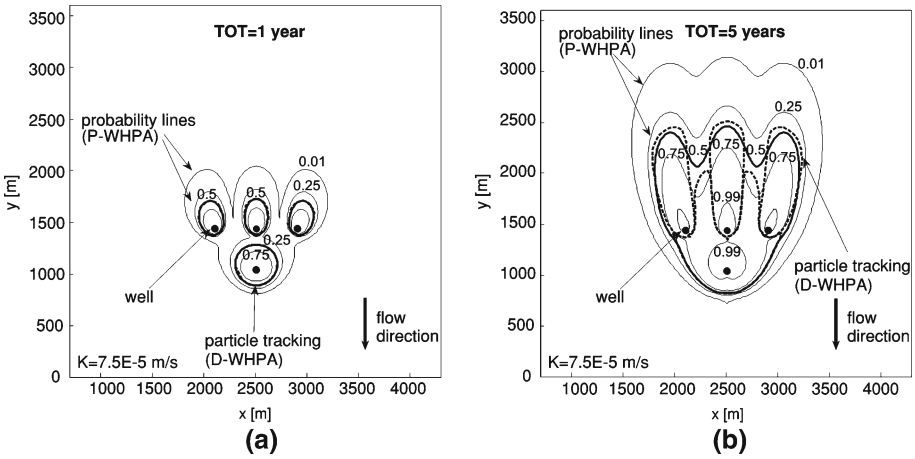


Fig. 2 *Case 2*: comparison between backward probability capture area (P-WHPA, 0.5 isoline as a thicker continuous line) and particle tracking capture area (D-WHPA, dashed line). P-WHPAs are plotted for several capture probabilities and their shape is compared to the D-WHPA perimeter, at TOT = 1 year (**a**), and TOT = 5 year (**b**). For TOT = 5 years, big discrepancies are evident, both in shape and extent

as suggested by Neupauer and Wilson [10] and Frind et al. [11], are no longer accurate, as the probability plume approaches the exit of the model domain, and the results are thus influenced. More in general, discrepancies with respect to the theoretical advective contribution increase as the dispersivity increases, i.e. when the capture probability plume is more dispersed and approaches the domain boundaries earlier. However, at intermediate travel times, at which the capture zones are usually calculated (between 180 days and 5 years), ΔA is closer to zero, not being strongly affected by boundary conditions at the pumping well or at the borders of the domain, nor by the load term definition.

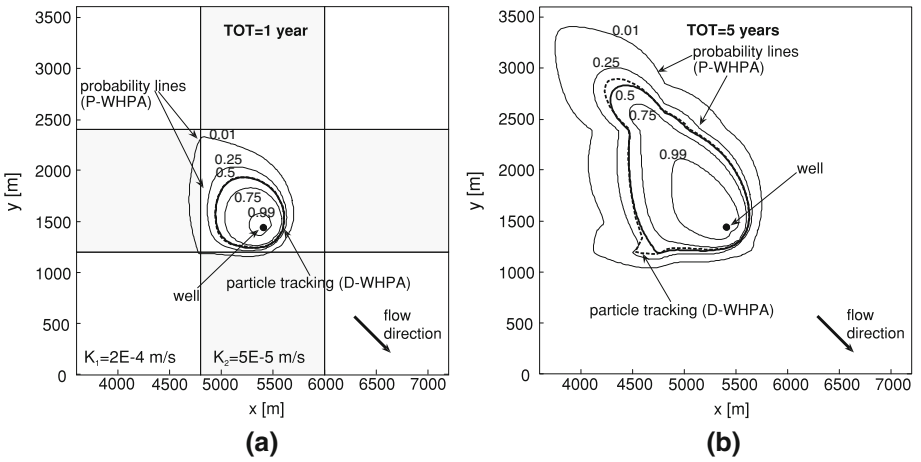


Fig. 3 *Case 3*: comparison between backward probability capture area (P-WHPA, 0.5 isoline as a thicker continuous line) and particle tracking capture area (D-WHPA, dashed line). P-WHPAs are plotted for several capture probabilities and their shape is compared to the D-WHPA perimeter, at TOT = 1 year (**a**), and TOT = 5 year (**b**). The spatial distribution of conductivity is also reported (**a**). For a TOT of 5 years, the P-WHPA is shown to be more smoothed, the D-WHPA being more affected by abrupt changes in conductivity

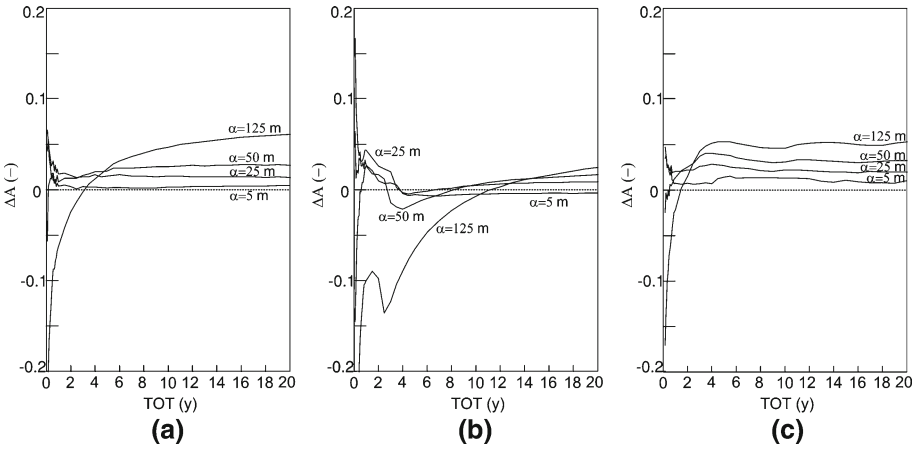


Fig. 4 Normalized differences ΔA are plotted for the 0.5 probability isoline of P-WHPAs for cases 1 (**a**), 2 (**b**) and 3 (**c**). The graphs highlight the influence of boundary conditions on the accuracy of the P-WHPA. For multiple pumping wells (**b**), the importance of dispersion in the P-WHPAs delineation is shown by the negative peaks of the curves, corresponding to the merging of probability plumes for different pumping wells. Dashed lines at $\Delta A = 0$ are reported to guide the eye

For case 2, differences in both shape and extent between the D-WHPA and the 0.5 probability isoline are more evident at intermediate travel times (Fig. 2). The differences in shape increase with the travel time as represented in Fig. 2b. For short times, the four P-WHPAs are completely or partially separated, and differences with respect to the D-WHPAs are not evident. For longer times, however, the 0.5 isolines of the single wells approach each other and merge, thus leading to an overall area larger than the one estimated by the D-WHPAs, and also different in shape. This corresponds to negative values of ΔA (Fig. 4b), observed at

travel times of a few years. Furthermore, the differences increase as the dispersivity increases, being generated by a physical phenomenon (dispersion), rather than numerical approximations (influence of the spatial discretization and the boundary conditions), as for cases 1 and 2. In the region between two different pumping wells, the transverse dispersive contribution to the transport of the capture probability of each well generates non-negligible values, which are summed in the overall P-WHPA. This region cannot be included in any way into the deterministic perimeter, thus it is neglected.

In case 3 (one pumping well, heterogeneous aquifer), for short travel times (Fig. 3a) there is a quite good agreement between the 0.5 probability isoline and the advective area. However, as the simulation time increases, shape differences become more evident: the deterministic capture zone extends more along and perpendicularly to the flow direction, as it enters new regions with a higher conductivity value (Fig. 3b). This means that particles move faster than the capture probability plume when they leave zones at lower conductivity and enter zones at higher conductivity. The final result is a more smoothed shape for the probability capture zone in case of abrupt heterogeneous in the aquifer, compared to the deterministic one, as already outlined by Frind et al. [11,12]. However, the normalized difference between the deterministic capture area and the 0.5 probability isoline (Fig. 4c) is always relatively small (<5% even at very long travel times). This means that, although the shape of the capture zone is quite different, its extent is not.

5.2 P-WHPAs: role of dispersivity and capture probability

Capture areas were then calculated for travel times up to about 20 years, over the extended model domain, to investigate the long-term influence of hydrodynamic dispersion and of the chosen limit probability on completely developed capture areas. The simulations were stopped when the 0.01 probability isoline approached the domain boundaries. Simulations were run for various dispersivities, as mentioned previously, and capture probability isolines were calculated for probabilities ranging from 0.01 to 0.99.

In Figs. 5, 6 and 7 the coefficient C_{Disp} of Eq. 8 is reported as a function of the travel time. C_{Disp} defines the importance of the advective growth rate of the capture area with respect of the theoretical value of the advective growth rate, Q/bn_e. Negative values are related to negative dispersive growth rates, positive values to positive growth rates, the advective contribution always being positive. Plots for case 1 (Fig. 5) are smoothed and more regular than the others, but a common trend can be identified in all three cases. The capture areas identified by the isolines for probabilities higher than 0.5 are characterized by negative coefficients, resulting in a total rate dA/dt lower than the constant advective contribution Q/bn_e, for any travel time and dispersivity employed (see Figs. 5, 6 and 7). Furthermore, for any fixed value of capture probability, the higher the dispersivity, the lower is C_{Disp}.

The opposite holds for probabilities lower than 0.5: C_{Disp} is positive, i.e. the growth rates are always higher than Q/bn_e, and increase as the dispersivity increases. This means that, for short travel times, the isolines are close one to each other (as evident in Figs. 1, 2 and 3), and become more and more distant as the travel time increases. The behaviour is more evident at high dispersivities. In all cases, for long travel times (corresponding to completely developed capture zones), the increase/decrease of the growth rate dA_{Disp}/dt is clearly constant.

For simple cases, with one pumping well and homogeneous parameter distributions (case 1), oscillations of C_{Disp} in time are not very significant, and lose importance for long-term WHPAs, when the capture zone is completely developed (Fig. 5). This general trend, however, is partially altered in cases 2 and 3. For multiple wells (case 2, Fig. 6), important, systematic variations in the general trend happen at low-intermediate travel times, for every

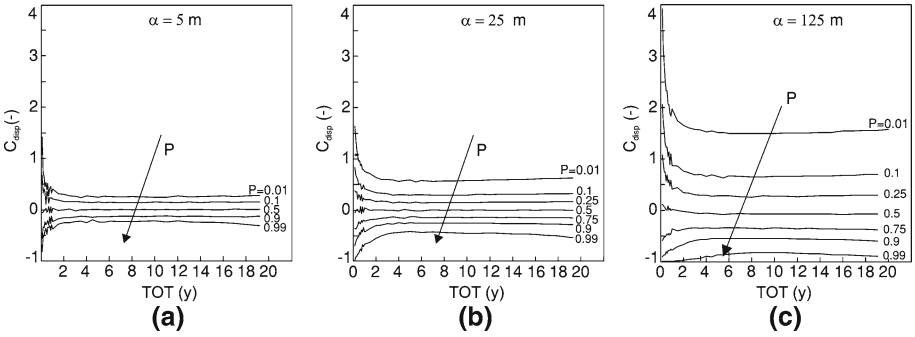


Fig. 5 *Case 1*: plots of the coefficient C_{Disp} in Eq. 8, representing the relative importance of the dispersive contribution to the capture area, with respect to the advective one. C_{Disp} is reported as a function of time, for small (**a**), medium (**b**), and large (**c**) dispersivities, and for different capture probabilities

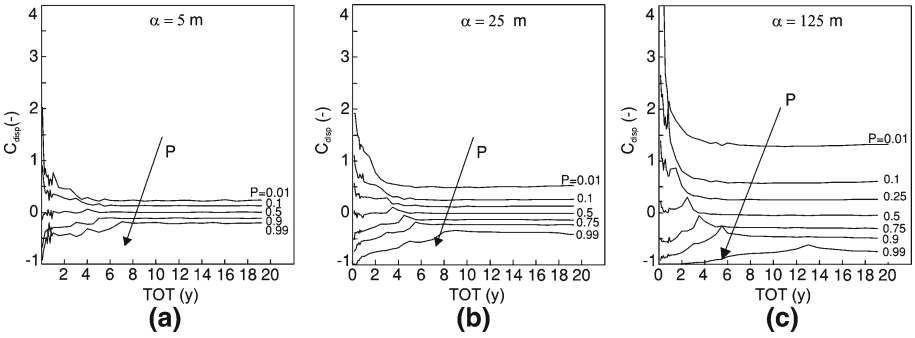


Fig. 6 *Case 2*: plots of the coefficient C_{Disp} in Eq. 8, representing the relative importance of the dispersive contribution to the capture area, with respect to the advective one. C_{Disp} is reported as a function of time, small (**a**), medium (**b**), and large (**c**) dispersivities, and for different capture probabilities. Peaks at small-medium travel times correspond to the merging of the probability plumes of the pumping wells

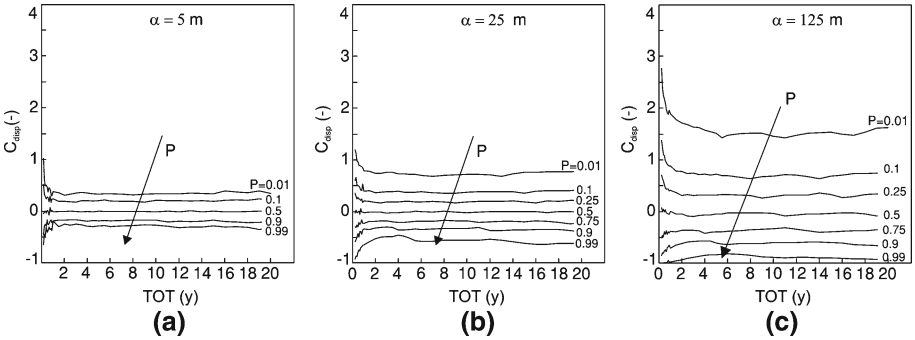


Fig. 7 *Case 3*: plots of the coefficient C_{Disp} in Eq. 8, representing the relative importance of the dispersive contribution to the capture area, with respect to the advective one. C_{Disp} is reported as a function of time, for small (**a**), medium (**b**), and large (**c**) dispersivities, and for different capture probabilities. Oscillations at long travel times reflect the inclusion in the capture zone of squared sub-domains with different conductivities

capture probability, and are mainly due to the merging of the capture zones of different wells, thus deriving from a real physical phenomenon, rather than numerical approximations. For every fixed capture probability, the main peak corresponds to the moment at which the isolines of the four probability plumes merge (see for comparison Fig. 4b). This time is of course smaller for small capture probabilities (corresponding to larger areas), if the dispersivity is fixed, and smaller for larger dispersivities, if the capture probability is fixed. In addition, a secondary peak at lower times can also be identified, which corresponds to the merging of the probability isoline of the two central wells. In case 3, on the contrary, the fluctuations are also present at long travel times, and reflect the heterogeneities in the conductivity distribution (one oscillation for every square sub-domain with high conductivity, see Figs. 3b, 7). It can also be noticed that such oscillations are more evident for high dispersivities: in this case the probability plume involves larger areas, and consequently a higher number of sub-domains with different conductivities.

6 Discussion and conclusions

The application of the backward probability model of Neupauer and Wilson [7] to capture zone delineation allows us to take into account dispersion without high computational costs. The backward probability model includes dispersion into WHPA delineation, and gives capture zones with only one backward simulation, which has the same time cost as an advective-dispersive transport simulation. However, if the model is to be applied to real systems, a good estimate of the dispersion coefficient will be necessary.

The implementation of this model is quite easy, because it can be done with any common transport simulation code, accordingly modified. However, numerical dispersion is to be carefully considered, and future perspectives can take into consideration the use of transport simulators less affected by numerical dispersion than those based on finite difference schemes. Since capture probability is calculated over the entire model domain, the boundaries of P-WHPAs for fixed probabilities can be automatically defined using any simple interpolation over the probability distribution, to obtain an implicit isoline (2D geometry) or iso-surface (3D geometry). On the contrary, the particle tracking methodology always requires post-processing interventions, for a manual or automatic delineation of the areas (eg. the APA algorithm presented in [5]), and the automatic definition of 3D limits involves quite complicated algorithms.

The introduction of a probability component in the model would allow us to fix the perimeter of the protection area in correspondence to a “limit probability”, according to the importance of the water resource, the dangerous centres near it and, more generally, the land management and use. It would be important to establish rules for the determination of this optimal capture probability: if a very low value is selected, the water resource will be more protected, but this could lead to excessive land use limitation. On the contrary, if a high value is adopted, higher than 0.5, the resulting capture zone is even smaller than the corresponding D-WHPA, which can lead to an absolutely inadequate protection of the water resource. However, the application of a probability model for WHPA delineation can be justified in case of multiple wells, also when the 0.5 probability isoline is taken as a limit probability. Public regulations require WHPAs to be delineated at relatively small travel times (180 days, 1 or 5 years), and in this case the differences, both in shape and in extent, between the D-WHPA and the 0.5 isoline of the P-WHPA can be apparent (see case 2 for TOT = 5 years, Figs. 2, 4b), and are due to the physics of the phenomenon (the dispersive component of the probability transport). Moreover, the backward probability model can also be successfully applied in case

of heterogeneities in the aquifer parameter distributions, even if the 0.5 capture probability is chosen as the limit probability. In this case, the method results in more smoothed and regular capture areas, if compared to the particle tracking results. Furthermore, the change in shape does not necessarily correspond to a change in the extent of the capture area, but can result in more efficiently defined WHPAs. As a future perspective, an extension of both methods in transient flow conditions can also be considered, which have been previously shown to have a relevant influence on WHPA delineation [18].

The test cases presented above, although simple 2D applications, show the importance of the choice of the spatial discretization and the extent of the model domain. The roughness of the discretization influences the accuracy of the results at low travel times (see the irregular shape of the first part of the curves in Figs. 5, 6 and 7), when the effect of the boundary condition at the pumping well is not negligible. On the other hand, the extent of the model domain influences the behaviour at long travel times, when the approximation of second-type, homogeneous boundaries is incorrect. Plots of the differences between the extent of the theoretical advective capture area (Q/bn_e) and the corresponding area included by the 0.5 probability isoline, like those reported in Fig. 4, highlight the existence of a region at medium travel times, where the influence of the boundary condition is minimal. Such plots could therefore be used in applications of the backward probability model, in order to check the goodness of the choice of the spatial discretization and the extent of the model domain.

References

1. Bear J, Jacobs M (1965) On the movement of water bodies injected into aquifers. J Hydrol (Amst) 3(1):37–57. doi:10.1016/0022-1694(65)90065-X
2. Javandel I, Tsang CF (1986) Capture-zones type curves: a tool for aquifer cleanup. Ground Water 24(5):616–625. doi:10.1111/j.1745-6584.1986.tb03710.x
3. Fienen MN, Luo J, Kitanidis PK (2005) Semi-analytical homogeneous anisotropic capture zone delineation. J Hydrol (Amst) 312(1–4):39–50. doi:10.1016/j.jhydrol.2005.02.008
4. Pollock D (1989) Documentation of computer programs to compute and display pathlines using results from the US geological survey modular three-dimensional finite-difference ground-water model. In: US geological survey, open file report 89-381, Reston, Virginia
5. Tosco T, Sethi R, Di Molfetta A (2008) An automatic, stagnation point based algorithm for the delineation of wellhead protection areas. Water Resour Res 44(7):W07419. doi:10.1029/2007WR006508
6. Shafer JM (1987) Reverse pathline calculation of time-related capture zones in nonuniform flow. Ground Water 25(3):283–289. doi:10.1111/j.1745-6584.1987.tb02132.x
7. Neupauer RM, Wilson JL (1999) Adjoint method for obtaining backward-in-time location and travel time probabilities of a conservative groundwater contaminant. Water Resour Res 35(11):3389–3398. doi:10.1029/1999WR900190
8. Cornaton F, Perrochet P (2006) Groundwater age, life expectancy and transit time distributions in advective-dispersive systems: 1. Generalized reservoir theory. Adv Water Resour 29(9):1267–1291. doi:10.1016/j.advwatres.2005.10.009
9. Cornaton F, Perrochet P (2006) Groundwater age, life expectancy and transit time distributions in advective–dispersive systems: 2. Reservoir theory for sub-drainage basins. Adv Water Resour 29(9):1292–1305. doi:10.1016/j.advwatres.2005.10.010
10. Neupauer RM, Wilson JL (2004) Numerical implementation of a backward probabilistic model of ground water contamination. Ground Water 42(2):175–189. doi:10.1111/j.1745-6584.2004.tb02666.x
11. Frind EO, Muhammad DS, Molson JW (2002) Delineation of three-dimensional well capture zones for complex multi-aquifer systems. Ground Water 40(6):586–598. doi:10.1111/j.1745-6584.2002.tb02545.x
12. Frind EO, Molson JW, Rudolph DL (2006) Well vulnerability: a quantitative approach for source water protection. Ground Water 44(5):732–742
13. Neupauer RM, Wilson JL (2001) Adjoint-derived location and travel time probabilities for a multidimensional groundwater system. Water Resour Res 37(6):1657–1668. doi:10.1029/2000WR900388

14. Tosco T, Sethi R, Di Molfetta A (2007) A backward probabilistic model to calculate well head protection areas. In: Candela et al (eds) Water pollution in natural porous media at different scales. Assessment of fate impact and indicators. WAPO2 (pp 731–737). Instituto Geologico y Minero de Espana, Barcelona, Spain
15. Tosco T, Sethi R, Di Molfetta A (2006) A probabilistic method for delineation of wellhead protection areas. In: Fifteenth international symposium on mine planning and equipment selection, MPES, Torino (Italy), 20–22 Sept 2006
16. Harbaugh AW, Banta ER, Hill MC, McDonald MG (2000) MODFLOW-2000, the US geological survey modular groundwater model—user guide to modularization concepts and the groundwater process. In: US geological survey, open file report 00-92, Reston, Virginia
17. Zheng C, Wang PP (1999) MT3DMS: a modular three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems; documentation and user's guide. US army engineer research and development center, Vicksburg
18. Vesselinov VV, Robinson BA (2006) Delineation of capture zones in transient groundwater flow systems. In: Bierkens M et al (eds) ModelCARE 2005 calibration and reliability in groundwater modeling: from uncertainty to decision making (Proceedings of ModelCARE 2005, The Hague, The Netherlands, June 2005). IAHS publication 304, wallingford, UK, pp 246–252

Environ Fluid Mech (2010) 10:91–102
DOI 10.1007/s10652-009-9142-7

ORIGINAL ARTICLE

Simulation of groundwater age evolution during the Wisconsinian glaciation over the Canadian landscape

Jean-Michel Lemieux · Edward A. Sudicky

Received: 6 April 2009 / Accepted: 15 June 2009 / Published online: 10 July 2009

Abstract The simulation of groundwater age (residence time) is used to study the impact of the Wisconsinian glaciation on the Canadian continental groundwater flow system. Key processes related to coupled groundwater flow and glaciation modeling are included in the model such as density-dependent flow, hydromechanical loading, subglacial infiltration, glacial isostasy, and permafrost development. It is found that mean groundwater ages span over a large range in values, between zero and 42 Myr; exceedingly old groundwater is found at large depths where there is little groundwater flow because of low permeabilities and because of the presence of very dense brines. During the glacial cycle, old, deep groundwater below the ice sheet mixes with the young subglacial meltwater that infiltrates into the subsurface; the water displacement due to subglacial recharge reaches depths up to 3 km. The depth of penetration of the meltwater is, however, strongly dependent on the permeability of the subsurface rocks, the presence of dense brines and the presence or absence on deep fractures or conductive faults. At the end of the simulation period, it was found that the mean groundwater age in regions affected by the ice sheet advance and retreat is younger than it was at the last interglacial period. This is also true for frozen groundwater in the permafrost area and suggests that significant parts of this water is of glacial origin. Finally, the simulation of groundwater age offers an alternative and pragmatic framework to understand groundwater flow during the Pleistocene and for paleo-hydrogeological studies because it records the history of the groundwater flow paths.

Keywords Groundwater age · Numerical modelling · Wisconsinian glaciation · Canada · Paleo-hydrogeology

J.-M. Lemieux (✉)
Geological Institute, ETH Zurich, 8092 Zurich, Switzerland
e-mail: jmlemieux@alumni.uwaterloo.ca

E. A. Sudicky
Department of Earth and Environmental Sciences, University of Waterloo, 200, University Avenue West, Waterloo, ON N2L 3G1, Canada
e-mail: sudicky@sciborg.uwaterloo.ca

1 Introduction

There are worldwide studies that have provided compelling evidence that significant volumes of Pleistocene waters (glacial meltwater from ice sheets or meteoric water), were driven into subsurface aquifers during past glaciation periods, and sometimes to great depths [11,15–17,21,22,31,33,36,37,41,43–45,53,57]. As pointed out by Person et al. [37], the origin of Pleistocene waters in subsurface aquifers is important to study for several reasons. Pleistocene waters have geochemical properties that are different than autochthonous waters; their quality is usually good, and in some locations, they are an important source of freshwater [16,53]. Glacial meltwater is also associated with the generation of biogenic gas resources that are exploited by the petroleum industry in the Michigan and Illinois basins [33]. Moreover, the lowering of sea level related to the growth of ice sheets have exposed coastal aquifers to freshwater recharge such that, in some locations, the freashwater/sea water interface observed today is not in equilibrium with the current sea level [1,36].

There are several studies that have attempted to quantify Pleistocene recharge rates and patterns using numerical models [5–10,23,34–40,54,57] because they are powerful tools that can take into account the numerous processes involved such as subglacial recharge, permafrost development, land surface isostatic evolution, mechanical loading, sea-level change, etc. However, in most cases, there is a paucity of data available for their calibration such that they are potentially unreliable. For this reason, most of these studies were conducted to explore alternative meltwater recharge scenarios to the groundwater system and to improve our understanding of the coupled processes involved. Potential datasets for the calibration of paleo-hydrogeological models include anomalous pressure measurements [e.g., 3,4], groundwater geochemistry [18], groundwater age estimates using environmental isotopes [15,32] and noble gas data [2,25].

Here, we use groundwater age as a tracer to delineate groundwater recharge zones and groundwater flow paths during a full 120 kyr glaciation cycle. Groundwater age, which can be defined as the time elapsed since the water infiltrated in a recharge zone, is a measure of the residence time of water in the subsurface and can be used to track the recharge history of an aquifer. The numerical model HydroGeoSphere [27,29,52] is used to simulate three-dimensional groundwater age evolution during the Wisconsinian glaciation over the Canadian landscape and up to a depth of 10 km. The objective is to delineate the recharge history of glacial meltwater into the subsurface during the last glaciation period and to better understand the impact of glaciations on the evolution of deep groundwaters. Presently, there is a paucity of deep groundwater age measurements over the Canadian landscape, particularly in remote regions. We feel that our calculations could serve as a guide for selecting strategic locations for the collection of age data to further our model development.

2 The computation of mean groundwater age

Groundwater age is usually defined as the time elapsed since the water infiltrated in a recharge zone [12,20,24]. Each water molecule enters the subsurface at different times and a zero age is assigned to molecules that enter the system at the recharge boundaries.

Water that circulates in the subsurface will follow different flow paths depending on the spatial structure of the geologic materials and the boundary conditions. Some flow paths may be short and shallow, but others may produce deep circulation patterns in the subsurface that will require a long period of time before discharging to the surface. The length of the flow path and the time elapsed since the water entered the subsurface will also affect the water

chemistry. Modelling groundwater age is then equivalent to simulating the advective and dispersive transport of conservative substances in groundwater. On the other hand, the age of a water sample is the average of all the water molecules collected in a sample, some of which may be from different origins. For this reason, groundwater age should also be viewed in a probabilistic framework rather than as a single value at each point in space and time. Cornaton and Perrochet [13] show, using a generalized reservoir theory, that the statistical distribution of the random variable groundwater "age" can be modelled using a formulation based on advective–dispersive transport of a conservative tracer, using appropriate boundary conditions. The transient age probability density function (PDF) is obtained for a finite reservoir, Ω, by solving the boundary value problem:

$$\frac{\partial \phi g_A}{\partial t} = -\nabla \cdot \mathbf{q} g_A + \nabla \cdot \phi \mathbf{D} \nabla g_A + q_I \delta(\tau) - q_O g_A \tag{1}$$

in which:

$$g_A(\mathbf{x}, 0, \tau) = f(\mathbf{x}, \tau) \quad \text{in } \Omega, \tag{2}$$

$$\mathbf{J}_A(\mathbf{x}, t, \tau) \cdot \mathbf{n} = (\mathbf{q} \cdot \mathbf{n}) \delta(\tau) \quad \text{on } \Gamma_- \tag{3}$$

where $g_A(x, t, \tau)$ is the transported transient age PDF, with t being the clock-time, τ the age dimension, ϕ is porosity or mobile water content, $\mathbf{q}$ is the fluid flux vector, $\mathbf{D}$ is the macrodispersion tensor, $\mathbf{J}_A(x, t)$ is the total age mass flux vector, $\mathbf{x} = (x, y, z)$ is the vector of Cartesian coordinates, $\mathbf{n}$ is a normal outward unit vector, and δ is the Dirac delta function which ensures an impulse flux on Γ_-. Γ_- represents the inlet boundaries and Γ_+ refers to the outlet boundaries. The terms q_I and q_O represent fluid sources and sinks, respectively, such as the internal production or extraction of groundwater. The total age mass flux vector $\mathbf{J}_A(x, t, \tau)$ is defined by the sum of the advective and dispersive age fluxes:

$$\mathbf{J}_A(\mathbf{x}, t, \tau) = \mathbf{q} g_A(\mathbf{x}, t, \tau) - \phi \mathbf{D} \nabla g_A(\mathbf{x}, t, \tau) \tag{4}$$

Equation 1 is a 5-D equation close to the one introduced by Delhez et al. [14] in the context of the theory of the age in ocean modelling. It however differs from Delhez et al.'s model by a porosity term required in the subsurface context.

The mean value $a(\mathbf{x}, t)$ of the probability density function $g_A(x, t, \tau)$ is defined by its first temporal moment:

$$a(\mathbf{x}, t) = \int_0^{+\infty} u g_A(\mathbf{x}, t, u) du \tag{5}$$

where u is the age value.

The first moment form of Eq. 1 is the mean age transport equation in which the mean age is defined as the average over a water sample containing water molecules having a range of individual ages. It is given by:

$$-\nabla \cdot \mathbf{q} a + \nabla \cdot \phi \mathbf{D} \nabla a - q_O a + \phi = \frac{\partial \phi a}{\partial t} \quad \text{in} \Omega \tag{6}$$

which is equivalent to that obtained by Goode [20] with a mass conservation approach. The mean groundwater age can be computed by prescribing a "concentration" input that is proportional to the water flux at the inflow boundaries. Equation 6 can be solved by assigning $a(\mathbf{x}, t) = 0$ on the inlet limits. Mean age is continuously generated during groundwater flow, since porosity acts as a source term. This source term indicates that groundwater is aging at the rate of one unit per unit time, on average. Cornaton and Perrochet [13] show that

mean age computations using Eq. 6 are well-suited for simulating isotopic age dates. The implementation of mean age calculations is straightforward in a numerical model that already solves the advection-dispersion equation because only a source term that equals the porosity needs to be included [13,56], as well as the use of proper boundary conditions ($a(\mathbf{x}, t) = 0$ on Γ_-).

3 Numerical model

The numerical model HydroGeoSphere [52] is used to solve the density-dependant groundwater flow equation along with the advective–dispersive solute transport of total dissolved solids (TDS) and mean groundwater age (Eq. 6). Key processes pertaining to groundwater flow modelling during a glaciation period such as hydromechanical loading, subglacial infiltration, isostasy, sea-level change and permafrost development are also included in the model (see [29,30] for details).

The limits of the model encompass the entire Canadian landscape as well as northern United States and Alaska; they were chosen to capture the entire region affected by the last glaciation and are sufficiently large such that the ice-sheet impact is negligible at the margins of the model (Fig. 1). Brick elements (25 × 25 km) are used to discretize the three-dimensional

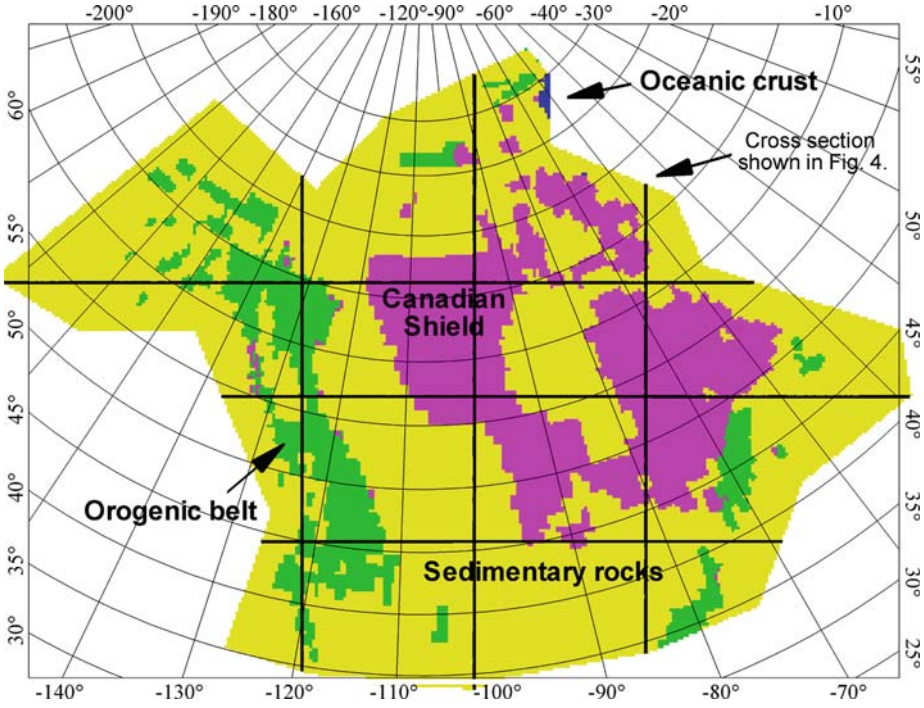


Fig. 1 Limits and hydrogeologic units of the numerical model. The geographic limits of the model are 172.5° W and 42.5° W on the western and eastern boundaries and 34.75° N and 84.75° N on the southern and northern boundaries. Simplified bedrock geology facies used in the numerical model shown in *pink*: Canadian Shield, *dark green*: orogenic belts, *dark blue*: oceanic crust and *yellow*: sedimentary rocks (Constructed from [26] and [58]). The north–south and east–west *black lines* indicate the location of cross sections shown in Fig. 2. Albers equal-area map projection

Table 1 Bedrock hydraulic properties by facies

Parameter	Facies			
	Sedimentary	Shield	Orogen	Oceanic crust
K (isotropic) (m/year)	60.0	30.0	3.0	0.1
Specific storage, S_s (m^{-1})	5.0×10^{-5}	3.0×10^{-6}	1.0×10^{-5}	1.0×10^{-5}
Porosity, ϕ (−)	0.2	0.001	0.05	0.01
Loading efficiency, ζ (−)	0.2	0.2	0.2	0.2
Longitudinal dispersivity, α_L (m)	5000	5000	5000	5000
Transverse dispersivity, α_T (m)	100	100	100	100
Transverse vertical dispersivity, α_V (m)	100	100	100	100
Rock-water mass transfer constant, K_{mt} ($year^{-1}$)	2.3×10^{-7}	2.3×10^{-8}	2.3×10^{-8}	2.3×10^{-9}

model (see [28] for discretization details) which covers about 10,000 km × 6000 km × 10 km with a total 2D area of 2.5 $\times 10^7$ km^2. There are 10 elements in the vertical dimension and the elements size is between 70 m near the surface to 1 km at the bottom of the model.

Hydraulic properties are assigned according to the main bedrock geological units which, at the scale of the Canadian landscape, are simplified into four major facies (see Fig. 1; Table 1). Although this simplification is dramatic, computational constraints, the highly complex three-dimensional architecture of Canadian geology and the lack of systematic hydraulic measurements over the Canadian landscape limit the use of a highly-detailed geological model. The hydrogeological units are: Canadian Shield, Sedimentary basins, Orogenic belts and Ocean crust for which average hydrogeological properties based upon a review of the Canadian hydrogeology are assigned (see Table 1 and [30]).

Boundary conditions used here are inferred from the ice and climatic chronologies during the Wisconsinian glaciation derived from the Memorial University of Newfoundland/University of Toronto Glacial System Model [48–51], as described in Lemieux et al. [28].

A steady-state simulation is first performed to obtain a set of initial conditions for the transient simulation of groundwater flow, solute transport and groundwater age evolution during the last glacial cycle. The transient simulation starts at the last interglacial (LIG), −120 kyr, until present time, 0 kyr, with a 0.1 kyr timestep. Flow boundary conditions are updated every timestep and the age boundary condition is a specified zero age at inflow nodes, both in the subglacial and periglacial environments.

4 Results

4.1 Mean groundwater age distribution at last interglacial

In order to perform the age calculation during the last glacial cycle, initial conditions are needed for the age distribution at LIG. In order to obtain the mean groundwater age distribution at LIG, a steady-state flow field is obtained using the climatic conditions of the LIG (see [30] for details) for which a steady-state age transport simulation is performed. The age transport boundary condition consist of a specified zero age at inflow nodes and a zero-order source term equal to the porosity for each of the rock facies is used. Although it was recognized by Lemieux et al. [30] that interglacial groundwater flow systems may not

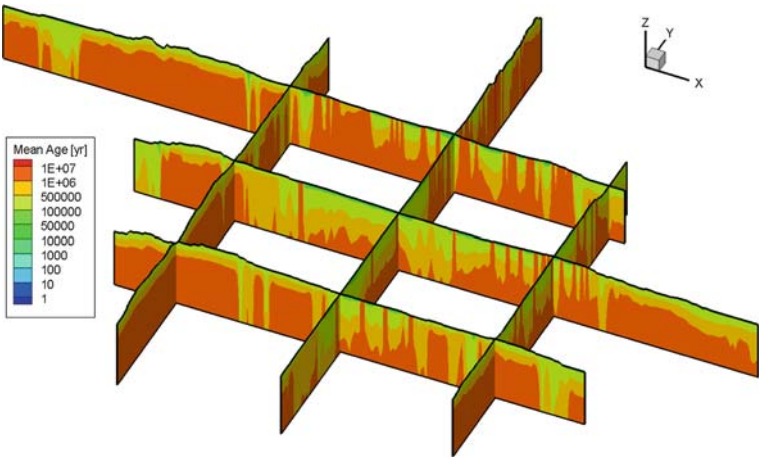


Fig. 2 Mean steady-state groundwater age distribution along selected cross sections for the last interglacial. The lateral extent of the figure is about 10,000 km × 6000 km. Vertical exaggeration is 60×

be in equilibrium with interglacial boundary conditions, it is believed that this assumption is suitable for the exploratory nature of this study. The simulation of several prior glacial cycles to establish a dynamic equilibrium over that last several glacial cycles would have been the most suited approach, but the computational efforts were preclusive.

The calculated three-dimensional mean groundwater age pattern at LIG is shown in Fig. 2 along select cross sections. A clear pattern appears where the mean groundwater age at LIG is younger near the surface and older with depth. The mean groundwater ages span a very large range in values, ranging from zero to 42 Myr. A value of 42 Myr occurs at great depth where there is little groundwater flow. Inspection of the results along the cross sections (Fig. 2) reveals that the youngest water occupies a relatively thin veneer near the surface. Because of the large density of the groundwater at depth, where there is little flow, a large mass of essentially stagnant groundwater exists in the bottom portion of the model.

The surface distribution of the mean groundwater age shows that most of the shallow water has a young age, which suggest that water has recently entered the system. A few locations, mainly along the coast of the province of Québec, Baffin Island and the north shore of Nunavut, show older water. They represent groundwater discharge regions and the age at the outlet is a mixture of deep old water and shallow, younger water. Such occurrences have been documented within the Canadian Shield [19,42] and is usually associated with rapid discharge of deep groundwater along fractures. Although there are no explicit fractures in our model, fractures may only have enhanced the natural discharge of deep brines that we observe in our simulation results.

4.2 Mean groundwater age evolution during the wisconsinian glaciation

Figure 3 shows the simulated hydraulic head and groundwater age at different depths z along with boundary conditions and climatic conditions applied on the surface over the glacial cycle at an observation point located in Timmins, Ontario. In Fig. 3a, subglacial conditions prevail above the horizontal line separating subglacial and periglacial conditions, while periglacial conditions prevail below. The black infilled region indicates that the subsurface is frozen and an unfilled box indicates frost-free conditions. For subglacial conditions, the the ice sheet

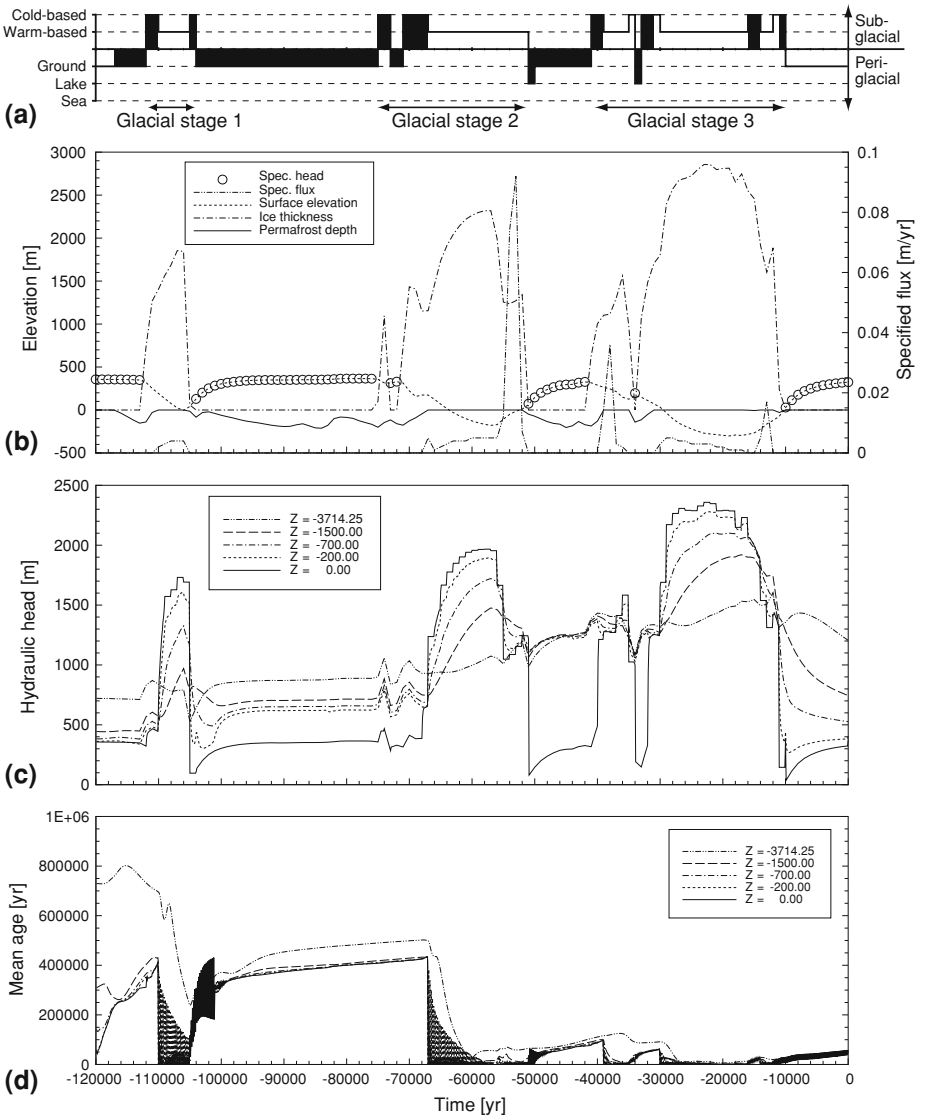


Fig. 3 Observation point located at Timmins, Ontario, Canada, showing (**a**) surface climatic conditions, (**b**) surface boundary conditions, (**c**) simulated hydraulic heads and (**d**) simulated mean age evolution for different depths (z). In (**a**) subglacial conditions prevail above the horizontal line while periglacial conditions prevail below. *Black infill* indicates frozen ground and *white infill* indicates frost-free conditions. For subglacial conditions, the bottom of the ice sheet can be warm-based or cold based. For periglacial conditions, the water table can be at ground surface level, at a periglacial/proglacial lake level or at sea level

can be warm-based, meaning there is subglacial melting, or cold based, meaning there is no subglacial melting. For periglacial conditions, the water table is either at ground surface level, at a periglacial/proglacial lake level or at sea level.

In Fig. 3b, the "elevation" axis on the left is used to designate the value of several variables such as specified hydraulic head and permafrost depth. Other variables displayed are the

ground surface elevation and ice-sheet thickness. The summation of the ice-sheet thickness and the ground surface elevation gives the ice-sheet surface elevation. Permafrost thicknesses are shown as negative values to symbolize there are progressing downward into the ground. The summation of the surface elevation and the permafrost thickness provides the elevation of the lower limit of the permafrost. Specified heads are designated using discrete symbols when this boundary condition is active in the model. A specified water flux, using values corresponding to the axis on the right, is prescribed when the specified head boundary condition is inactive and subglacial meltwater is being produced.

In Figs. 3a and b, it can be seen that there were three glacial stages within the last glacial cycle. At the beginning of every glacial stage, the ice-sheet flows southward from the Arctic and cold temperatures associated with the climate evolution causes freezing of the ground such that thick permafrost develops (see [28,30] for details). As the ice sheet flows, it covers the frozen ground and isolate it from the cold air temperature. The upward geothermal energy migration and frictional heat at the bottom of the ice sheet contribute to the melting of the subglacial permafrost and subglacial meltwater can be driven into the subsurface at great depths under ice-sheet pressure. This can be seen in Fig. 3c where subglacial simulated hydraulic heads increase up to 3 km when the ice sheet overrides the ground surface. A portion of the pore pressure increase is due to the subglacial meltwater recharge and a part is due to the surface loading of the km-thick ice sheet.

The groundwater age evolution is intimately related to the recharge of meltwater into the subsurface. It can be seen in Fig. 3d that the groundwater age decreases significantly, and up to 3 km, when a large recharge event occurs, such as during glacial stage 3. As the glacial period ends and time progresses towards present day, the groundwater age increases steadily due to the lack of mixing with young water derived from the meltwater. At the end of the simulation, the groundwater age is far from its initial values, especially at great depths. It should be noted that subglacial meltwater does not circulate to a depth of 3 km, but displaces water up to this depth, depending on local rock permeabilities, which modifies the groundwater age.

Figures 4a–e show the mean groundwater age distribution and permafrost depth along the eastern north-south cross section shown in Fig. 1. It can be seen that the initial mean age (at $t = -120$ kyr) is very old at depth and becomes younger near the surface, where transit times are shorter. When the ice sheet grows, at -30 kyr, large volumes of the subglacial meltwater mix with the old groundwater such that the mean groundwater age becomes younger. The mixing below the ice sheet occurs at great depths, and locations where the mean groundwater age was originally older than about 1 Myr now has a value between 10 kyr and 100 kyr. A thin veneer of young meltwater is clearly evident below the ice sheet at and following LGM. At the end of the simulation period, it can be seen that mean groundwater ages below regions that were affected by the ice sheet advance and retreat are much younger than they were at LIG.

The results also show that current frozen water within the permafrost area (Fig. 4e) is much younger than at LIG (Fig. 4a). This suggest that a large proportion of this frozen water is composed of subglacial meltwater infiltrated during the advance of the ice sheet and that the geochemical signature of the permafrost should point toward a glacial origin. A geochemical study conducted at the Lupin Mine (NWT, Canada) in a 500 m thick permafrost environment [46,47] was, however, not able to unequivocally confirm this finding.

Finally, it should be noted that little numerical oscillations were observed during the simulation of the large subglacial recharge events (see Fig. 3d), but the overall stability of the model was satisfactory and converging rapidly at every timestep [30].

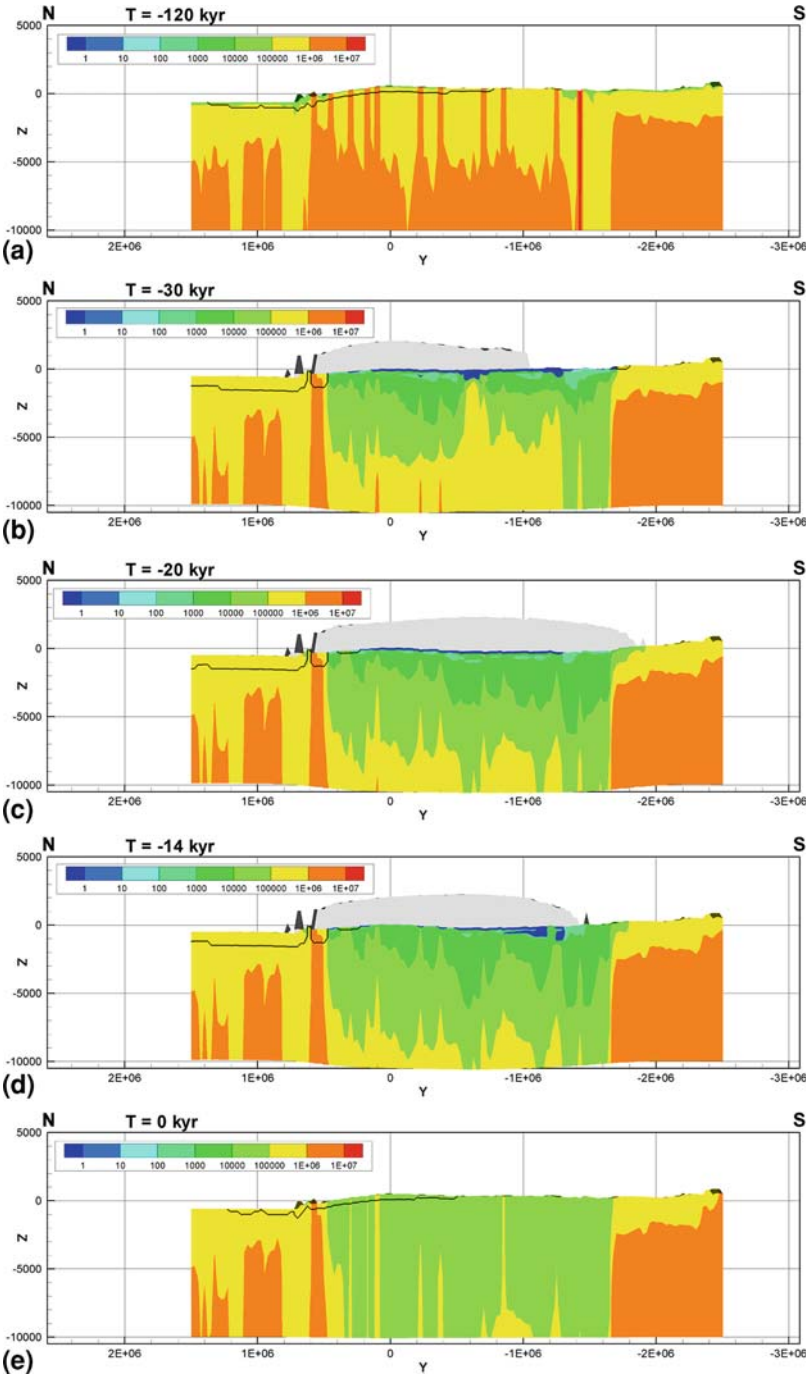


Fig. 4 Mean groundwater age and permafrost depth (*solid line*) along a selected north–south cross section (shown on Fig. 1) at selected time slices: **a** −120 kyr; **b** −30 kyr; **c** −20 kyr; **d** −14 kyr and **e** 0 kyr. Vertical exaggeration: 100×

5 Conclusions

Numerical simulation of groundwater flow is a powerful tool in order to understand the impact of past climates on groundwater resources. In this paper, the concept of groundwater age was applied to compute mean groundwater ages across the Canadian landscape at LIG and throughout the last glacial cycle. It was found that at LIG, mean groundwater ages are younger near the surface and become much older with depth. The mean groundwater ages span over a large range in values, between zero and 42 Myr. Exceedingly old groundwater is found at large depths where there is little groundwater flow because of low permeabilities and because of the presence of very dense brines.

During the glacial cycle, old, deep groundwater below the ice sheet mixes with the young subglacial meltwater that infiltrates into the subsurface and the resulting mean groundwater age becomes younger. Our calculations indicate that glacial meltwaters may have penetrated to to depths as large as one kilometer; however, this is based on a gross regionalization of the subsurface hydrogeology. These calculations are however in agreement with meltwater penetration depths observed in the Illinois, Michigan and Williston basins [22,31,37] and in the Canadian Shield [11,15,41]. They also agree with age estimates reported by McIntosh and Walter [32] in the Great Lake region, which extends up to 50 kyr. Although not presented here (see [30]), simulated total dissolved solids profiles in the Canadian Shield also agree with reported measurements of Frape and Fritz [18].

Clearly, local-scale analysis which bring to bear local geological conditions would serve to refine our calculations. Nevertheless, the work presented here can serve as a framework to refine such calculations and can be useful for choosing potential field sampling locations for further study, sitting of nuclear waste repositories, and identify potential reservoirs of high-quality drinking water.

Acknowledgments Financial support for this study was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Ontario Graduate Scholarship (OGS) Program as scholarships to J.-M. Lemieux and also from NSERC and the Canada Research Chairs program as research funding to E.A. Sudicky. We would like to thank guest-editor Fabien Cornaton for insightful comments on groundwater age simulation along with two anonymous reviewers who provided constructive comments on the original manuscript.

References

1. Akouvi A, Dray M, Violette S, de Marsily G, Zupp GM (2008) The sedimentary coastal basin of Togo: example of a multilayered aquifer still influenced by a palaeo-seawater intrusion. Hydrogeol J 16(3): 419–436. doi:10.1007/s10040-007-0246-1
2. Andrews JN, Lee DJ (1979) Inert gases in groundwater from the Bunter sandstone of England as indicators of age and palaeoclimatic trends. J Hydrol 41:233–252
3. Bahr JM, Moline GR, Nadon G (1994) Anomalous pressures in the deep Michigan Basin. In: Ortoleva P (ed) Basin compartments and seals, AAPG Memoir 61, AAPG, pp 153–165
4. Bekele EB, Rostron BJ, Person MA (2003) Fluid pressure implications of erosion unloading, basin hydrodynamics and glaciation in the Alberta Basin. J Geochem Explor 78(79):143–147
5. Bense VF, Person MA (2008) Transient hydrodynamics within inter-cratonic sedimentary basins during glacial cycles. J Geophys Res 113:F04005. doi:10.1029/2007JF000969
6. Boulton GS, Slot T, Blessing K, Glasbergen P, Leijnse T, van Gijssel K (1993) Deep circulation of groundwater in overpressured subglacial aquifers and it's geological consequences. Quat Sci Rev 12:739–745
7. Boulton GS, Caban PB, van Gijssel K (1995) Groundwater flow beneath ice sheets: Part I - Large scale patterns. Quat Sci Rev 14:545–562
8. Boulton GS, Caban PB, van Gissel K, Leijnse A, Punkari M, van Weert FHA (1996) The impact of glaciation on the groundwater regime of Northwest Europe. Global Planet Change 12(1–4):397–413

9. Breemer CH, Clark PU, Haggerty R (2002) Modeling the subglacial hydrology of the late Pleistocene Lake Michigan Lobe, Laurentide Ice Sheet. Geol Soc Am Bull 114(6):665–674
10. Chan T, Christiansson R, Boulton GS, Ericsson LO, Hartikainen J, Jensen MR, Ivars DM, Stanchell FW, Vistrand P, Wallroth T (2005) DECOVALEX III BMT3/BENCHPAR WP4, the thermo-hydro-mechanical responses to a glacial cycle and their potential implications for deep geological disposal of nuclear fuel waste in a fractured crystalline rock mass. Int J Rock Mech Min 42(5-6):805–827
11. Clark ID, Douglas M, Raven K, Bottomley D (2000) Recharge and preservation of Laurentide glacial melt water in the Canadian Shield. Ground Water 38(5):735–742
12. Cornaton F (2004) Deterministic models of groundwater age, life expectancy and transit time distributions in advective–dispersive systems. PhD thesis, University of Neuchâtel, Switzerland
13. Cornaton F, Perrochet P (2006) Groundwater age, life expectancy and transit time distributions in advectivefbdispersive systems: 1. Generalized reservoir theory. Adv Water Resour 29(9):1267–1291
14. Delhez EJM, Campin JM, Hirst AC, Deleersnijder E (1999) Toward a general theory of the age in ocean modelling. Ocean Modelling 1:17–27
15. Douglas M, Clark ID, Raven K, Bottomley D (2000) Groundwater mixing dynamics at a Canadian Shield mine. J Hydrol 235(1–2):88–103
16. Edmunds WM (2001) Paleowaters in European coastal aquifers—the goals and main conclusions of the PALAEAUX project. In: Edmunds WM, Milne CJ (eds) Paleowaters in coastal Europe: evolution of groundwater since the late Pleistocene, Geological Society, London, pp 1–16, Special Publications, 189
17. Ferguson GAG, Betcher RN, Grasby SE (2007) Hydrogeology of the Winnipeg formation in Manitoba, Canada. Hydrogeol J 15: 573–587. doi:10.1007/s10040-006-0130-4
18. Frape SK, Fritz P (1987) Geochemical trends for groundwaters from the Canadian Shield. In: Fritz P, Frape SK (eds) Saline water and gases in crystalline rocks, Geological Association of Canada Special paper 33, pp 19–38
19. Gascoyne M, Sheppard MI (1993) Evidence of terrestrial discharge of deep groundwater on the Canadian Shield from Helium in soil gases. Environ Sci Technol 27(12):2420–2426
20. Goode DJ (1996) Direct simulation of groundwater age. Water Resour Res 32(2):289–296
21. Grasby SE, Chen Z (2005) Subglacial recharge into the western Canada sedimentary basin—impact of Pleistocene glaciation on basin hydrodynamics. Geol Soc Am Bull 117(3/4):500–514
22. Grasby S, Osadetz K, Betcher R, Render F (2000) Reversal of the regional-scale flow system of the Williston basin in response to Pleistocene glaciation. Geology 28(7):635–638
23. Hoaglund JRIII, Kolak JJ, Long DT, Larson GJ (2004) Analysis of modern and Pleistocene hydrologic exchange between Saginaw Bay (Lake Huron) and the Saginaw Lowlands area. Geol Soc Am Bull 116(1–2): 3–15
24. Kazemi GA, Lehr JH, Perrochet P (2006) Groundwater age. Wiley, Hoboken
25. Klump S, Grundl T, Purtschert R, Kipfer R (2008) Groundwater and climate dynamics derived from noble gas, ^{14}C, and stable isotope data. Geology 36(5):395–398
26. Laske G, Masters G (1997) A global digital map of sediment thickness. EOS Trans AGU 78:F483
27. Lemieux JM (2006) Impact of the Wisconsinian glaciation on Canadian continental groundwater flow. PhD thesis, University of Waterloo, Waterloo, ON, Canada, available online at http://hdl.handle.net/10012/2654
28. Lemieux JM, Sudicky EA, Peltier WR, Tarasov L (2008) Dynamics of groundwater recharge and seepage over the Canadian landscape during the Wisconsinian glaciation. J Geophys Res 113:F01011. doi:10.1029/2007JF000838
29. Lemieux JM, Sudicky EA, Peltier WR, Tarasov L (2008) Simulating the impact of glaciations on continental groundwater flow systems: 1. Relevant processes and model formulation. J Geophys Res 113:F03017. doi:10.1029/2007JF000928
30. Lemieux JM, Sudicky EA, Peltier WR, Tarasov L (2008) Simulating the impact of glaciations on continental groundwater flow systems: 2. Model application to the Wisconsinian glaciation over the Canadian landscape. J Geophys Res 113:F03018. doi:10.1029/2007JF000929
31. McIntosh JC, Walter LM (2005) Volumetrically significant recharge of Pleistocene glacial meltwaters into epicratonic basins: constraints imposed by solute mass balances. Chem Geol 222(3-4):292–309
32. McIntosh JC, Walter LM (2006) Paleowaters in silurian-devonian carbonate aquifers: geochemical evolution of groundwater in the great lakes region since the late pleistocene. Geochim Cosmochim Acta 70:2454–2479
33. McIntosh JC, Walter LM, Martini AM (2002) Pleistocene recharge to midcontinent basins: Effects on salinity structure and microbial gas generation. Geochim Cosmochim Acta 66(10):1681–1700
34. McIntosh JC, Garven G, Hanor JS (2005) Modeling variable-density fluid flow and solute transport in glaciated sedimentary basins. Eos Trans AGU 86(52), fall Meet. Suppl., Abstract H11D-1310

35. Moeller CA, Mickelson DM, Anderson MP, Winguth C (2007) Groundwater flow beneath Late Weichselian glacier ice in Nordfjord, Norway. J Glaciol 53(180):84–90
36. Person M, Dugan B, Swenson JB, Urbano L, Stott C, Taylor J, Willett M (2003) Pleistocene hydrogeology of the Atlantic continental shelf, New England. Geol Soc Am Bull 115(11):1324–1343
37. Person M, McIntosh J, Bense V, Remenda VH (2007) Pleistocene hydrology of North America: the role of ice sheets in reorganizing groundwater flow systems. Rev Geophys 45:RG3007. doi:10.1029/2006RG000206
38. Piotrowski JA (1997) Subglacial groundwater flow during the last glaciation in northwestern Germany. Sediment Geol 111:217–224
39. Piotrowski JA (1997) Subglacial hydrology in north-western Germany during the last glaciation: groundwater flow, tunnel, valleys and hydrological cycles. Quat Sci Rev 16:169–185
40. Provost AM, Voss CI, Neuzil CE (1998) Glaciation and regional ground-water flow in the Fennoscandian shield; Site 94. SKI Report 96:11, Swedish Nuclear Power Inspectorate, Stockholm, Sweden
41. Raven KG, Bottomley DJ, Sweezey RA, Smedley JA, Ruttan TJ (1987) Hydrogeological characterization of the East Bull Lake research area. Tech. rep., National Hydrology Research Institute Paper No. 31, Inland Waters Directorate Series No. 160, Environment Canada, Ottawa, Canada
42. Sheppard MI, Thibault DH, Milton GM, Reid JAK, Smith PA, Steve K (1995) Characterization of a suspected terrestrial deep groundwater discharge area on the Canadian Precambrian shield. J Contam Hydrol 18(1):59–84
43. Siegel DI (1989) Geochemistry of the Cambrian-Ordovician aquifer system in the Northern Midwest, United States. U.S. Geological Survey Professional Paper 1405-D, USGS, Denver, Colorado
44. Siegel DI (1991) Evidence for dilution of deep, confined ground water by vertical recharge of isotopically heavy Pleistocene water. Geology 19(5):433–436
45. Siegel DI, Mandle RJ (1983) Isotopic evidence for glacial meltwater recharge to the Cambrian-Ordovician aquifer, North-Central United States. Quart Res 22(3):328–335
46. Stotler RL (2008) Evolution of Canadian Shield groundwaters and gases: influence of deep permafrost. PhD thesis, University of Waterloo, Waterloo, ON, Canada. Available online at http://hdl.handle.net/10012/4020
47. Stotler RL, Frape SK, Ruskeeniemi T, Ahonen L, Onstott TC, Hobbs MY (2009) Hydrogeochemistry of groundwaters in and below the base of thick permafrost at Lupin, Nunavut, Canada. J Hydrol 373(1–2):80–95. doi:10.1016/j.jhydrol.2009.04.013
48. Tarasov L, Peltier WR (1999) Impact of thermomechanical ice sheet coupling on a model of the 100 kyr ice age cycle. J Geophys Res 104(D8):9517–9545
49. Tarasov L, Peltier WR (2002) Greenland glacial history and local geodynamic consequences. Geophys J Int 150:198–229
50. Tarasov L, Peltier WR (2004) A geophysically constrained large ensemble analysis of the deglacial history of the North American ice-sheet complex. Quat Sci Rev 23(3–4):359–388
51. Tarasov L, Peltier WR (2006) A calibrated deglacial drainage chronology for the North American continent: Evidence of an Arctic trigger for the Younger Dryas. Quat Sci Rev 25(7–8):659–688
52. Therrien R, McLaren R, Sudicky EA, Panday S (2006) HydroGeoSphere, a three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. Groundwater Simulations Group, Waterloo, Ontario, Canada
53. Vaikmäe R, Vallner L, Loosli HH, Blaser PC, Julliard-Tardent M (2001) Paleogroundwater of glacial origin in the Cambrian-Vendian aquifer of northern Estonia. In: Edmunds WM, Milne CJ (eds) Paleowaters in coastal Europe: evolution of groundwater since the late Pleistocene. Geological Society, London, pp 17–27, Special Publications, 189
54. van Weert FHA, van Gijssel K, Leijnse A, Boulton GS (1997) The effects of Pleistocene glaciations on the geohydrological system of Northwest Europe. J Hydrol 195(1–4):137–159
55. Vidstrand P, Wallroth T, Ericsson LO (2008) Coupled HM effects in a crystalline rock mass due to glaciation: indicative results from groundwater flow regimes and stresses from an FEM study. Bull Eng Geol Environ 67(2): 187–197. doi:10.1007/s10064-008-0123-8
56. Voss CI, Provost AM (2001) Recharge-area nuclear waste repository in southeastern Sweden—demonstration of hydrogeologic siting concepts and techniques. SKI Report 01:44, Swedish Nuclear Power Inspectorate, Stockholm, Sweden
57. Weaver TR, Frape SK, Cherry JA (1995) Recent cross-formational fluid flow and mixing in the shallow Michigan Basin. Geol Soc Am Bull 107(6):697–707
58. Wheeler JO, Hoffman PF, Cardand KD, Davidson A, Sanford BV, Okulitch AV, Roest WR (1997) Geological map of Canada, Map D1860A. Geological Survey of Canada, Ottawa, Canada

Environ Fluid Mech (2010) 10:103–120
DOI 10.1007/s10652-009-9147-2

ORIGINAL ARTICLE

Quantification of advective solute travel times and mass transport through hydrological catchments

Amélie Darracq · Georgia Destouni · Klas Persson · Carmen Prieto · Jerker Jarsjö

Received: 29 March 2009 / Accepted: 10 August 2009 / Published online: 25 August 2009

Abstract This study has investigated and outlined the possible quantification and mapping of the distributions of advective solute travel times through hydrological catchments. These distributions are essential for understanding how local water flow and solute transport and attenuation processes affect the catchment-scale transport of solute, for instance with regard to biogeochemical cycling, contamination persistence and water quality. The spatial and statistical distributions of advective travel times have been quantified based on reported hydrological flow and mass-transport modeling results for two coastal Swedish catchments. The results show that the combined travel time distributions for the groundwater-stream network continuum in these catchments depend largely on the groundwater system and model representation, in particular regarding the spatial variability of groundwater hydraulic parameters (conductivity, porosity and gradient), and the possible contributions of slower/deeper groundwater flow components. Model assumptions about the spatial variability of groundwater hydraulic properties can thus greatly affect model results of catchment-scale solute spreading. The importance of advective travel time variability for the total mass delivery of naturally attenuated solute (tracer, nutrient, pollutant) from a catchment to its downstream water recipient depends on the product of catchment-average physical travel time and attenuation rate.

Keywords Hydrology · Travel time · Solute transport · Natural attenuation · Catchment · Groundwater–surface water interactions

1 Introduction

Travel time distributions (or also called transit time distributions, system response functions, weighting functions [1]) are useful descriptors of how small-scale physical transport processes and their dynamics combine to determine larger-scale transport behavior in

A. Darracq (✉) · G. Destouni · K. Persson · C. Prieto · J. Jarsjö
Department of Physical Geography and Quaternary Geology, Stockholm University, Stockholm, Sweden
e-mail: amelie.darracq@natgeo.su.se

catchments. A travel time distribution can be determined from the mass flow response or breakthrough of an instantaneous, conservative tracer input in a catchment area with zero background tracer concentration [2]. This integrates the physical transport of tracer in all the pathways that carry it through the catchment into a single distribution of the timescales of the tracer transport through the catchment. This distribution quantifies the physical spreading of tracer mass in that catchment-scale transport process and can aid significantly in the understanding and quantification of the processes involved in the catchment-scale water flow and solute (tracer, nutrient, pollutant) transport [3–7]. These processes control also biogeochemical cycling, contamination persistence and water quality [8].

Purely physical, advective solute travel times through a catchment depend on the transport velocities and transport pathway lengths between the solute input and output locations. These physical transport quantities and associated solute travel times may vary widely for different solute input locations, an influence that may be referred to as geomorphologic dispersion in the stream networks [9] and analogously in the subsurface transport process from the land surface to the streams [3,5]. Even for solute input at a single well-defined point-source location in a stream, the downstream solute transport and travel times through the stream network are subject to dispersion due to transport velocity variations among and along different transport pathways [10,11]. The solute may also undergo diffusive mass transfer between mobile and immobile water in the hyporheic [12] and dead zones [13].

Different factors and mechanisms may control the dynamics and timescales of hydrological mass transport through drainage basins [14–17]. The travel time variability that exists at all scales in all catchments may to smaller or greater degree mask some important effects of these factors and mechanisms and lead to disparities between different solute transport models and results for different measurement and model scales [4,18–20]. Such disparities limit our capability to incorporate field knowledge and to interpret and transfer results in and between different modeling frameworks and catchments.

In general, realistic distributions of solute travel times in catchments have been pointed out as essential information for accurately representing the catchment-scale process of solute transport, yet commonly difficult to quantify and constrain [1]. In this paper, we investigate the possible quantification of solute travel time distributions in catchments, by the use of reported results on catchment-scale hydrological flow and mass transport modeling in two well-investigated, coastal Swedish catchments areas (Fig. 1): the Norrström drainage basin [19–24] and the Forsmark catchment area [25–28]. In particular, we investigate here the role of different possible groundwater system representations for the quantification of solute travel times through catchments. We further investigate the effects of different travel time distribution quantifications for the resulting solute mass transport from the catchments to downstream recipients.

2 Materials and methods

This section describes the general approach to quantify travel times and their spatial and statistical distributions in catchments, and the solute mass delivery from the catchments. The modeling details for the two specific investigated catchment areas are given in the Appendix.

2.1 General quantification approach

Numerous studies have in the past decades developed and used theoretical conceptualization and quantification approaches that account for the large-scale, physical spreading of solute

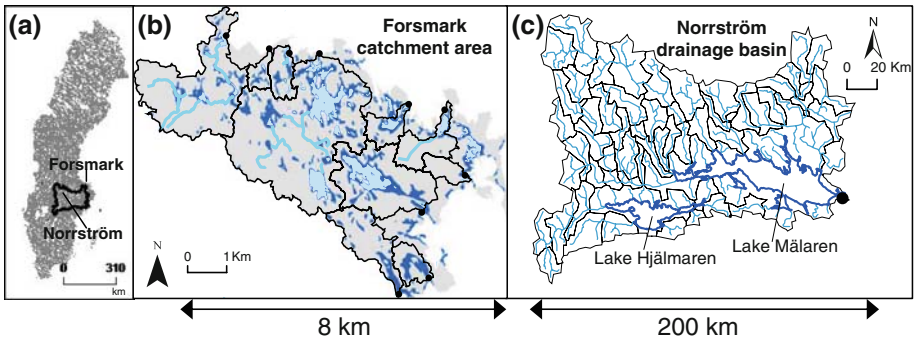


Fig. 1 **a** Location of the Forsmark and Norrström catchment areas within Sweden; **b** the Forsmark catchment area with its: surface water system, including streams (*blue lines*), lakes (*soft blue*), wetlands (*dark blue*) and ten main stream outlets to the coast (*black dots*), the subcatchment boundaries of which are drawn with black lines; and **c** the Norrström drainage basin with its: river network (*soft blue*), major lakes (*dark blue*) and subcatchment boundaries (*black lines*), and outlet to the sea (*black dot*). The small near-coastal zones in between the main stream outlets of the Forsmark catchment area discharge mainly groundwater to the sea

transport in heterogeneous geological formations in terms of prevailing advection variability; see for instance Dagan [29] and Rubin [30] for reviews of such different approaches. Some of these approaches have particularly developed the use of advective solute travel times and their distributions as a main basis for Lagrangian conceptualizations and derivations of field-scale solute transport and spreading in different subsurface water systems (unsaturated soil and groundwater, e.g. [2,31–47]). Parallel studies also extended the theoretical basis of the Lagrangian travel time-based approaches to link the solute transport through the different water subsystems (unsaturated soil, groundwater, streams and stream networks) that are hydraulically connected at the larger scales of hydrological catchments [4,5,44,48–50].

The advective travel time distributions that have been used in most previous studies have been approximated by assuming some common type of probability density function (e.g., log-normal, inverse Gaussian), which can be fully parameterized based on knowledge of only the possible mean and variance of solute travel times in the considered transport system. In this study, we adopt the Lagrangian advective travel time-based approach and extend it to quantify and investigate entire distributions of advective solute travel times in the two Swedish catchment cases and their different water subsystems, by the use of the flow and mass transport results that have already been modeled, tested against all available monitoring data and reported in a series of previous published studies of these catchment areas [19–28].

In general, studies that use Lagrangian advective travel time-based approaches do so because they focus on the macro-dispersion of solute transport due to large-scale advection variability. In such large-scale contexts, the local mixing that occurs between and along different advection pathways due to pore-scale dispersion and molecular diffusion in mobile water is often neglected [29,30]. However, if and where account of these processes is needed, they can be linked to the advective travel-time based model representations, with such linked studies showing that neglecting local dispersion and diffusion within mobile water does not much affect the large-scale mean mass flow rate or concentration, but may lead to the overestimation of local mass flux and concentration variances [51–54].

Jarsjö et al. [26] have also specifically investigated the effect of local random variability around mean advective travel time, e.g. due to local dispersion and diffusion, for the Forsmark catchment area, which constitutes one of the two specific catchment cases of the present

study. The results of Jarsjö et al. [26] confirm that also in this specific catchment case, the effect of such local variability within mobile water is small on the expected large-scale solute transport. The present paper therefore focuses on the quantification of advective solute travel times as the main, first and necessary step towards quantifying catchment-scale pollutant transport and its dominant timescales.

With regard to more significant effects of diffusive mass transfer between mobile and immobile water zones, it is one of the main advantages of Lagrangian travel time-based approaches that their first-step quantification of advective solute travel time distributions can readily be coupled with relevant process models of diffusive mass transfer [5,33,34, 36,37,46,49], as well as with biogeochemical reaction process models of various degrees of complexity [34,38–43,45,47,50,55]. The resulting coupled advection-sorption and/or advection-reaction models account then both for the physical solute spreading effect of advection variability and the diffusive mass transfer and/or biogeochemical reaction process effects on large-scale pollutant transport. In this study, this extension possibility will only be illustrated for a generic, hypothetical and simple case of solute undergoing first-order attenuation. This illustration is made to show some general first-order effects of the advective solute travel time variability and distributions on the large-scale solute mass delivery from different parts of a catchment area and the whole catchment to a downstream water recipient. More complex investigations of diffusive mass transfer and/or reactive transport of specific pollutants are outside the scope of the present study, but we note with reference to the above-cited diffusive-reactive transport studies that such investigations are facilitated by the present, first-step quantification of advective solute travel time distributions.

In the present quantification of these distributions, we further neglect the travel time components in the essentially vertical transport from the soil surface down to the groundwater table, for simplicity and in comparison to the dominant, large travel times in the groundwater system, from the groundwater table to the groundwater–stream interface. This is by no means any necessary neglect requirement in the Lagrangian advective travel time-based approach. On the contrary, this approach has already been developed and used for linking the travel times and travel time distributions of the essentially vertical unsaturated zone transport with the essentially horizontal transport in the groundwater zone and its travel times and travel time distributions, in order to represent the large-scale solute transport through the integrated soil–groundwater system [44,49]. If and where the advective travel times through the unsaturated zone are quantified or expected to be significant in relation to the groundwater travel times, the same methodology can readily be used to extend the present quantification results to consider and integrate the unsaturated zone travel time components in the combined catchment-scale travel time distribution.

Furthermore, the previously reported flow and transport modeling of the specific two Swedish catchment cases considered in this study certainly include soil properties and processes [19–28]. The main reason and motivation for the present primary focus on quantifying and linking the groundwater and stream network travel times is that the soil depth of quaternary deposits above the bedrock is generally small (around 1–2 m, up to maximum 5 m) in both these catchment areas, with the groundwater table being on average about one meter below the soil surface. In contrast, the horizontal transport lengths are three to five orders of magnitude greater in both the small Forsmark catchment area of 30 km^2 and the much larger Norrström catchment area of 22000 km^2 (Fig. 1). We believe that these conditions justify a primary focus on the advective travel times of the horizontal transport through the groundwater–surface water continuum, especially with the particular aim of the present study to investigate the role of different possible groundwater system characterizations and model representations for the quantification of solute travel times through catchments. Also in this

respect, the present results facilitate follow-up studies that can incorporate the additional travel time components of vertical transport through the unsaturated zone and investigate their effects on the combined total distributions of travel times through whole catchments.

2.2 General travel time and mass delivery fraction calculations

We consider solute mass releases from different sources on the land surface and/or directly into the streams, lakes of each catchment area, which discharges its water and waterborne solute mass flows into a downstream recipient. This recipient is the coastal water for both catchment areas investigated here, with a single coastal outlet in the Norrström basin, and multiple stream outlets as well as zones of direct groundwater discharge to the coast in the Forsmark catchment area (Fig. 1).

The advective solute travel time from any mass input location a_{gw} along the mean groundwater flow direction x_{gw} to a given control plane location x_{CP} along that direction (e.g., at the nearest groundwater–stream or groundwater–coast interface), and a_s along the mean stream/surface water flow direction x_s to the outlet x_{out} is quantified as $\tau_{gw} = \int_{a_{gw}}^{x_{CP}} \frac{dX_{gw}}{v_{gw}(X_{gw})}$ and $\tau_s = \int_{a_s}^{x_{out}} \frac{dX_s}{v_s(X_s)}$, with $v_{gw}(X_{gw})$ and $v_s(X_s)$ being the local transport velocity in the x_{gw} and x_s direction at advective solute transport position X_{gw} and X_s along x_{gw} and x_s, respectively. For any solute input location at the catchment surface, a total flow-weighted average travel time T to the recipient can be quantified as $T = (1-\beta_{gw})\tau_s + \beta_{gw}(\tau_{gw} + \tau_s)$, or just $T = \tau_{gw}$ in near-coastal catchment zones with only groundwater flow to the coast (see Forsmark area in Fig. 1b), where β_{gw} is the flow fraction of the total precipitation surplus (precipitation minus actual evapotranspiration) at the catchment surface that infiltrates the soil–groundwater system, and $(1 - \beta_{gw})$ is the complementary fraction that flows directly into the recipient through only surface runoff and stream flow. In general, $\beta_{gw} = 0$ in all catchment area parts that are covered by surface water, while it may generally have different values at different land surface locations. In the present calculations, explained further in the specific catchment sections below, β_{gw} is assumed steady in time and is estimated mainly from available land cover information for the Forsmark catchment area, and both land cover and river network information for the Norrström drainage basin.

The quantification of delivered solute mass fraction from the catchment surface to the coast is made here for solute that undergoes first-order attenuation $exp(-\lambda_{gw}\tau_{gw})$ in the subsurface and $exp(-\lambda_s\tau_s)$ in the stream network system of the catchment. For simplicity, because we do not investigate any specific tracer, nutrient or pollutant transport situation, we illustrate results for $\lambda_{gw} = \lambda_s = \lambda$, so that the mass delivery fraction α from any input location to the coast is quantified as $\alpha = exp(-\lambda T)$. The total delivered mass fraction from the whole catchment to the coast is quantified by the mean value, $\overline{\alpha}$, of α for uniform mass input over the whole catchment surface.

2.3 The Forsmark and Norrström catchment areas

The Forsmark catchment area is relatively small (30 km^2) and characterized by uniquely high-resolved (on 10 m × 10 m grid cells) measured and modeled hydrological data [25–28]. The Norrström drainage basin is relatively large (22000 km^2), with much coarser (1 km × 1 km) resolution of available measured and modeled data [19–24]. This section shortly describes the main flow and transport characteristics of these areas. More details on the modeling and calculations for each area are given in Appendix.

2.3.1 Forsmark catchment area

The Forsmark catchment area (Fig. 1b) contains the subcatchments (black contours, Fig. 1b) of ten main stream-outlets to the Baltic Sea, with small near-coastal catchment zones in between discharging mainly groundwater to the sea. The Forsmark catchment area is currently of particular interest due to its consideration by the Swedish Nuclear Fuel and Waste Management Company as a possible suitable location for a deep repository for spent nuclear fuel, e.g., [26,56]. Quaternary deposits cover a major part of the surface and are dominated by till (mainly sandy). The land surface is mainly covered by forest. There are also many lakes and wetlands, with the wetlands being sometimes partially forested. Figure 1b shows the ten main connected stream networks (with outlets to the coast shown with black dots) and their catchments (with boundaries drawn with black lines), where the dominating flow and transport pathways from the land surface to the coastal waters go through the coupled groundwater–stream system to the nearest surface water (stream, lake, wetland) and through the associated stream network to the coast. The remaining surface area in the Forsmark catchment represents the about 11% of the total catchment surface area that is covered by the small, near-coastal catchment zones where groundwater discharges directly into the coastal waters.

In general, infiltration excess overland flow may occur in this catchment area but only over short distances [56,57], implying a negligible surface runoff contribution to the total runoff from the catchment and thereby $\beta_{gw} \approx 1$ in the land surface grid cells, which cover about 85% of the catchment area. The remaining 15% is covered by surface water (streams, lakes and wetlands), for which $\beta_{gw} = 0$.

The fine data and model resolution for this catchment area allows us to investigate the role of the groundwater hydraulic gradient quantification, by using the same underlying fine-resolved (10 m $\times$ 10 m grid) ground slope data as in previously reported hydrological modeling [26,28] in two different ways. Specifically, we estimate the hydraulic gradient in each grid cell in the groundwater system as equal to: either (i) the arithmetic mean value of all the local, fine-resolved ground slopes in the subcatchment area of the outlet (to the nearest stream or directly to the sea) that is associated with the grid cell; this slope is then constant among the different grid cells within each subcatchment area and referred to as the subcatchment-average slope and hydraulic gradient; or (ii) the fine-resolved local ground slope at each grid cell location, which we refer to as the local ground slope and groundwater hydraulic gradient. Grid cell lengths through each model cell are generally for both gradient approaches calculated in the horizontal plane, based on estimated flow path directions and the size of model grid cells. Elevation is thus not accounted for in the transport distance calculations, which implies that any result differences between the different gradient estimation approaches depend on associated transport velocity and not transport length differences.

Field measurements of hydraulic conductivity (by 36 slug tests and 2 pumping tests) throughout the Forsmark catchment area yielded highly variable conductivity values, which were generally higher at the interface between the quaternary deposits and the underlying bedrock than in the soil above that interface [57]. For the investigation purposes of the previous hydrological modeling studies of this area [26,28], a uniform hydraulic conductivity value (equal to the reported mean value from measurements [57]) was used to mainly represent the solute transport through the high-conductivity layer at the soil–bedrock interface. The same model representation is used also here, allowing us to investigate the effect of different assumptions with regard to the spatial variability of hydraulic conductivity, by comparison with Norrström basin results under similar mean travel time conditions.

2.3.2 Norrström drainage basin

The Norrström drainage basin (Fig. 1c) is defined by the coastal outlet location of Norrström in the Swedish capital, Stockholm, and contains many (sixty shown in Fig. 1c) main subcatchments that drain their water through the major lake Mälaren to that common outlet and further into the Baltic Sea. The basin is rather flat with a basin-average topographic slope of 1.5% and a steepest topographic slope of 10% in a single 1 km × 1 km model grid cell, and low-lying with numerous lakes, underlain by granitic and gneiss-granitic bedrock covered by clay or till deposits. On the resolution scale of 1 km × 1 km, land-cover is classified to consist of 4% built-up areas, 36% agricultural and open land, 49% forest, 1.5% wetlands and 9.5% major inland surface waters.

Due to the coarse spatial model resolution of this basin, there are generally unresolved streams and other surface water features also within the model grid cells that are classified as land. Given a relevant stream density for these grid cells based on several paper and digital sources for river network delineation [19], the previous hydrological model studies of the Norrström basin [19–24] have quantified the total flow from the land–soil–groundwater system that feeds the surface water system to be, on average, about 75% of the total water flow through the basin. The remaining flow of about 25% goes then only through the surface water system. In these surface water cells defined by land cover and river network information, $\beta_{gw} = 0$. In the land–soil–groundwater system grid cells, $\beta_{gw} = 1$ because the pure surface runoff contribution to the total (surface and land–soil–ground) water flow is negligible (about 0.02%) in Norrström, as in Forsmark.

The previous, underlying hydrological modeling of the Norrström basin [21] conceptualized the groundwater flow to be partitioned between a relatively highly conductive (shallow, e.g., of quaternary deposits) and a less conductive (deeper, e.g., the bedrock) groundwater subsystem with the average total thickness of the two groundwater systems being set to 50 m following de Wit [58]. In this study, we investigate specifically the advective travel time effects of accounting for or neglecting the possible flow partitioning into the groundwater subsystem of slower/deeper flow.

Due to the coarse spatial resolution, the grid cell-average hydraulic gradient quantification for the groundwater system in the Norrström basin is more consistent with the subcatchment-average than with the local gradient estimate in the Forsmark catchment area. In contrast to the Forsmark application, the groundwater hydraulic conductivity in Norrström is modeled to vary between grid cells, depending on the available data of soil characteristics [19–24].

3 Results and discussion

Figure 2 illustrates the spatial and statistical distributions of advective travel times through the different water subsystems, and in total through the catchment area to the coast, from all the 10 m × 10 m grid cells in Forsmark, with the different hydraulic gradient quantifications: (i) the subcatchment-average gradient (Fig. 2a, b), and (ii) the local gradient (Fig. 2c, d). The different gradient quantifications yield large travel time differences in both the spatial distribution (Fig. 2a, c) and the spreading of the statistical distribution (Fig. 2b, d) of travel times. The differences depend on the contributions of very long travel time components in the local gradient approach from the large flat-topography parts of the Forsmark area. The arithmetic averaging involved in the subcatchment-average gradient approach reduces the weight of small local gradient values and provides more realistic estimates of the prevailing hydraulic gradient, which is not likely to fluctuate as much as the local ground slope.

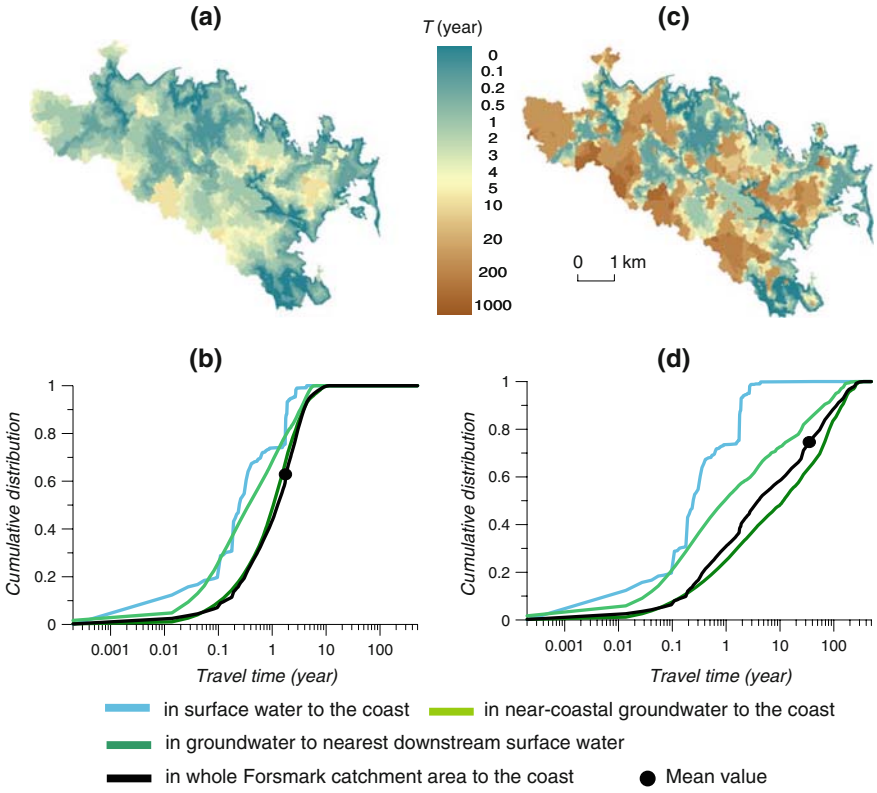


Fig. 2 Spatial (**a**, **c**) and statistical (**b**, **d**) distributions of advective travel time from all grid cells in Forsmark catchment area to the coast, for the stream network and groundwater subsystems and the whole catchment, with hydraulic gradient quantification from **a**, **b** subcatchment-average ground slope; and **c**, **d** local ground slope

The differences in Fig. 2 underline the essential role of the model representation of groundwater hydraulics, here reflected by the different possible approaches to estimate the hydraulic gradient. The fact that infiltration excess overland flow is negligible in Forsmark [56,57] explains the strong hydraulic gradient control on calculated advective travel times through this catchment area, which was also found by McGuire et al. [15]. A contrasting and counter-intuitive positive relationship between catchment transit times and ground slope has been found by Tetzlaff et al. [59] for the flat Swedish Krycklån boreal basin. This result is explained by the total runoff being dominated by relatively fast overland flow, rather than by groundwater flow as in Forsmark, in the flatter, poorly drained peat soils of the Krycklån basin [60].

Figure 3 illustrates the spatial and statistical distributions of advective travel times through the different water subsystems and the whole basin to the coast from all the 1 km × 1 km grid cells in Norrström. Results are illustrated for the alternative model representations that neglect (Fig. 3a, b) or account for (Fig. 3c, d) the possible contribution of slow/deep groundwater flow. The travel time differences obtained by these alternative model representations are large in terms of both the spatial distribution (Fig. 3a, c) and the statistical spreading (Fig. 3b, d) of travel times in the basin. Since flow path directions and flow pathway lengths are the same in

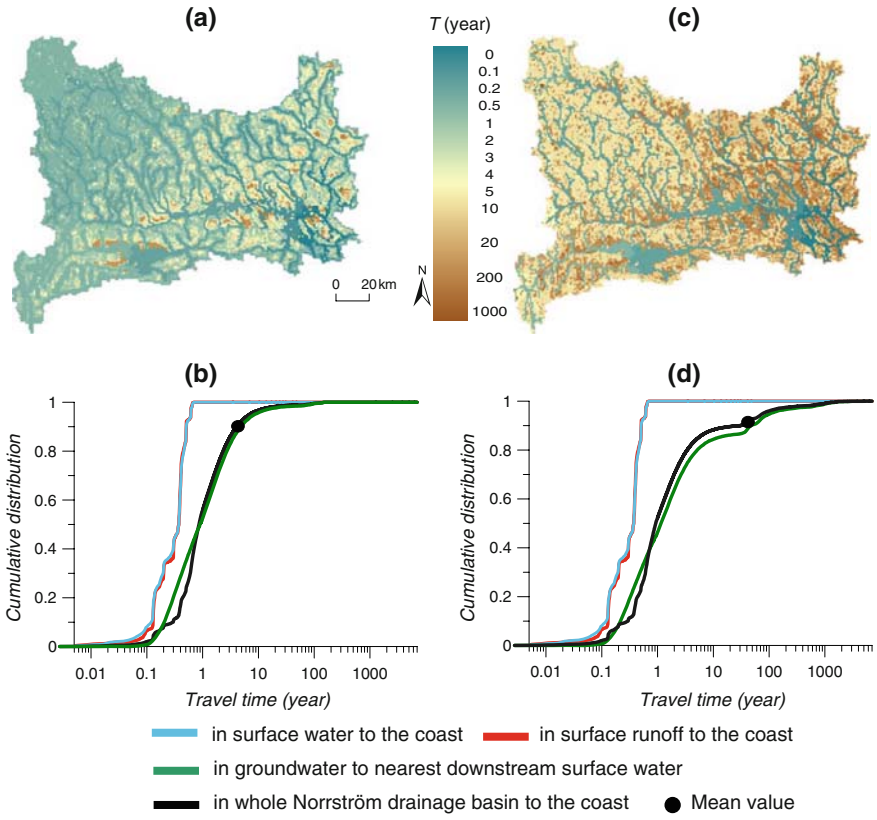


Fig. 3 Spatial (**a**, **c**) and statistical (**b**, **d**) distributions of advective travel time from all grid cells in the Norrström drainage basin to the coast, for the surface runoff, surface water and groundwater subsystems and the whole catchment for **a**, **b** neglect and **c**, **d** account for the possible contribution of a slow/deep groundwater subsystem

both conceptualizations, these travel time distribution differences are only due to the assumed partitioning (in Fig. 3c, d) or the no-partitioning (in Fig. 3a, b) of groundwater flow between the two groundwater subsystems with distinctly different advective velocities. Also these differences emphasize thus the importance of relevant groundwater system characterization for relevant and accurate assessment of solute travel time distributions in catchments.

The importance of groundwater controls on catchment-scale travel times has also been reported in other studies, which have found greater travel time dependence on bedrock seepage [61,62] than on the more directly intuitive catchment size. The results illustrated in Fig. 3 show that also in the Norrström basin, the relatively small proportion of about 12% of the total runoff recharging the slow/deep groundwater subsystem is sufficient for significantly increasing the total mean travel time to the coast. With this deep recharge fraction, the mean total travel time increases from about 3 to about 30 years (Fig. 3). Furthermore, the travel time variability, quantified in terms of the travel time standard deviation, increases from about 10 to about 60 years (Fig. 3).

Table 1 summarizes the most directly comparable catchment-scale travel time statistics for the two catchment areas: those for the case of neglecting the slow/deep groundwater

Table 1 Catchment-scale mean value, standard deviation and coefficient of variation of advective travel times T from all grid cells in the Forsmark and Norrström catchment areas to the coast

Advective travel time to the coast in the Forsmark catchment area		Advective travel time to the coast in the Norrström drainage basin	
Mean value (years)	1.7	Mean value (years)	3.4
Standard deviation (years)	1.7	Standard deviation (years)	11.2
Coefficient of variation	1.0	Coefficient of variation	3.3

Travel times in Forsmark are for the subcatchment-average hydraulic gradient quantification. Travel times in Norrström neglect the possible contribution of slow/deep groundwater

flow contribution in Norrström (Fig. 3a, b), and the case of subcatchment-average hydraulic gradient in Forsmark (Fig. 2a, b). Under these conditions, the resulting total mean travel time is similar for both catchment areas. To explain this similarity, Tables 2 and 3 in Appendix summarize some characteristic flow and transport parameter statistics for these Forsmark and Norrström cases, respectively. A comparison between these tables shows that, beyond the similar precipitation surplus because both catchments are in the same hydro-climatic region, the similarity in mean advective travel times between the two cases depends on their similar mean combined times for advective groundwater transport, expressed as the mean value of the ratio between groundwater flow path length and groundwater flow velocity (in turn quantified as the product of hydraulic conductivity and slope divided by porosity). This time scale is similar even though the separate groundwater flow and transport characteristics are quite different between the two cases, and irrespectively of the very different catchment area sizes. The independence of mean travel time on catchment scale is consistent with similar findings for diffuse solute transport by McGuire et al. [15], Tetzlaff et al. [59] and Rodgers et al. [63], however depending on different types of flow and transport controls in the different catchment case studies.

The main groundwater system control of the Forsmark and Norrström travel time results is emphasized by the order-of-magnitude smaller standard deviation and the three times smaller coefficient of variation of advective travel times (from different input positions to the coast) in Forsmark than in Norrström (Table 1). Specifically, comparison between Tables 2 and 3 with regard to the coefficients of variation of different flow and transport parameters shows that the uniform hydraulic conductivity and porosity assumption for the groundwater system in Forsmark is primarily responsible for its small travel time variability (in terms of both standard deviation and coefficient of variation). This variability difference implies a much greater spatio-temporal spreading (macro-dispersion) of solute around its centre of mass in Norrström than in Forsmark, and emphasizes the importance of spatial groundwater variability assumptions for the distributions of advective solute travel time and the associated physical spreading of solute mass in catchment-scale hydrological transport.

Figure 4 finally illustrates the effect of these variability differences for the delivered solute mass fraction from different input locations to the coast in the comparable (in terms of similar mean travel time) Norrström (Fig. 4a–c) and Forsmark (Fig. 4d–f) cases, for different scenarios of the product of catchment-average physical travel time $\check{T}$ and biogeochemical attenuation rate λ. For each catchment area and $\lambda\check{T}$ scenario, Fig. 4 shows also the total resulting catchment-scale delivery fraction $\overline{\alpha}$ of solute mass to the coast.

The results in Fig. 4 indicate that the differences in solute travel time variability implied by the different spatial variability assumptions for the groundwater hydraulic parameters in the two catchment case quantifications are primarily important in solute-catchment situations where $0.1 < \lambda\check{T} < 10$. For the interval $0.1 < \lambda\check{T} < 10$, any uniform, instantaneous solute mass

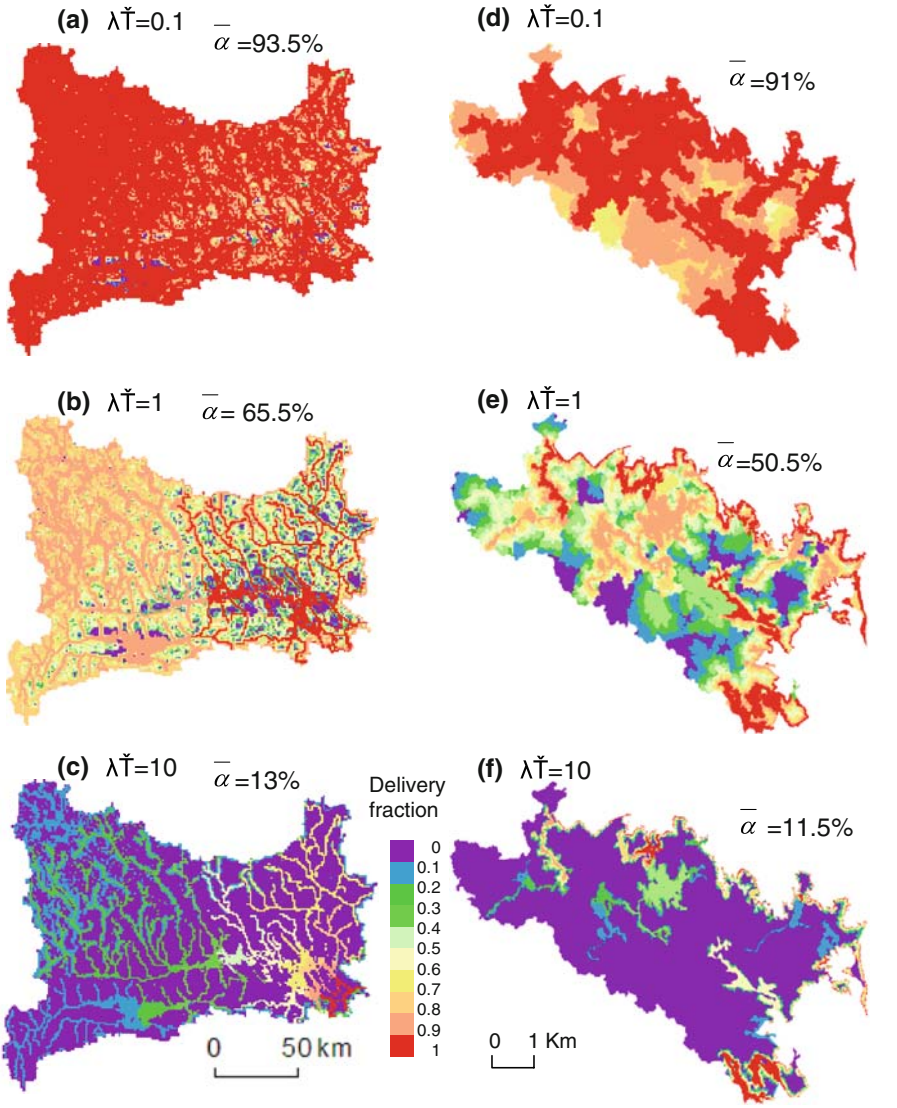


Fig. 4 Map of delivered mass fraction from each grid cell location to the coast in the Norrström drainage basin (**a**–**c**) and Forsmark catchment area (**d**–**f**), for different scenarios (0.1, 1 and 10) of the product of catchment-average advective travel time $\check{T}$ and attenuation rate λ. The total delivered mass fraction $\bar{\alpha}$ from the whole catchment area is also quantified in the figure for each $\lambda\check{T}$ scenario. Travel times in Norrström (**a**–**c**) neglect the possible contribution of slow/deep groundwater. Travel times in Forsmark (**d**–**f**) are for the subcatchment-average hydraulic gradient

input leads to a delivered mass fraction that is about 30% (or 15 percentage units) greater for the Norrström case with the larger advective travel time variability than for the Forsmark case with the smaller travel time variability. Smaller or greater mean $\lambda\check{T}$ scenarios than this $0.1 < \lambda\check{T} < 10$ interval imply nearly non-attenuated or nearly totally attenuated solute mass, respectively, essentially regardless of the prevailing variability of advective solute travel times.

4 Conclusion

This study has outlined the possible quantification of advective solute travel-time distributions in different catchment areas. The specific catchment cases in the study differ largely in terms of their scale, data-model resolutions, and process representations in the travel time modeling. Yet the comparative analysis of these cases has provided some important general insights.

The results show that the groundwater system characterization and model representation largely controls the resulting distributions of advective travel times through these hydrological catchments. For groundwater assumptions that yielded similar catchment-average travel times in the different catchment cases, the spatial variability in groundwater hydraulics played an essential role for the travel time variance, which determines the physical spreading (macrodispersion) of non-reactive solute mass transported through the catchment.

For solute that is physically or biogeochemically attenuated along its different transport pathways through the catchment, the product of the catchment-average advective travel time and the solute-dependent biogeochemical attenuation rate was shown to largely determine the effects of travel time variance on the total solute mass delivery from the catchments. These effects were found to be primarily important for the product interval $0.1 < \lambda\check{T} < 10$. For hazardous contaminants, where even very small solute concentrations and concentration differences may be essential for environmental and health risks, however, the travel time variability effects may be important and need to be further investigated also for $\lambda\check{T} \geq 10$.

Furthermore, the primary importance-interval $0.1 < \lambda\check{T} < 10$ applies to the investigated conditions of variability only in the physical, advective travel time T. The interval may widen significantly if also the attenuation rate λ is variable, and depending on its possible cross-correlation with the advective travel time T [e.g., 34,64–66]. Further investigations and realistic quantifications are needed for the spatial variability of biogeochemical attenuation rates and their correlation with the physics of flow and transport in both the surface water and not least the groundwater systems of hydrological catchments.

Acknowledgements Financial support for this work has been provided by the Swedish Research Council (VR), the Swedish Rescue Services Agency (Räddningsverket), and the Swedish Nuclear Fuel and Waste Management Company (SKB).

Appendix

Modeling of the Forsmark catchment area

The previous hydrological modeling of the Forsmark catchment area [25–28], described in detail by Jarsjö et al. [25,26,28], provides the spatial distribution of total (surface and subsurface) annual average runoff (over 30 years), as estimated from the modeled precipitation surplus, which is the difference between annual average precipitation and modeled actual evapotranspiration in each model cell. The direction of the water flow and solute transport pathway through each cell of the modeled catchment area is estimated from the local ground slope, which is in turn estimated from a detailed digital elevation model of the area, as explained in more detail by Jarsjö et al. [25,26,28]. The present Forsmark application is based on average results from two different empirical approaches [67,68], which were used for actual evapotranspiration calculations in the previous, underlying hydrological modeling and yielded consistent resulting spatial water flow distribution with each other and consistent water flow results with independent runoff data from the catchment area.

To obtain the travel time from each input cell a_{gw} along the associated groundwater pathway (as estimated from the ground slope direction) to the control plane at distance x_{cp} (of the nearest stream and/or the coast), the travel time contribution $\Delta\tau_{gw} = \Delta x_{gw}/v_{gw}$ of each model cell is estimated from the cell length Δx_{gw} and the transport velocity $v_{gw} = K \cdot I/n$, where K is hydraulic conductivity, I is hydraulic gradient, and n is effective porosity. The total mean $\tau_{gw}(a_{gw}, x_{CP})$ is the sum of $\Delta\tau_{gw}$ for all cells along the transport pathway to x_{cp}. The hydraulic conductivity and effective porosity are assumed to be $1.5 \cdot 10^{-5}$ m/s and 0.05, respectively, over the whole catchment area, as reported by Johansson et al. [57] for the quaternary deposits/bedrock interface. In the subcatchment-average gradient approach (i) in the main text, the hydraulic gradient in each grid cell is set equal to the arithmetic average of all the local ground slopes in the subcatchment area of the associated grid-cell outlet to the nearest stream or directly to the sea. The hydraulic gradient is then constant among the different grid-cells within each subcatchment area, including in cells with nearly zero local ground slope. In the local gradient approach (ii) in main text, the local hydraulic gradient in each cell equals the local ground slope in that cell.

The stream network includes all the interconnected bodies of surface water, streams, lakes and wetlands, through which the waterborne solute mass may be transported all the way to the coast. For obtaining the travel time from each input cell a_s along the associated stream network pathway to the outlet at x_{out}, the travel time contribution $\Delta\tau_s = L_s/v_s$ of each stream stretch is estimated from its length Ls and mean flow velocity $v_s = Q/A_{cs}$, where Q is the mean annual flow rate and A_{cs} is the mean cross-sectional area of the stream. The total mean $\tau_s(a_s, x_{out})$ is the sum of $\Delta\tau_s$ for all the stream stretches, lakes and wetlands along the whole stream network pathway to x_{out}; the estimation of $\Delta\tau_s$ in lakes and wetlands is explained below. In streams where the mean cross-sectional area is measured and known (from 0.29 to 0.43 m^2), Q is assumed to be equal to the modeled mean annual water flow at the model cell location where the stream cross-section area was measured. Otherwise, a generic value of $A_{cs} = 0.3\,m^2$ and the modeled [25,26,28] mean annual flow value at the mouth of the stream are used in the calculation of mean flow velocity. The $\Delta\tau_s$ contribution of a lake or a wetland is estimated as $\Delta\tau_s = A_{L/W} \cdot d_{eff}/Q$, where $A_{L/W}$ is the area of the lake or wetland, Q is the mean annual flow rate through the lake or wetland, and d_{eff} is the mean depth in lakes and is defined as the product of the depth and the water content (typically around 0.9) in wetlands. The mean flow rate Q is the modeled [25,26,28] mean annual runoff (precipitation minus evaporation) generated in the lake or wetland plus the modeled [25,26,28] runoff into the lake or wetland from all upstream cells. For isolated lakes and wetlands that are not part of any connected stream network pathway all the way to the coast, their travel time contribution is calculated from the length of a topographically estimated transport pathway through the lake or wetland divided by an average velocity $v_{L/W} = Q\sqrt{[4A_{L/W}/\pi]}/[A_{L/W} \cdot d_{eff}]$, where $\sqrt{[4A_{L/W}/\pi]}$ is the diameter of a circle with the same area $A_{L/W}$ as the lake or wetland. The travel time contribution obtained is added to the total groundwater travel time τ_{gw} along the main groundwater pathway that crosses the isolated lake or wetland. Details on measured lake and wetland depths are given by Johansson [69] and Brunberg et al. [70].

Modeling of the Norrström drainage basin

The previous hydrological modeling of the Norrström drainage basin [19–24], described in detail by Darracq et al. [21] after de Wit [58,71] and Greffe [72], provides the spatial distribution of total (surface and subsurface) annual average runoff (over 30 years),

as estimated from the modeled precipitation surplus in each model cell; as in Forsmark, the precipitation surplus is also here defined as the difference between annual average precipitation and actual evapotranspiration, modeled based on an empirical function of precipitation and potential evapotranspiration [67,73]. A digital elevation map available at a resolution of 1 km × 1 km was used for assigning water flow and solute-transport pathway directions through each cell of the modeled catchment area. Previous hydrological modeling results in the Norrström basin [21] are also used to obtain: the contribution of the flow through the land–soil–groundwater system to the total water flow through the basin, based on the ratio between the long-term average groundwater recharge and the total precipitation surplus in each grid cell, as functions of land cover and topographic slope [68]; and the travel time $\tau_{gw}(a_{gw}, x_{CP})$ from each grid cell a_{gw} along the groundwater pathway (as estimated from the ground slope direction) to the control plane at distance x_{cp} of the nearest stream.

As mentioned also in the main text, the previous hydrological modeling of the Norrström basin [21] conceptualized the groundwater flow to be partitioned between a relatively highly conductive (shallow, e.g., of quaternary deposits) and a less conductive (deeper, e.g., the bedrock) groundwater subsystem, with the groundwater flow depending on aquifer type, soil texture, groundwater level, slope, land use and average January temperature [73] based on empirical estimates by de Wit [58] and Mourad [74]. Following Wendland [68], the travel time $\tau_{sgw}(a_{gw}, x_{CP})$ from each input cell a_{gw} along the associated pathway (as estimated from the ground slope direction) in the fast/shallow groundwater subsystem to the nearest stream at x_{cp} is quantified as: $\tau_{sgw} = lp/v$, where v is the groundwater flow velocity and lp is the average length of the groundwater flow path, as a function of conductivity of the aquifer ca, hydraulic gradient h(with topographic slope in each 1000 m grid cell used as an estimate of h), primary effective aquifer porosity pp and modeled total runoff Q. Specifically, $v = ca \cdot h/pp$ and $lp = 0.5/ns$, where ns is stream density quantified as $ns = 2$ in wetlands and $(Q/450)^{0.8}$ (with Q given in mm year^{-1}) elsewhere.

Furthermore, following Meinardi et al. [73], the travel time $\tau_{dgw}(a_{gw}, x_{CP})$ from each grid cell a_{gw} along the pathway in the slow/deep groundwater subsystem to the nearest stream at x_{cp} (the transport length of which is estimated similarly as in the fast/shallow groundwater system, from the ground slope direction) is calculated as the product between total effective porosity of the aquifer tp, thickness of groundwater flow formation at and the inverse of the long-term average recharge of the slow/deep groundwater subsystem. The average total thickness of both groundwater systems over the whole Norrström basin is set to 50 m following de Wit [58]. Values for aquifer conductivity, primary and total porosity are empirically related [58,68,73] to the aquifer type and the soil and bedrock groundwater flow yields, which were obtained for the Norrström basin from the Swedish Geological Survey mapping of groundwater in soil and bedrock.

The resulting total groundwater travel time $\tau_{gw}(a_{gw}, x_{CP})$ is quantified as: $\tau_{gw} = (1 - \beta_{dgw}) \cdot \tau_{sgw} + \beta_{dgw} \cdot \tau_{dgw}$, where β_{dgw} is the recharge of the slow/deep groundwater subsystem, in terms of flow fraction of the total groundwater flow into the surface water system; that fraction was on average about 12% of the total flow in the previous model simulations [19–24], implying that $\beta_{dgw} = 0.12$ in the present results that account for the possible slow/deep groundwater flow contribution, and $\beta_{dgw} = 0$ in the results that neglect it (Tables 2, 3).

The stream network includes all the interconnected bodies of surface water, streams and lakes through which the waterborne mass may be transported all the way to the coast. For obtaining the travel time from each cell a_s along a stream network pathway to the outlet at x_{out}, the travel time contribution $\Delta\tau_s = L_s/v_s$ of each stream stretch is estimated from

Table 2 Mean value, standard deviation and coefficient of variation of precipitation surplus (i.e., precipitation minus actual evapotranspiration), topographic slope, groundwater system porosity, hydraulic conductivity and flow path length, and the combined characteristic time for groundwater transport expressed by the fraction: $\frac{\text{groundwater\_flow\_path\_length}}{\frac{\text{conductivity}^{*}\text{slope}}{\text{porosity}}}$ in the Forsmark catchment area

	Mean value	Standard deviation	Coefficient of variation
Precipitation surplus (mm/year)	226	35	0.2
Slope	0.03	0.01	0.3
Porosity	0.05	0	0
Hydraulic conductivity (m/day)	1.3	0	0
Groundwater flow path length (m)	394	491	1.2
$\frac{\text{groundwater\_flow\_path\_length}}{\frac{\text{conductivity}^{*}\text{slope}}{\text{porosity}}}$ (years)	1.5	1.8	1.2

Table 3 Mean value, standard deviation and coefficient of variation of precipitation surplus (i.e., precipitation minus actual evapotranspiration), topographic slope, groundwater system porosity, hydraulic conductivity and flow path length, and the combined characteristic time for groundwater transport expressed by the fraction: $\frac{\text{groundwater\_flow\_path\_length}}{\frac{\text{conductivity}^{*}\text{slope}}{\text{porosity}}}$ in the Norrström drainage basin

	Mean value	Standard deviation	Coefficient of variation
Precipitation surplus (mm/year)	233	50	0.2
Slope	0.01	0.01	1.0
Porosity	0.18	0.3	1.6
Hydraulic conductivity (m/day)	125	295	2.4
Groundwater flow path length (m)	1042	299	0.3
$\frac{\text{groundwater\_flow\_path\_length}}{\frac{\text{conductivity}^{*}\text{slope}}{\text{porosity}}}$ (years)	3.4	12	3.6

its length L_s and mean flow velocity v_s, empirically estimated from an expression given in [20,75], as: $v_s = 0.36Q^{0.241}$ in streams and $v_s = 0.36(Q/A_L)^{0.241}$ in lakes, where Q is mean annual flow rate in m^3/s as obtained from previous hydrological modeling [21] and A_L is lake surface area. The total τ_s (a_s, x_{out}) is the sum of $\Delta\tau_s$ for all the stream stretches and lakes along the whole stream network pathway and topographically estimated transport pathway through lakes to x_{out}.

The travel time contributions $\Delta\tau_{sr=}L_{sr}/v_{sr}$ in the surface runoff subsystem is estimated in analogy with the stream network subsystem, from the surface runoff pathway length L_{sr} and mean flow velocity $v_s = 0.36(Q)^{0.241}$, where Q is modeled surface runoff flow in m^3/s. In the combined total travel time distribution through the whole basin, the weight of the surface runoff contribution is negligible (0.02%) compared to the groundwater flow, so that results are directly comparable between Norrström and Forsmark.

References

1. McGuire KJ, McDonnell JJ (2006) A review and evaluation of catchment transit time modeling. J Hydrol 330:543–563
2. Maloszewski P, Zuber A (1982) Determining the turnover time of groundwater systems with the aid of environmental tracers. 1. Models and their applicability. J Hydrol 57:207–231
3. Simic E, Destouni G (1999) Water and solute residence times in a catchment: stochastic model interpretation of 18O transport. Water Resour Res 35(7):2109–2120
4. Lindgren GA, Destouni G (2004) Nitrogen loss rates in streams: scale-dependence and up-scaling methodology. Geophys Res Lett. doi:10.1029/2004GL019996
5. Lindgren GA, Destouni G, Miller AV (2004) Solute transport through the integrated groundwater–stream system of a catchment. Water Resour Res. doi:10.1029/2003WR002765
6. Botter G, Bertuzzo E, Bellin A, Rinaldo A (2005) On the Lagrangian formulations of reactive solute transport in the hydrologic response. Water Resour Res. doi:10.1029/2004WR003544
7. Fiori A, Russo D (2008) Travel time distribution in a hillslope: insight from numerical simulations. Water Resour Res. doi:10.1029/2008WR007135
8. Schnoor JL (1996) Environmental modeling: fate and transport of pollutants in water, air and soil. Wiley, New York
9. Rinaldo A, Marani A, Rigon R (1991) Geomorphological dispersion. Water Resour Res 27(4):513–525
10. White AB, Kumar P, Saco PM, Rhoads BL, Yen BC (2004) Hydrodynamic and geomorphologic dispersion: scale effects in the Illinois River Basin. J Hydrol 288:237–257
11. Saco PM, Kumar P (2002) Kinematic dispersion in stream networks 1. Coupling hydraulic and network geometry. Water Resour Res. doi:10.1029/2001WR000695
12. Valett HM, Morrice JA, Dahm CN, Campana ME (1996) Parent lithology, surface-groundwater exchange, and nitrate retention in headwater streams. Limnol Oceanogr 41(2):333–345
13. Ensign SH, Doyle MW (2005) In-channel transient storage and associated nutrient retention: evidence from experimental manipulations. Limnol Oceanogr 50(6):1740–1751
14. Haggerty R, Wondzell SM, Johnson MA (2002) Power-law residence time distribution in the hyporheic zone of a 2nd-order mountain stream. Geophys Res Lett. doi:10.1029/2002GL014743
15. McGuire KJ, McDonnell JJ, Weiler M, Kendall C, McGlynn BL, Welker JM, Seibert J (2005) The role of topography on catchment-scale water residence time. Water Resour Res 41(5):W05002.1–W05002.14
16. Boano F, Packman AI, Cortis A, Revelli R, Ridolfi L (2007) A continuous time random walk approach to the stream transport of solutes. Water Resour Res. doi:10.1029/2007WR006062
17. Wörman A, Packman AI, Marklund L, Harvey J, Stone S (2007) Fractal topography and subsurface water flows from fluvial bedforms to the continental shield. Geophys Res Lett. doi:10.1029/2007GL029426
18. Malmström ME, Destouni G, Banwart SA, Strömberg BHE (2000) Resolving the scale-dependence of mineral weathering rates. Environ Sci Technol 34:1375–1378
19. Darracq A, Destouni G (2005) In-stream nitrogen attenuation: model-aggregation effects and implications for coastal nitrogen impacts. Environ Sci Technol 39(10):3716–3722
20. Darracq A, Destouni G (2007) Physical versus biogeochemical interpretations of nitrogen and phosphorus attenuation in streams and its dependence on stream characteristics. Glob Biogeochem Cycles. doi:10.1029/2006GB002901
21. Darracq A, Greffe F, Hannerz F, Destouni G, Cvetkovic V (2005) Nutrient transport scenarios in a changing Stockholm and Mälaren valley region. Water Sci Technol 51(3-4):31–38
22. Destouni G, Darracq A (2006) Response to comment on "In-stream nitrogen attenuation: model aggregation effects and implications for coastal nitrogen impacts". Environ Sci Technol 40(7):2487–2488
23. Lindgren GA, Destouni G, Darracq A (2007) The inland subsurface water system role for coastal nitrogen load dynamics and abatement responses. Environ Sci Technol 41(7):2159–2164
24. Darracq A, Lindgren GA, Destouni G (2008) Long-term development of phosphorus and nitrogen loads through the subsurface and surface water systems of drainage basins, Global Biogeochem Cycles. doi:10.1029/2007GB003022
25. Jarsjö J, Shibuo Y, Destouni G (2004) Using the PCRaster-POLFLOW approach to GISbased modelling of coupled groundwater–surface water hydrology in the Forsmark Area. Swedish Nuclear Fuel and Waste Management Company Report R-04-54, Stockholm, Sweden
26. Jarsjö J, Destouni G, Persson K, Prieto C (2007) Solute transport in coupled inland-coastal water systems. General conceptualization and application to Forsmark. Swedish Nuclear Fuel and Waste Management Company Report R-07-65, Stockholm, Sweden
27. Destouni G, Shibuo Y, Jarsjö J (2008) Freshwater flows to the sea: spatial variability, statistics and scale dependence along coastlines. Geophys Res Lett. doi:10.1029/2008GL035064

28. Jarsjö J, Shibuo Y, Destouni G (2008) Spatial distribution of unmonitored inland water discharges to the sea. J Hydrol 348(12):59–72
29. Dagan G (1989) Flow and transport in porous formations. Springer Verlag, Berlin
30. Rubin Y (2003) Applied stochastic hydrogeology. Oxford University Press, New York
31. Simmons CS (1982) A stochastic-convective transport representation of dispersion in one-dimensional porous media. Water Resour Res 18:1193–1214
32. Shapiro AM, Cvetkovic V (1998) Stochastic analysis of solute travel time in heterogeneous porous media. Water Resour Res 24:1711–1718
33. Cvetkovic V, Shapiro AM (1990) Mass arrival of sorptive solute in heterogeneous porous media. Water Resour Res 26:2057–2067
34. Destouni G, Cvetkovic V (1991) Field-scale mass arrival of sorptive solute into the groundwater. Water Resour Res 27:1315–1325
35. Destouni G (1993) Stochastic modeling of solute flux in the unsaturated zone at the field scale. J Hydrol 143:45–61
36. Cvetkovic V, Dagan G (1994) Transport of kinetically sorbing solute by steady random velocity in heterogeneous porous formations. J Fluid Mech 265:189–215
37. Destouni G, Sassner M, Jensen KH (1994) Chloride migration in heterogeneous soil: 2, stochastic modeling. Water Resour Res 30:747–758 (Correction, Water Resour Res 31:1161, 1995)
38. Ginn TR, Simmons CS, Wood BD (1995) Stochastic-convective transport with nonlinear reaction: biodegradation with microbial growth. Water Resour Res 31:2689–2700
39. Simmons CS, Ginn TR, Wood BD (1995) Stochastic-convective transport with nonlinear reaction: mathematical framework. Water Resour Res 31:2675–2688
40. Berglund S, Cvetkovic V (1996) Contaminant displacement in aquifers: coupled effects of flow heterogeneity and nonlinear sorption. Water Resour Res 32:23–32
41. Cvetkovic V, Dagan G (1996) Reactive transport and immiscible flow in geochemical media: 2 applications. Proc R Soc Lond Ser A 452:303–328
42. Eriksson N, Destouni G (1997) Combined effects of dissolution kinetics, secondary mineral precipitation, and preferential flow on copper leaching from mining waste rock. Water Resour Res 33: 471–483
43. Yabusaki SB, Steefel CI, Wood BD (1998) Multidimensional, multicomponent, subsurface reactive transport in nonuniform velocity fields: code verification using an advective reactive streamtube approach. J Contam Hydrol 30(3):299–331
44. Foussereau X, Graham W, Aakpoji A, Destouni G, Rao PSC (2001) Solute transport through a heterogeneous coupled vadose-saturated zone system with temporally random rainfall. Water Resour Res 37(6):1577–1588
45. Tompson AFB, Bruton CJ, Pawloski GA, Smith DK, Bourcier WL, Shumaker DE, Kersting AB, Carle SF, Maxwell RM (2002) On the evaluation of groundwater contamination from underground nuclear tests. Environ Geol 42:235–247
46. Cvetkovic V, Haggerty R (2002) Transport with multiple-rate exchange in disordered media. Phys Rev E. doi:10.1103/PhysRevE.65.051308
47. Malmström ME, Destouni G, Martinet P (2004) Modeling expected solute concentration in randomly heterogeneous flow systems with multicomponent reactions. Environ Sci Technol 38:2673–2679
48. Rinaldo A, Marani A (1987) Basin scale model of solute transport. Water Resour Res 23:2107–2118
49. Destouni G, Graham W (1995) Solute transport through an integrated heterogeneous soil–groundwater system. Water Resour Res 31:1935–1944
50. Botter G, Bertuzzo E, Bellin A, Rinaldo A (2005) On the Lagrangian formulations of reactive solute transport in the hydrologic response. Water Resour Res. doi:10.1029/2004WR003544
51. Dagan G, Fiori A (1997) The influence of pore-scale dispersion on concentration statistical moments in transport through heterogeneous aquifers. Water Resour Res 33(7):1595–1606
52. Fiori A, Dagan G (2000) Concentration fluctuations in aquifer transport: a rigorous first-order solution and applications. J Contam Hydrol 45:139–163
53. Fiori A, Berglund S, Cvetkovic V, Dagan G (2002) A first-order analysis of solute flux statistics in aquifers: the combined effect of pore-scale dispersion, sampling, and linear sorption kinetics. Water Resour Res. doi:10.1029/2001WR000678
54. Janssen GMCM, Cirpka OA, Van der Zee EATM (2006) Stochastic analysis of nonlinear biodegradation in regimes controlled by both chromatographic and dispersive mixing. Water Resour Res. doi:10.1029/2005WR004042
55. Malmström ME, Berglund S, Jarsjö J (2008) Combined effects of spatially variable flow and mineralogy on the attenuation of acid mine drainage in groundwater. Appl Geochem 23(6):1419–1436

56. Lindborg T (2005) Description of surface systems. Preliminary site description Forsmark area—version 12. Swedish Nuclear Fuel and Waste Management Company Report R-05-03, Stockholm, Sweden
57. Johansson P-O, Werner K, Bosson E, Juston J (2005) Description of climate, surface hydrology, and near-surface hydrology. Preliminary site description. Forsmark area—version 1.2. Swedish Nuclear Waste Management Company (SKB) Report R-05-06, Stockholm, Sweden
58. de Wit MJM (1999) Nutrients fluxes in the Rhine and Elbe basins. Dissertation, Royal Dutch Geographical Society, Utrecht, Netherlands
59. Tetzlaff D, Seibert J, McGuire KJ, Laudon H, Burns DA, Dunn SM, Soulsby C (2009) How does landscape structure influence catchment transit time across different geomorphic provinces? Hydrol Process 23:945–953
60. Laudon H, Sjöblom V, Buffam I, Seibert J, Mörth CM (2007) The role of catchment scale and landscape characteristics for runoff generation of boreal streams. J Hydrol 344:198–209
61. Asano Y, Uchida T, Ohte N (2002) Residence times and flow paths of water in steep unchannelled catchments, Tanakami, Japan. J Hydrol 261:173–192
62. Dunn SM, McDonnell JJ, Vaché KB (2007) Factors influencing the residence time of catchment waters: a virtual experiment research. Water Resour Res. doi:10.1029/2006WR005393
63. Rodgers P, Soulsby C, Waldron S (2005) Stable isotope tracers as diagnostic tools in upscaling flow path understanding and residence time estimates in a mountainous mesoscale catchment. Hydrol Process 19:2291–2307
64. Jarsjö J, Bayer-Raich M, Ptak T (2005) Monitoring groundwater contamination and delineating source zones at industrial sites: uncertainty analyses using integral pumping tests. J Contam Hydrol 79:107–134
65. Jarsjö J, Bayer-Raich M (2008) Estimating plume degradation rates in aquifers: effect of propagating measurement and methodological errors. Water Resour Res. doi:10.1029/2006WR005568
66. Cunningham JA, Fadel ZJ (2007) Contaminant degradation in physically and chemically heterogeneous aquifers. J Contam Hydrol 94:293–304
67. Turc L (1954) The water balance of soils. Relation between precipitation, evaporation and flow. Annal Agron 5:491–569
68. Wendland F (1992) Die Nitratbelastung in den Grundwasserlandschaften der 'alten' Bundesländer (BRD), Berichte aus der Okologischen Forschung, Band 8, Forschungszentrum Jülich, Jülich
69. Johansson P-O (2003) Drilling and sampling in soil. Installation of groundwater monitoring wells and surface water level gauges. Swedish Nuclear Fuel and Waste Management Company Report P-03-64, Stockholm, Sweden
70. Brunberg A-K, Carlsson T, Blomqvist P, Brydsten L, Strömgren M (2004) Identification of catchments, lake-related drainage parameters and lake habitats. Swedish Nuclear Fuel and Waste Management Company Report P-04-25, Stockholm, Sweden
71. de Wit MJM (2001) Nutrients fluxes at the river basin scale I: the PolFlow model. Hydrol Process 15: 743–759
72. Greffe F (2003) Material transport in the Norrström drainage basin: integrating GIS and hydrological process modelling. Master Thesis, Royal Institute of Technology, Stockholm, Sweden
73. Meinardi C, Beusen A, Bollen M, Klepper O (1994) Vulnerability to diffuse pollution of European soils and groundwater. National Institute of Public Health and Environmental Protection (RIVM), Report 4615001002, Bilthoven, Netherlands
74. Mourad DSJ (2002) Application of GIS-based modelling to assess nutrient loads in rivers of the Estonian part of the lake Peipsi basin. MANTRA-East working paper 5.1
75. Alexander RB, Elliott AH, Shankar U, McBride GB (2002) Estimating the sources and transport of nutrients in the Waikato River Basin, New Zealand. Water Resour Res 38:1268–1290

Environ Fluid Mech (2010) 10:121–135
DOI 10.1007/s10652-009-9148-1

ORIGINAL ARTICLE

Estimating travel time of recharge water through a deep vadose zone using a transfer function model

Samuel Mattern · Marnik Vanclooster

Received: 30 March 2009 / Accepted: 13 August 2009 / Published online: 28 August 2009

Abstract We estimate the travel time of percolating water through a deep vadose zone at the regional scale using a transfer function model and a physical based conceptual flow model (Hydrus-1D), thereby exploiting the time series of precipitation, actual evapotranspiration and groundwater piezometry and generic vadose zone data. With the transfer function model we observe a high variability of estimated travel time varying from 0.9 to 3.1 years, corresponding to estimated vertical water flux velocities varying from 6.6 to 28.0 m/year. These results were compared with the travel time estimated from the physical based conceptual model. With the flow model, estimated travel time varies between 4.7 and 15.5 years, corresponding to water flux velocities varying between 1.7 and 4.1 m/year. The estimated travel time calculated with the flow model were therefore about five times larger than those estimated with the transfer function model. This could be explained by the fact that the transfer function model considers heterogeneous recharge from the vadose zone as well as from the vicinity of the piezometer through the so called "pushing effect". In addition, the flow model requires various hydrogeological and hydrodynamic parameters which were estimated using generic parametrisation approaches, that are largely affected by uncertainty and may not reflect the local conditions. In contrast, the transfer function model only exploits available measurable time series and has the advantage of being site-specific.

Keywords Transfer function model · Travel time · Soil · Groundwater level fluctuations · Brusselian sands

1 Introduction

The vadose zone is a key component within terrestrial systems and plays an essential role in the hydrological cycle. The vadose zone encompasses the unsaturated soil root zone and controls thereby the fluxes of water, matter and energy between the atmosphere, land surface

S. Mattern (✉) · M. Vanclooster
Université catholique de Louvain, Croix du Sud, 2, bte 2, 1348 Louvain-la-Neuve, Belgium
e-mail: samuel.mattern@uclouvain.be

and subsurface water bodies. Knowing the travel time of water in the vadose zone is therefore a prerequisite for quantifying the recharge and load of groundwater bodies in terms of variable boundary conditions at the land surface interface [12].

Characterizing the travel time of water and associated matter in the vadose zone is complicated and this because of many reasons [2]. First, a wide set of physical, chemical and biological processes affect flow in the vadose zone. Second, the size and geometry of the vadose zone in natural conditions is often not well known. Third, the physical, chemical and biological properties of the vadose zone are variable in space and in time which makes the experimental assessment tedious. Fourth, as compared to technologies and tools available for characterizing surface hydrological and hydrogeological properties, much less technologies have been designed to characterize the partially saturated vadose zone. Fifth, water flow in the vadose zone is a highly non-linear process and depends very much on the saturation degree of water which on its turn is controlled by the variable water input at the land surface. Given this complexity, assessing travel time of water through the vadose zone has received far less attention as compared to travel time assessments in surface and groundwater bodies. The vadose zone is therefore still often a large gray box in our current hydrological knowledge.

Travel times of water through the vadose zone depend very much on the thickness of the vadose zone. In flat alluvial regions, where only a thin vadose zone is present, travel time may be small such that pressures exerted at the land surface will propagate fast to the groundwater body. In undulating landscapes, however, where the size of the vadose zone will generally be much larger, higher travel time is expected. In arid environments, for instance, the vadose zone can be hundreds of meters thick, and infiltration fluxes very low [10], resulting in the residence time of water ranging from several hundreds to thousands of years [5]. However, also in such a deep vadose zone, flow can be extremely fast. For instance Levitt et al. [29] showed that a surface applied tracer could reach a 78 m deep groundwater tables in 8–14 days, proving the small travel time in such large formations.

Different methods exist to assess vadose zone travel times. Analysing at a given depth in the vadose zone the breakthrough of a land surface applied tracer allows determining the travel time from the moments of the time dependent breakthrough curve. Use have been made of single ionic inert tracers such as chloride [1,9,20,34,35], stable and radiogenic isotopes [3,15], nitrate [37], or a combination of different ionic tracers [41]. Yet, given the complexity to carry out large scale tracer breakthrough experiments and the long time frames involved, an alternative approach relies on the inversion of conceptual flow and transport models. With such approaches use is made of some measured vadose zone property which is modelled using a transport model in which the travel time is considered, directly or indirectly, as the fitting model parameter. With the inverse approach, measured moisture content or resident tracer profiles have been used as fitting data by Wang et al. [49], McElroy and Hubbell [32], Hubbell et al. [17], Flint and Ellett [13] and Wu et al. [50]. Robinson et al. [39] combines measurements of moisture content with the analysis of a tracer experiment. Constantz et al. [7], proposes the logging of temperature profiles in combination with a coupled flow and heat transport model.

When using the inverse approach, physically based conceptual flow and transport models can be used that are based on the physical laws of mass, momentum and energy conservation and flow. The use of such flow and transport models is often prohibitive given the lack of data that are available for parametrising such models. Where nothing but only few time series of vadose zone properties are available, time series models are often preferred to the conceptual flow and transport models [27]. One of these, the transfer function model, was proposed by Jury [25] to characterize field scale solute transport through an unsaturated soil. Using a transfer function model, it is considered that the internal transport mechanisms are

unknown. Jury [25] derived a simple theory for estimating the average and extreme behaviour of solutes moving through the vadose zone as a function of a simple field-measurable travel time probability density function (pdf). Inverting the transfer function for the parameters of the travel time pdf allows estimating the travel time through the vadose zone. This approach was applied for example by Vanderborght et al. [46] and Javaux and Vanclooster [20] to determine the travel time through large undisturbed soil columns from resident concentration data; by Gasser et al. [14] to predict nitrate leaching under potato crops of an agricultural field; by Stewart and Loague [43] to predict pesticide load of ground water at the regional scale; by van der Velde et al. [47] to estimate the travel time through a volcanic soil on a small atoll in the Pacific Ocean.

Notwithstanding the availability of these methods, uncertainties in the estimation of the travel time through the vadose zone remain very high [11,16]. For instance, Cook et al. [8] compared ionic tracer studies with radiogenic isotope tracer studies and concluded on the inconsistency between the estimates of the travel time with both methods. Given these uncertainties and inconsistencies, quite some scope exist to improve our knowledge on the travel time of water through the vadose zone. Upon our knowledge, while the Jury transfer function has been used to simulate solute transfer through the vadose zone, no study has been presented so far which exploits the easily measurable temporal dynamics of the vadose zone lower boundary condition, i.e. the groundwater position, as fitting data in the inversion process.

The objective of this study is to evaluate the travel time of percolating water through a deep unsaturated zone using transfer function theory, thereby exploiting the measured time series of precipitation, actual evapotranspiration and groundwater piezometry. The underlying hypothesis is that the water flux wave at the soil surface will propagate through the vadose zone and, when it hits the ground water body, result in a ground water depth perturbation. Modeling this traveling wave by means of the transfer function model allows inferring the large scale travel time through the vadose zone. The results obtained by means of the transfer function model will further be compared with the travel time estimated by means of a physical based conceptual model (Hydrus-1D).

2 Material and methods

2.1 Site and data description

This study focuses on the vadose zone of the Brusselian aquifer situated in the center of Belgium. The aquifer has a surface area of 965 km^2 and encompasses an unconfined ground water body which is of primary importance for drinking water supply in the region. The aquifer is composed of Tertiary sedimentary sands and is overlain by a Quaternary loess layer of variable thickness (0–15 m). The Brusselian sands outcrop mainly in the valleys where sandy and sandy loam soils develop. Transmissivity of the aquifer varies from 2.9×10^{-5} to 1.2×10^{-2} m^2/s and its permeability varies from 1.4×10^{-6} to 6×10^{-3} m/s [19].

Daily precipitation measured at Uccle (Belgium) from 1961 to 2007 was provided by the European Climate Assessment & Dataset [44], and was aggregated to monthly totals. A time series of monthly actual evapotranspiration was calculated by Oger [36], using daily weather data collected at Gembloux (situated in the study region) from 1950 to 1989. In order to extrapolate these data to the period for which precipitation series were available (1961–2007), we calculated the mean actual evapotranspiration for each month and completed the time series of monthly evapotranspiration rates with the calculated mean per month. Completing the

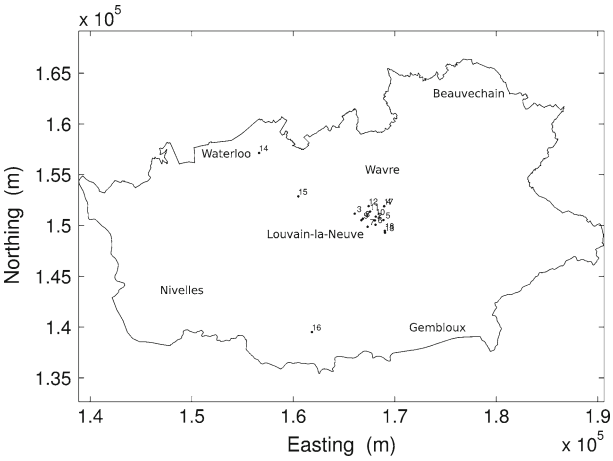


Fig. 1 Location of the piezometers in the Brusselian sands ground water body

missing values as such is justified by the high reproducibility of the actual evapotranspiration cycle from 1 year to another. The time series of the piezometric height were provided by the regional administration [33] and from the Hydrological Service of the Université catholique de Louvain. In the study area, ground water levels are measured on 241 piezometers. However, the vast majority of these piezometers cannot be used in the context of this study since either the length of the time series is insufficient, or the frequency with which the measurements are performed does not allow to capture the dynamics of the system. From the 241 available piezometers, only a subset of 18 piezometers were selected for which water levels are measured for a long period (1967–2007) and at a high measurement frequency (Fig. 1).

2.2 The transfer function model

We use a transfer function model to describe water transport through the vadose zone. The underlying hypothesis is that a water flux wave at the soil surface will propagate through the heterogeneous vadose zone and perturbate the water level when it hits the groundwater body. The vadose zone in this case study includes the biologically active soil root zone, a loamy substratum, and the unsaturated sandy substratum. The water flux wave at the soil surface is determined by infiltration of rain water in the soil, from which evapotranspiration rates are subtracted. The model further assumes that no lateral run-off occurs and that no dispersion of percolating water takes place other than that which is implicitly represented by the travel time variations of the system [25]. Water transport is assumed to be convective with a stochastic component expressing the variability of water travel time, due to heterogeneous pathways in the soil and subsoil [14]. We consider only the vertical flux since the vertical travel time in the vadose zone is generally 1–3 orders of magnitude smaller than the horizontal one [26].

The transfer function model uses a simple travel time probability density function (pdf) to relate an output time series to an input time series. The approach of Jury [25] derives a distribution function based on the distribution of the travel time of solutes from the soil surface to a reference depth. Since the vadose zone between the surface and the aquifer smooth out the variations of water flux, a transfer function model can be used to describe the relation between the variation in rainfall and evapotranspiration and the variations of the level of

the groundwater table. We use the transfer function model to describe the convolution of the variation in the rainfall minus the actual evapotranspiration with the transfer function to predict variation in the level of the ground water table:

$$P(t) - \mu_P = \alpha \int_0^t f(t)(I(t-\tau) - \mu_I)dt \tag{1}$$

where $I(t) = R(t) - E(t)$ is the input time series (percolating water, mm/month); $R(t)$ is the rainfall (mm/month); $E(t)$ is the actual evapotranspiration (mm/month); μ_I is the mean of I (mm/month); $I(t) - \mu_I$ is the variation of input around μ_I (mm/month); $f(t)$ is the transfer function equal to the travel time probability density function (pdf) (months^{-1}); t is time (months); τ is the time lag (months); α is the scaling factor (−); $P(t)$ is the piezometric level (mm/month); μ_P is the mean of P (mm/month); $P(t) - \mu_P$ is the variation of P around μ_P (mm/month).

The integral in Eq. 1 expresses the convolution between the variations around the mean percolating water ($I(t) - \mu_I$) with the travel time probability density function to predict the variations around the mean of the piezometric level ($P(t) - \mu_P$). Here t is time and τ the lag time between inputs and piezometric variations, α is a scaling factor, $f(t; \mu, \sigma)$ is the travel time probability density function. A log-normal probability density function was chosen for the transfer function $f(t; \mu, \sigma)$ to reflect the transport and mixing of the percolating water as it moves through a porous medium exhibiting a log-normal distribution of pore-water velocities [38]. Hence:

$$f(t; \mu, \sigma) = \frac{1}{\sigma t\sqrt{2\pi}} \exp\left(-\frac{(\ln(t) - \mu)^2}{2\sigma^2}\right). \tag{2}$$

The model $f(t; \mu, \sigma)$ corresponds to a log-normal probability density function where μ and σ respectively correspond to the mean and standard deviation of the logarithm of the travel time through the vadose zone.

The transfer function was used to determine the modal transfer time, defined as the time where the probability density function reaches its maximum value, and the 5th and 95th percentiles, defined as the time needed by 5 and 95% of the percolating water to reach groundwater table. The parameters α, μ, and σ were obtained by minimizing the sum of squared errors for all observations within each piezometer [47]:

$$\Phi(\alpha, \mu, \sigma) = \frac{1}{n} \sum_{i=1}^{n} \left[(P - \mu_P)_{i,\text{simulated}} - (P - \mu_P)_{i,\text{observed}}\right]^2 \tag{3}$$

where $\Phi(\alpha, \mu, \sigma)$ is the objective function to be minimized; $(P - \mu_P)_{i,\text{simulated}}$ is the simulated variation of P around μ_P (mm); $(P - \mu_P)_{i,\text{observed}}$ is the observed variation of P around μ_P (mm); n is the number of observations in each piezometer.

To avoid ill-posedness, an a-priori parameter domain was divided into a regular grid, covering a range from 0 to 5.5 for μ (Eq. 2) (corresponding to mean travel time from 1 month to 20 years), from 0 to 1 for σ (Eq. 2) and from 0 to 150 for the scaling factor α (Eq. 1). The objective function (Eq. 3) was calculated for each combination of the grid selected parameters. Afterward, the parameters minimizing the objective function were used as an input of the Nelder–Mead multidimensional unconstrained nonlinear minimization algorithm available in Matlab™ for a more accurate estimation of the optimal parameters.

We identified first the travel time pdf for each piezometer separately. Subsequently we identified an effective travel time by inverting all data simultaneously and compared these to

the mean travel time pdf obtained from the local estimates. In case the traveling wave process through the vadose zone is a linear process, than the local and effective travel time should converge.

2.3 The physical based conceptual flow model

We compared the estimated travel time by means of transfer function theory with predictions made by means of a physically based flow and transport model. We used the Hydrus-1D model [42] which solves numerically the 1-D governing flow model (Richards equation) for the vadose zone:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K(h)\left(\frac{\partial h}{\partial z} + 1\right)\right] - S \tag{4}$$

where z is the vertical coordinate (positive upwards), t is time (day), S is a sink term (day^{-1}), h is the pressure head (m), θ is the volumetric moisture content and K is the unsaturated hydraulic conductivity function (m/day); and the convection dispersion solute transport model:

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial z^2} - \nu\frac{\partial C}{\partial z} \tag{5}$$

where C is the solute concentration ($\mathrm{m}^3\,\mathrm{m}^{-3}$), D is the dispersion coefficient ($\mathrm{m}^2\,\mathrm{day}^{-1}$) and ν is the water flux ($\mathrm{m\,day}^{-1}$). Hydrus-1D solves Eqs. 4 and 5 for vertical heterogeneous porous media and transient boundary conditions by means of the Galerkin finite element method. The soil hydraulic functions used in Hydrus-1D are modelled according to the van Genuchten–Mualem model [48].

A variable flux condition was imposed on the soil surface, considering the same inputs $I(t)$ than those used in the transfer function model (Eq. 1). At the bottom of the flow domain a constant pressure head equal to zero was imposed, representing the position of the water table. The initial conditions were defined through a warming up procedure. In a first loop, the model was run from 1961 to 2007. The obtained moisture and pressure head profile at the end of this warming up loop was subsequently used as initial moisture and pressure head profile for the tracer simulation loop. In this second loop, a tracer pulse was applied on the top of the soil column during one time unit and followed through the profile allowing to infer convective solute travel time.

In contrast to the transfer function approach, we need information about the lithology of the sub-soil (i.e. a sandy layer overlain by a sandy loam layer) and the physical properties of each sub-soil layer to parametrise the formation profile in Hydrus-1D. The thickness of the different layers could be inferred from lithological logs that were collected when installing the piezometers. However, these logs were only available for 12 out of the 18 piezometers. No data were available to parametrise the soil physical properties, and therefore use was made of three generic approaches to estimate these parameters (Table 1). The first set comes from the Hydrus catalog [6], the second one from Rosetta [40] and the third one from literature of measured hydraulic properties of media in the study area [18,22,23]. Three typical different parameter sets describing the soil properties were used since we didn't know the real soil properties at the location of the piezometers.

Except for the soil parameters, the generic parameters were used (e.g. single porosity, no hysteresis), because of the absence of data needed for a better parametrisation of the physically-base model.

Table 1 Soil properties used in the Hydrus-1D simulations

Source	Soil layer	Depth (m)	θ_r (−)	θ_s (−)	α (m^{-1})	n (−)	λ (−)	K_{sat} (m/year)	Dispersivity (m^{-1})
Hydrus catalog	Sandy loam	0–4	0.065	0.41	7.5	1.89	0.5	387.265	0.1
	Sand	4–14.3	0.045	0.43	14.5	2.68	0.5	2601.72	1
Rosetta	Sandy loam	0–4	0.0387	0.387	2.67	1.4484	0.5	139.612	0.1
	Sand	4–14.3	0.053	0.3747	3.53	3.1798	0.5	2346.88	1
Literature	Sandy loam	0–4	0	0.3894	0.4084	1.204	0.5	108.6	0.1
	Sand	4–14.3	0.0818	0.33	2.87	5.64	−0.555	189.8	1

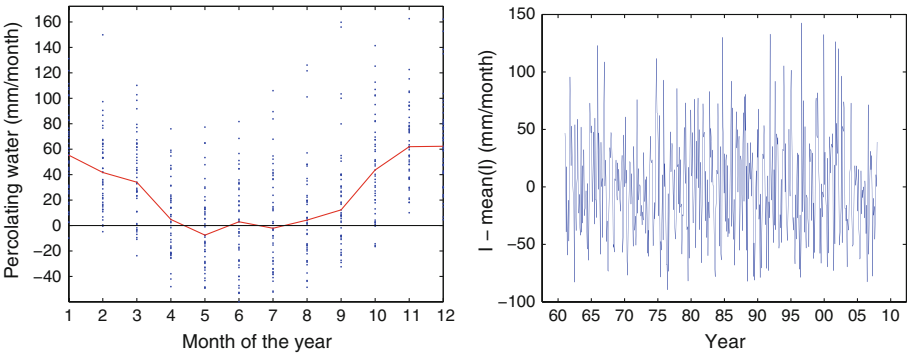


Fig. 2 **a** Monthly total (*dots*) and monthly median of the monthly total (*line*) percolating water flux at the soil–atmosphere interface (mm/month), calculated from 1965 to 2007 and **b** variation of the monthly total percolation water flux around its mean value (mm/month) used as input time series of the transfer function model

3 Results

3.1 Input of the transfer function model

The monthly median of the monthly total percolating water flux at the soil–atmosphere interface, $I(t)$, calculated from 1965 to 2007, varies from nearly null from April to August, to 62 mm in December (Fig. 2a). The variability of the monthly total percolating water flux for each month is large. The input time series of the transfer function model was obtained by subtracting the average of the percolating water flux calculated for the complete time series from the monthly total percolating water flux time series $I(t)$ (Fig. 2b).

3.2 Travel time estimated from the transfer function model

Figure 3 presents, for two piezometres randomly sampled in the study area, the results of simulated versus measured variation of piezometry around the mean (Fig. 3a, b) together with its corresponding optimal log-normal transfer functions (Fig. 3c, d). We observe that the transfer function model allows to describe the phase of the dynamic signal propagating through the deep vadose zone, but has more difficulty to predict correctly its amplitude.

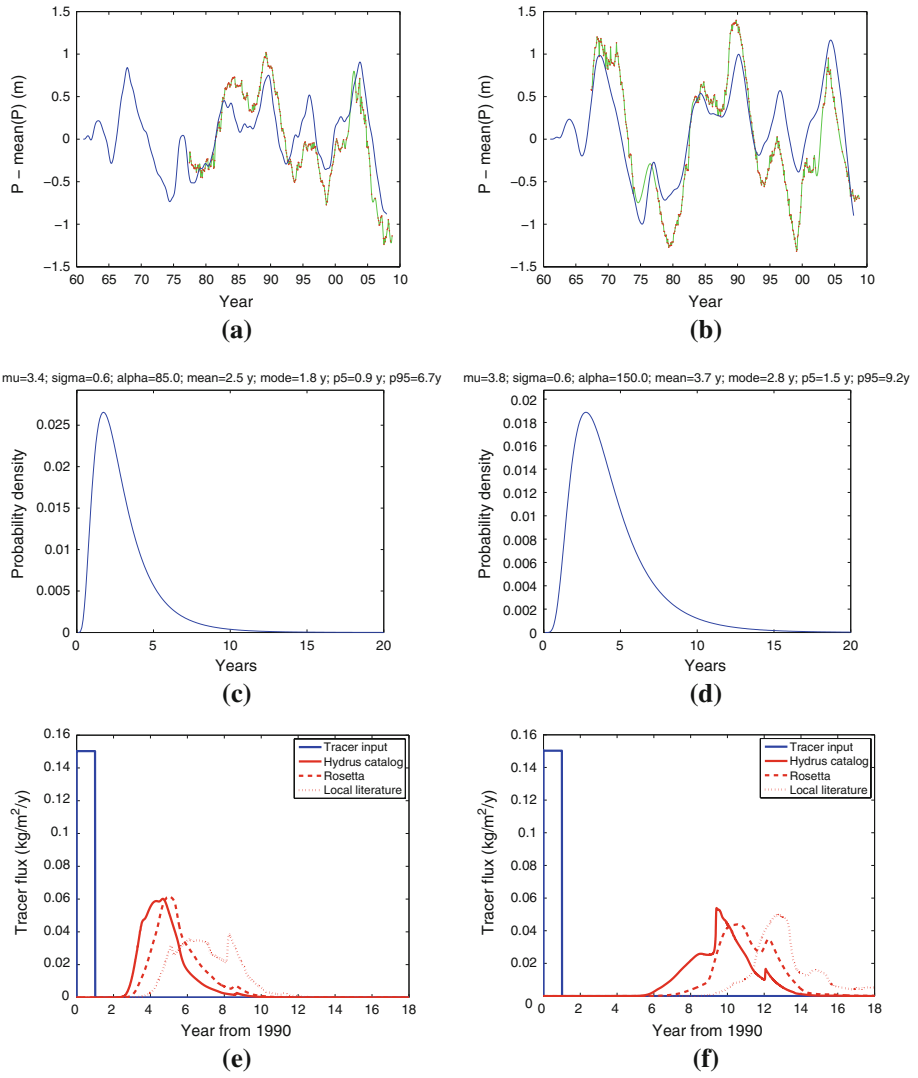


Fig. 3 Measured (*red dots*) and simulated with the transfer function model (*line*) piezometric time series (**a, b**), their optimal log-normal transfer functions (**c, d**) and the breakthrough curves calculate with Hydrus-1D (**e, f**) for two different piezometers

We also observe a systematic overestimation of the simulation between 1993 and 1998 which can be explained by the exploitation regime of ground water in this area, and which is not considered in the transfer function model. However, only the phase is important for estimating the travel time.

The modal value of the travel time of recharge water through the unsaturated zone, calculated from the transfer function model for 18 individual piezometers varies from 0.9 to 3.1 years (Table 2). The average of the modal transfer time of these 18 piezometers was 2.3 years, with a standard deviation of 0.6 years. Furthermore, the fifth percentile of the travel time for the 18 piezometers varies from 0.5 to 1.8 years and the 95th percentile from 5.2 to

Table 2 Depth to the water table, travel time and water flux velocities calculated with the transfer function model and Hydrus-1D (with the three sets of soil properties) for the 18 piezometers

Piezometer	Depth	Transfer function		Hydrus (Hydrus set)		Hydrus (Rosetta set)		Hydrus (Local literature set)	
		Travel time	Velocity	Travel time	Velocity	Travel time	Velocity	Travel time	Velocity
	(m)	(year)	(m/year)	(year)	(m/year)	(year)	(m/year)	(year)	(m/year)
1	37.7	2.92	12.9	12.4	3.04	13.4	2.82	15.5	2.43
2	36.7	2.25	16.3	10.4	3.53	12.5	2.94	13.3	2.76
3	41.4	2.92	14.2	10.5	3.94	10.0	4.14	12.0	3.45
4	14.3	1.75	8.2	4.7	3.04	5.0	2.86	8.3	1.72
5	33.8	2.25	15.0	10.4	3.25	12.6	2.69	13.4	2.53
6	26.9	2.17	12.4	9.2	2.92	9.1	2.96	9.3	2.89
7	39.5	1.92	20.6	10.1	3.91	9.9	3.99	11.7	3.37
8	16.9	2.58	6.6	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
9	17.7	2.25	7.8	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
10	36.5	2.33	15.7	10.2	3.58	10.7	3.41	12.6	2.9
11	33.7	2.92	11.6	9.5	3.55	9.6	3.51	11.6	2.91
12	49.1	3.08	15.9	12.6	3.9	12.9	3.81	13.5	3.64
13	29.7	2.75	10.8	9.4	3.16	10.6	2.8	12.8	2.32
14	25.6	0.92	28.0	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
15	39.3	2.25	17.4	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
16	22.9	1.33	17.2	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
17	14.5	2.17	6.7	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
18	29.8	2.50	11.9	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

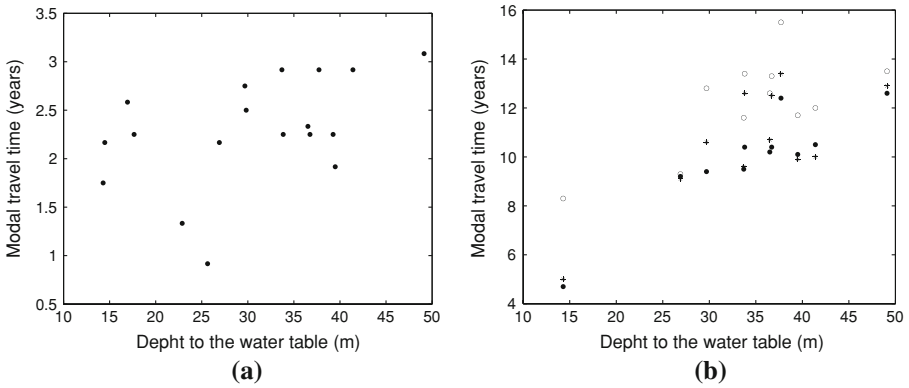


Fig. 4 Relation between the depth to the water table and the modal travel time calculated with **a** the transfer function model and **b** Hydrus-1D (*black circles*, *stars* and *white circles* correspond to travel time calculated with the soil parameters of the Hydrus catalog, Rosetta and local literature, respectively)

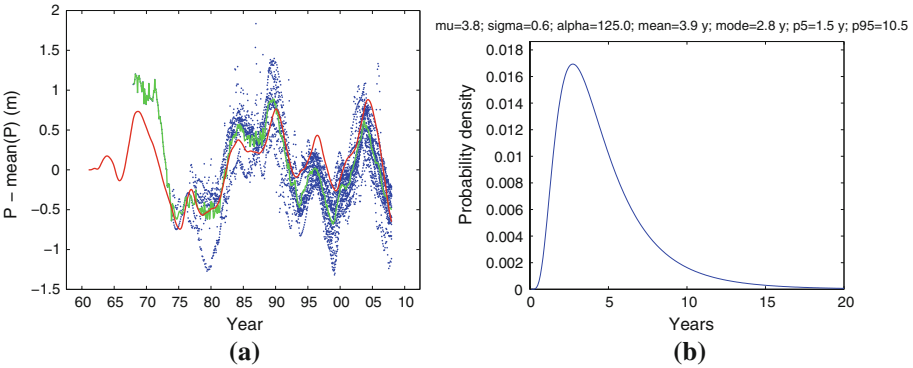


Fig. 5 Measured data of all the piezometers (*blue dots*), their mean (*green line*) and the simulated piezometry (*red line*) (**a**) and corresponding optimal log-normal transfer function (**b**)

15.1 years, which shows that, for a same location, the water wave propagates very quickly through the unsaturated zone while at other places wave propagations is much slower. The variability of the travel time between the piezometers can be explained by the differences in the thickness of the unsaturated zone, variability of hydrogeological properties, and flow phenomena that are not considered within the transfer function model. For instance, we observe that travel time increases with depth to the water table (Fig. 4a). By dividing the depth to the water table (Table 2) by the modal travel time (Table 2) for each piezometer, we can estimate that the water travel velocity through the vadose zone varies from 6.6 to 28.0 m/year (Table 2).

We also attempted to derive a regional transfer function model by fitting one log-normal transfer function to the mean of all the data of the 18 piezometers (Fig. 5a). The resulting optimal log-normal pdf is characterized by a modal travel time of 2.8 years, which is close to the mean of the modal travel time of the 18 individual piezometers (2.3 years) (Fig. 6). The 5th percentile of the travel time estimated on the mean of all piezometers was 1.5 year and the 95th was 10.5 years (Fig. 5b).

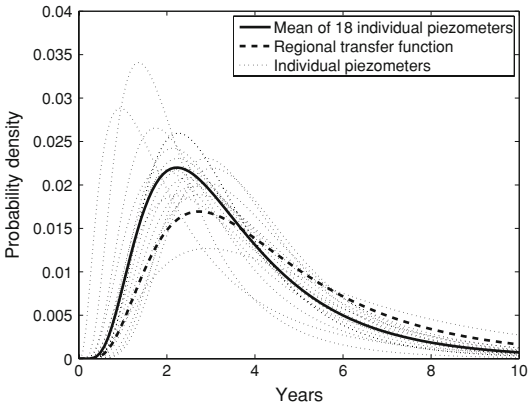


Fig. 6 Transfer function models of the 18 individual piezometers (*dotted lines*), their mean (*line*) and regional transfer function model (*broken line*)

3.3 Travel time estimated from the conceptual flow and transport model

Travel time was also calculated from the breakthrough of a conservative tracer, simulated for 12 piezometers using the Hydrus-1D model. All 18 piezometers of the previous section could not be used since information about the thickness of the sandy and sandy loam layer was not available for 6 out of them. The breakthrough curve of the inert tracer at the bottom of the unsaturated zone column had a log-normal-like shape with a modal transfer time going from 4.7, 5.0 and 8.3 years to 12.6, 13.4 and 15.5 years for the first, second and third set of soil properties, respectively (Table 2). Figure 3e, f presents the simulated breakthrough curves for two selected piezometers. Obviously, we observe also that travel time increases with depth to the water table (Fig. 4b).

4 Discussion

The values of the travel time estimated with the physical based conceptual model are about five times larger than those estimated with the transfer function model. Differences can be explained by conceptual differences between both modeling approaches and the uncertainty associated with the parametrisation of the physically based conceptual flow and transport model.

Indeed, the travel time estimated from the simulated breakthrough curve of the inert tracer by means of the physical based conceptual model, considers 1D convective mass transport of an inert tracer with the moving water body through an unsaturated column on top of the piezometer. In this case, it has been hypothesized that the full water body within the vadose zone contributes to the conservative mass transport, and that the estimated travel time of the recharging water corresponds to the simulated travel time of the conservative tracer. In reality however, the measured piezometric position will not only be influenced by the local vertical recharge through the 1D vadose zone column on top of the piezometer, but also by the lateral recharge through the vadose zone and the groundwater body itself.

The transfer function model considers the geologic formation between the observation point (i.e. the local piezometer) and the surface input as a black box with no well defined geometry. With this modeling approach the flow domain does not necessarily pertain to a 1D unsaturated soil column on top of the observation point, but may be equivalent to a geometrically irregular flow domain, encompassing stream tubes through the vadose zone and

the groundwater body itself, where water is transported to the observation point in different directions. Hence, the transfer function model may conceptually encompass the recharge of the water coming from the vadose zone on top of the observation point as well as water that laterally feeds the observation point. The transfer function model will also consider the water that feeds the observation point through the so-called "pushing effect". Indeed, when pressure waves propagate through the vadose zone, resident water in the vicinity of the piezometer may be mobilized and contribute to piezometric variation. Pressure waves generally propagates through natural porous media at velocities which are much faster than mass transport waves [4], and piezometric variation does not necessarily reflect mass transport through the vadose zone. In case the lateral recharge is more important than the vertical recharge through the vadose zone, or in case the "pushing effect" is significant, then travel time inferred from the transfer function modeling approach may easily be smaller than estimated from a physically based 1D conceptual mass transport model through the vadose zone.

Further, vertical flow rate in the vadose zone may be faster than estimated from the conceptual mass transport model. Many experimental evidence show that only part of the available water in the vadose zone contributes to effective mass transport. Using small scale sandy columns collected from the same region as in the present case study, Vanclooster et al. [45] showed that mass transport could much better be described with a two porosity flow model rather than a single porosity flow model, considering mobile water fractions ranging between 0.2 and 0.8 of total resident volumetric soil moisture. In case the mobile water fraction would indeed be 0.2, then effective velocities would be five times larger which is consistent with the estimates from the transfer function model in the present study. Based on detailed solute transport experiments at variable flow rates on an undisturbed 0.5 m^3 size sandy soil monolith, collected from the sandy substratum of the Brusselian aquifer, Javaux and Vanclooster [21] showed that effective pore water velocity contributing to the mass transport of the inert tracer ranges between 0.8 and 1.6 times the theoretical pore water velocity, suggesting mobile water fractions between 0.6 and 1.2 times the total volumetric moisture content. Also in the majority of the experiments of this last study, larger effective velocities than pore velocities estimated from a single region mass transfer model were obtained.

The unavailability of pore water to contribute to effective mass transport will further be exacerbated when heterogeneous flow develops in the flow domain feeding the piezometer. In-situ heterogeneous flow may first be induced by the local variability of the vadose zone transport parameters. The local scale variability of the vadose zone flow properties is indeed very large. Mallants et al. [30], for instance, showed that the local scale saturated hydraulic conductivity of the loamy soil specific for the present study area could vary with six orders of magnitude, while saturated porosity could vary between 0.38 and 0.51. Together with the local scale variability of the unsaturated hydraulic properties, these induce variability of fluxes of a factor 12. Similar results were obtained when analyzing the local scale variability of the sandy substratum of the present case study area [23]. Heterogeneous flow may further be promoted at the formation scale by the presence of distinguished structural features, such as clay lenses and discontinuous clay layers. Indeed, the sandy substratum of the study area is intersected with different discontinuous clay layers which reflect the genesis of these natural formations. Flow across different textured layers will be unstable and induce heterogeneous flow. Within the unsaturated sandy soil core analyzed by Javaux et al. [24], a distinguished structural feature was present which induced clearly heterogeneous flow at the macroscopic core scale. Similarly, when analyzing inert tracer transport through a 6 m deep observation well below an artificial lake, different clay layers were observed. The measured local flow velocity across the different layers varied between 2.5 and 13.1 m/year which is in the same order of magnitude of flow velocities that are calculated in the present study.

The difference in estimated travel time with both methods could also partially be explained by the uncertainty on the hydrogeological and hydrodynamic data feeding the 1D conceptual flow model. The 1D flow model requires a good knowledge of the lithology of the vadose zone and the unsaturated flow hydraulic properties. In case the hydraulic properties are not know, they can be estimated based on the knowledge of the granulometry and bulk density using so-called pedo-transfer functions. For the considered piezometers in this case study, the lithology of the vadose zone could rather correctly be reconstructed using the lithological logs that were collected when installing the piezometers. However, the hydraulic properties are far more uncertain. To consider the uncertainty in vadose zone hydraulic properties, three parametrisation strategies were considered in the present study. The different parametrisation schemes yield estimated travel time variations ranging between 4.7 and 15.5 years, corresponding to velocity variations from 1.7 to 4.1 m/year. Although largely significant, this suggests that uncertainty in hydraulic parametrisation only marginally explains the difference in estimated travel between both methods.

As a final remark, the important travel time through the vadose zone may have considerable consequences for water management in the study area. Indeed, the unconfined groundwater body of the Brusselian aquifer is exploited for drinking water purposes and is subjected to many pressures. Mattern et al. [31] for instance describe the current pollution of the groundwater body by nitrates, while Leterme et al. [28] discuss pollution by pesticides. Within the context of the implementation of the European Water Framework Directive and the European Nitrate Directive, the regional administration currently implements a set of land use management measures, envisaging to reduce emissions of pollutants from the land surface and the root active zone to the groundwater body. The important travel time through the vadose zone suggests that the impact of these measures on ground water quality can not be expected at the short term.

5 Conclusions

In this study, we used a transfer function model to estimate the travel time of percolating water through a deep vadose zone (Brusselian sands, Belgium) and we showed a high variability across the study area, with modal travel time varying from 0.9 to 3.1 years and transfer velocities varying from 6.6 to 28.0 m/year. These water flux velocities were compared with estimates made by means of a physically based conceptual flow model (Hydrus-1D). The transfer velocity calculated with the transfer function model was about five times higher than those simulated with the flow model. This could be explained by the fact that the flow model requires various hydrogeological inputs which are affected by considerable uncertainty and that it doesn't take the local soil characteristics and heterogeneities into account. In contrast, the transfer function only needs easily measurable time series and has the advantage to be site-specific. However, in case the "pushing effect", which intrinsically is modeled with the transfer function model, would be important, then the travel time estimated from the transfer function would not reflect real mass transport from the surface to the groundwater body. The importance of the "pushing effect" could not be quantified with the present set-up. It is therefore suggested to complete this study with more advanced mass tracer experiments.

Acknowledgements S. Mattern is a Research Fellow supported by the "Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture" (F.R.I.A. Belgium). The authors are grateful to the "Direction Générale des Ressources Naturelles et de l'Environnement" (Ministère de la Région Wallonne) for delivering the "10-Sous" database and to Gérard Peeters (Université catholique de Louvain) for delivering hydrogeological and piezometric data.

References

1. Allison G, Stone W, Hughes M (1985) Recharge in karst and dune elements of a semi-arid landscape as indicated by natural isotopes and chloride. J Hydrol 76(1–2):1–25
2. Alvarez-Benedi J, Munoz-Carpena R (eds) (2005) Soil-water-solute processes in environmental systems. Monitoring, characterization and modelling. CRC Press, Boca Raton
3. Barnes C, Allison G (1988) Tracing of water movement in the unsaturated zone using stable isotopes of hydrogen and oxygen. J Hydrol 100(1–3):143–176
4. Bear J (1972) Dynamics of fluids in porous media. Dover, New York
5. Birdsell KH, Newman BD, Broxton DE, Robinson BA (2005) Conceptual models of vadose zone flow and transport beneath the Pajarito Plateau, Los Alamos, New Mexico. Vadose Zone J 4(3):620–636
6. Carsel RF, Parrish RS (1988) Developing joint probability-distributions of soil-water retention characteristics. Water Resour Res 24(5):755–769
7. Constantz J, Tylerb SW, Kwicklisc E (2003) Temperature-profile methods for estimating percolation rates in arid environments. Vadose Zone J 2:12–24
8. Cook PG, Jolly ID, Leaney FW, Walker GR, Allan GL, Fifield LK, Allison GB (1994) Unsaturated zone tritium and Cl-36 profiles from Southern Australia—their use as tracers of soil-water movement. Water Resour Res 30(6):1709–1719
9. Dettinger MD (1989) Reconnaissance estimates of natural recharge to desert basins in Nevada, U.S.A., by using chloride-balance calculations. J Hydrol 106(1–2):55–78
10. Dyck MF, Kachanoski RG, de Jong E (2005) Spatial variability of long-term chloride transport under semiarid conditions: Pedon scale. Vadose Zone J 4(4):915–923
11. Faust AE, Ferre TPA, Schaap MG, Hinnell AC (2006) Can basin-scale recharge be estimated reasonably with water-balance models? Vadose Zone J 5(3):850–855
12. Feddes R, de Rooij G, van Dam J (eds) (2004) Unsaturated zone modelling: progress, challenges and applications. Kluwer, Dordrecht
13. Flint AL, Ellett KM (2004) The role of the unsaturated zone in artificial recharge at San Gorgonio Pass, California. Vadose Zone J 3(3):763–774
14. Gasser M, Caron J, Lagace R, Laverdiere M (2003) Predicting nitrate leaching under potato crops using transfer functions. J Environ Qual 32(4):1464–1473
15. Gee GW, Hillel D (1988) Groundwater recharge in arid regions—review and critique of estimation methods. Hydrol Process 2(3):255–266
16. Holt RM, Nicholl MJ (2004) Uncertainty in vadose zone flow and transport prediction. Vadose Zone J 3(2):480–484
17. Hubbell JM, Nicholl MJ, Sisson JB, McElroy DL (2004) Application of a Darcian approach to estimate liquid flux in a deep vadose zone. Vadose Zone J 3(2):560–569
18. Hupet F, van Dam JC, Vanclooster M (2004) Impact of within-field variability in soil hydraulic properties on transpiration fluxes and crop yields: a numerical study. Vadose Zone J 3(4):1367–1379
19. IBW (1987) Étude des ressources en eau du Brabant Wallon, contrat Région Wallonne. Technical report, Intercommunale du Brabant Wallon
20. Javaux M, Vanclooster M (2003) Robust estimation of the generalized solute transfer function parameters. Soil Sci Soc Am J 67(1):81–91
21. Javaux M, Vanclooster M (2003) Scale- and rate-dependent solute transport within an unsaturated sandy monolith. Soil Sci Soc Am J 67(5):1334–1343
22. Javaux M, Vanclooster M (2004) In situ long-term chloride transport through a layered, nonsaturated subsoil. 2. Effect of layering on solute transport processes. Vadose Zone J 3(4):1331–1339
23. Javaux M, Vanclooster M (2006) Scale-dependency of the hydraulic properties of a variably saturated heterogeneous sandy subsoil. J Hydrol 327(3–4):376–388
24. Javaux M, Vanderborght J, Kasteel R, Vanclooster M (2006) Three-dimensional modeling of the scale-and flow rate-dependency of dispersion in a heterogeneous unsaturated sandy monolith. Vadose Zone J 5(2):515–528
25. Jury WA (1982) Simulation of solute transport using a transfer-function model. Water Resour Res 18(2):363–368
26. Jury WA, Flühler H (1992) Transport of chemicals through soil: mechanisms, models, and field applications, vol. 47. Academic Press, New York 141–201
27. Kim SJ, Hyun Y, Lee KK (2005) Time series modeling for evaluation of groundwater discharge rates into an urban subway system. Geosci J 9(1):15–22
28. Leterme B, Vanclooster M, Rounsevell MD, Bogaert P (2006) Discriminating between point and non-point sources of atrazine contamination of a sandy aquifer. Sci Total Environ 362(1–3):124–142

29. Levitt DG, Newell DL, Stone WJ, Wykoff DS (2005) Surface water-groundwater connection at the Los Alamos Canyon weir site: Part 1. Monitoring site installation and tracer tests. Vadose Zone J 4(3):708–717
30. Mallants D, Jacques D, Vanclooster M, Diels J, Feyen J (1996) A stochastic approach to simulate water flow in a macroporous soil. Geoderma 70(2–4):299–324
31. Mattern S, Bogaert P, Vanclooster M (2008) Advances in subsurface pollution of porous media—indicators, processes and modelling: IAH selected papers, vol 14. Taylor and Francis. Introducing time variability and sampling rate in the mapping of groundwater contamination by means of the Bayesian Maximum Entropy (BME) method. IAH—Selected Papers on Hydrogeology, ISBN:9780415476904
32. McElroy DL, Hubbell JM (2004) Evaluation of the conceptual flow model for a deep vadose zone system using advanced tensiometers. Vadose Zone J 3(1):170–182
33. MRW-DGRNE Ddes (2008) Banque de données "10-sous". Technical report, 15, Avenue Prince de Liège, B-5100 Jambes
34. Nativ R, Adar E, Dahan O, Geyh M (1995) Water recharge and solute transport through the vadose zone of fractured chalk under desert conditions. Water Resour Res 31(2):253–261
35. O'Geen AT, McDaniel PA, Boll J (2002) Chloride distributions as indicators of vadose zone stratigraphy in Palouse loess deposits. Vadose Zone J 1:150–157
36. Oger R (1991) Rétrospective climatologique de la période 1950–1989, poste d'Ernage-Gembloux, 106 pp. Technical report, Centre de Recherches Agronomiques de Gembloux
37. Onsoy YS, Harter T, Ginn TR, Horwath WR (2005) Spatial variability and transport of nitrate in a deep alluvial vadose zone. Vadose Zone J 4(1):41–54
38. Ott WR (1990) A physical explanation of the lognormality of pollutant concentrations. J Air Waste Manag Assoc 40(10):1378–1383
39. Robinson BA, Cole G, Carey JW, Witkowski M, Gable CW, Lu ZM, Gray R (2005) A vadose zone flow and transport model for Los Alamos Canyon, Los Alamos, New Mexico. Vadose Zone J 4(3):729–743
40. Schaap M, Leij F, Genuchten MV (2001) Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J Hydrol 251(3–4):163–176
41. Si BC, de Jong E (2007) Determining long-term (decadal) deep drainage rate using multiple tracers. J Environ Qual 36(6):1686–1694
42. Simunek J, van Genuchten MT, Sejna M (2008) Development and applications of the HYDRUS and STANMOD software packages and related codes. Vadose Zone J 7(2):587–600
43. Stewart IT, Loague K (1999) A type transfer function approach for regional-scale pesticide leaching assessments. J Environ Qual 28(2):378–387
44. Tank AMGK, Wijngaard JB, Konnen GP, Bohm R, Demaree G, Gocheva A, Mileta M, Pashiardis S, Hejkrlik L, Kern-Hansen C, Heino R, Bessemoulin P, Muller-Westermeier G, Tzanakou M, Szalai S, Palsdottir T, Fitzgerald D, Rubin S, Capaldo M, Maugeri M, Leitass A, Bukantis A, Aberfeld R, Van Engelen AFV, Forland E, Mietus M, Coelho F, Mares C, Razuvaev V, Nieplova E, Cegnar T, Lopez JA, Dahlstrom B, Moberg A, Kirchhofer W, Ceylan A, Pachaliuk O, Alexander LV, Petrovic P (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int J Climatol 22(12):1441–1453
45. Vanclooster M, Mallants D, Diels J, Feyen J (1993) Determining local-scale solute transport parameters using time domain reflectometry (tdr). J Hydrol 148(1–4):93–107
46. Vanderborght J, Vanclooster M, Mallants D, Diels J, Feyen J (1996) Determining convective lognormal solute transport parameters from resident concentration data. Soil Sci Soc Am J 60(5):1306–1317
47. van der Velde M, Javaux M, Vanclooster M, Clothier BE (2006) El Nino-Southern Oscillation determines the salinity of the freshwater lens under a coral atoll in the Pacific Ocean. Geophys Res Lett 33(21):L21403
48. Van Genuchten M (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci Soc Am J 44:892–898
49. Wang W, Neuman SP, Yao Tm, Wierengac PJ (2003) Simulation of large-scale field infiltration experiments using a hierarchy of models based on public, generic, and site data. Vadose Zone J 2:297–312
50. Wu YS, Lu GP, Zhang K, Bodvarsson GS (2004) A mountain-scale model for characterizing unsaturated flow and transport in fractured tuffs of Yucca Mountain. Vadose Zone J 3(3):796–805

Environ Fluid Mech (2010) 10:137–156
DOI 10.1007/s10652-009-9119-6

ORIGINAL ARTICLE

RANS-based simulation of transverse turbulent mixing in a 2D geometry

Carlo Gualtieri

Received: 26 November 2008 / Accepted: 31 January 2009 / Published online: 6 March 2009

Abstract Although transverse mixing is a significant process in river engineering when dealing with the discharge of pollutants from point sources or the mixing of tributary inflows, no theoretical basis exists for the prediction of its rate, which is indeed based upon the results of experimental works carried on in laboratory channels or in streams and rivers. The paper presents the preliminary results of a numerical study undertaken to simulate the transverse mixing of a steady-state point source of a tracer in a two-dimensional rectangular geometry, which is expected to reproduce a shallow flow. This geometry is that of Lau and Krishnappan (J Hydraul Div 13(HY10):1173–1189, 1977), who collected turbulent mixing data for a shallow flow. In the numerical study an approach based on the Reynolds Averaged Navier–Stokes (RANS) equations was applied, where the closure problem was solved by using turbulent viscosity concept. Particularly, the classical two-equations k–ε model was used. Two methods were applied to the model results to evaluate the turbulent transverse mixing coefficient. The effect on transverse mixing of a grid located upstream the tracer source was also studied. Numerical results were generally higher than the experimental data. This overestimation could be explained considering the hypothesis of isotropic turbulence underlying the k–ε model, which can lead to large turbulent viscosities and rate of mixing. However, RANS-based results may still be considered acceptable also providing the large uncertainties associated with literature predictive equations.

Keywords Environmental hydraulics · Turbulence · Transverse turbulent mixing · Numerical simulations · Grid-turbulence · Experimental verification

C. Gualtieri (✉)
Department of Hydraulic, Geotechnical and Environmental Engineering (DIGA), University of Napoli "Federico II", Via Claudio, 21, 80125 Naples, Italy
e-mail: carlo.gualtieri@unina.it

1 Introduction

The mixing of contaminants in streams and rivers is a significant problem in environmental fluid mechanics and rivers engineering since to understand the impact and the fate of pollutants in these water bodies is a primary goal of water quality management [1]. Since most rivers have a high aspect ratio, that is the width to depth ratio, discharged pollutants become vertically mixed within a short distance from the source and vertical mixing is only important in the so-called near-field. As a rule of thumb, neutrally buoyant solute becomes fully mixed vertically within 50–75 depths from the source [2]. Notably, vertical mixing analysis relies on well-known theoretical basis, that is Prandtl mixing length model, which assumes the hypothesis of plane turbulent shear flow and provides theoretical predictions of the vertical turbulent diffusivity which closely match experimental results [2,3].

In the mid-field, the vertical concentration gradients are negligible and both subsequent transverse and longitudinal changes of the depth-averaged concentrations of the pollutants should be addressed. In the literature, for the application of one-dimensional water quality models the majority of research efforts were devoted to estimate the rate of longitudinal mixing of a contaminant, that is the development of a plume resulting from a temporally varying pollutant source once it has become cross-sectionally well-mixed, in the far-field [4]. However, recent analysis of predictive equations of longitudinal dispersion coefficient demonstrated that most of available formulas exhibit a limited accuracy [5,6]. Most of them are not applicable to all stream hydrodynamic conditions, but, on the contrary, they remain specific for the conditions where they were developed.

Although transverse mixing process is of significant importance when dealing with wastewater treatment plants discharges, cooling water returns and the mixing of tributary inflows, this process has received less attention by the researchers than longitudinal mixing [2,7]. In such situations, since steady-state conditions are approximated, that is temporal concentration gradients are small, the spreading across the channel is important and accurate modeling and prediction of transverse mixing is required [4]. Also, it is well known that transverse mixing is important in determining the rate of longitudinal mixing because it tends to control the exchange between regions of different longitudinal velocity. Particularly, transverse mixing and longitudinal mixing are inversely proportional. A strong transverse mixing tends to erase the effect of differential longitudinal advection and pollutants particles migrate across the velocity profile so fast that they essentially all move at the mean speed of the flow, causing only a weak longitudinal spreading. On the other hand, a weak transverse mixing implies a long time for differential advection to take effect, so the pollutants patch is highly distorted while it diffuses moderately in the transverse direction and longitudinal mixing is large [8]. Despite its importance, no established theory exists to predict transverse mixing rates, turbulent diffusion coefficient and its dependence on the various flow parameters must be determined from experimental works.

This paper presents preliminary results of a numerical study undertaken to simulate the transverse mixing of a steady-state point source of a tracer in a two-dimensional rectangular geometry, which is expected to reproduce a shallow flow. A shallow flow can be defined as a predominantly horizontal flow in a fluid domain for which the two horizontal dimensions greatly exceed the vertical dimension [9]. The considered geometry is that of Lau and Krishnappan [10], who collected transverse turbulent mixing data in a rectangular channel. Two methods were applied to the results obtained with Multiphysics 3.4^{TM}, a commercial modeling environment [11], to evaluate the transverse turbulent mixing coefficient. Numerical results were compared with Lau and Krishnappan experimental data. Finally, the effect on transverse mixing of a grid inserted upstream the point of injection was estimated.

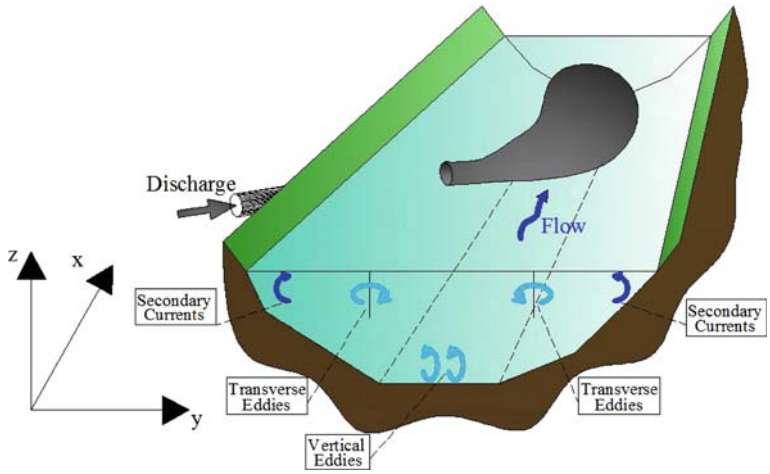


Fig. 1 Sketch of transverse mixing in a natural channel

2 Transverse turbulent mixing: literature review

2.1 Theory on transverse mixing

It is believed that transverse or lateral mixing is due to following causes [2,12,13] (Fig. 1):

- turbulence generated by the channel boundaries, which involves many eddies of various sizes and intensities, all embedded in one another [14]. These eddies are responsible for both momentum and mass transfer, according the Reynolds analogy [15] resulting in a contaminants mixing far exceeding that occurring in the molecular scale. Also it could be expected that in a turbulent flow the largest eddies regulate the rate of turbulent diffusion. in a river, lateral mixing is due to the transverse eddies that rotate horizontally, about a vertical axis;
- vertical variations in the transverse velocity (velocity shear), which are significant in the vicinity of channel banks and further contribute to transverse spreading of contaminants;
- secondary currents, which causes contaminants to move in opposite directions at different depths increasing the rate of mixing [2,16,17].

Prediction of transverse mixing rate is needed when using in the mid-field the two-dimensional depth-averaged advection-diffusion equation [2]:

$$h\frac{\partial \bar{C}}{\partial t}+\frac{\partial}{\partial x}\left(h\bar{u}\bar{C}\right)+\frac{\partial}{\partial y}\left(h\bar{v}\bar{C}\right)=\frac{\partial}{\partial x}\left(h\left(D_L+D_{t-x}\right)\frac{\partial \bar{C}}{\partial x}\right)+\frac{\partial}{\partial y}\left(h\left(D_T+D_{t-y}\right)\frac{\partial \bar{C}}{\partial y}\right) \quad (1)$$

where x and y are longitudinal and transverse distances, h is local depth, $\bar{C}$ is depth-averaged *contaminants concentration*, $\bar{u}$ and $\bar{v}$ *are depth-averaged velocities in the* x and y directions, D_{t-x} and D_{t-y} are turbulent diffusion coefficients in the x and y directions. Moreover, D_L is longitudinal dispersion coefficient, which accounts for the effects of vertical variations of longitudinal velocity, and D_T is transverse dispersion coefficient, which accounts for the

effects of vertical variations of transverse velocity. Equation 1 confirms that transverse turbulent diffusion and transverse dispersion are different processes that produce the spreading of a solute in the lateral direction. Nevertheless, since in the field observations it is difficult to distinguish their own effects on contaminants spreading in the lateral direction, the effect of these processes is usually combined into a single mixing coefficient for convenience, that is D_{t-y} [2]. However, in some conditions, it is possible to argue that one process would be more important than the other one. Also, the effects of the secondary currents may play their role. For example, at bends of natural channels there are strong secondary currents which carry solute sideways in different directions at different depths increasing significantly the transverse mixing process [2]. On the other hand, in straight rectangular laboratory channels, the flow does not depart significantly from plane shear flow and transverse dispersion is negligibly small. Thus, in these conditions mixing is mainly due to turbulent diffusion even if secondary currents could occur increasing the rate of mixing in the channel. Nevertheless, since no established theory exists to predict transverse mixing rates even in those channels, turbulent diffusion coefficient and its dependence on the various flow parameters must be determined from experimental works. A good number of experimental works are available in the literature [2,10,12,13,18].

As previously recalled, a theoretical model for the transverse turbulent diffusion does not yet exist. Prandtl mixing length hypothesis relates vertical turbulent diffusivity D_{t-z} with vertical gradient of longitudinal velocity [13]:

$$D_{t-z} = L_m^2 \left| \frac{\partial \langle u \rangle}{\partial z} \right| \tag{2}$$

where z is vertical distance, D_{t-z} is vertical turbulent diffusion coefficient, $\langle u \rangle$ is ensemble mean longitudinal velocity and L_m is the mixing length, that is a turbulent length scale. Drawing an analogy for the transverse turbulent mixing process yields:

$$D_{t-y} = L_m^2 \left| \frac{\partial \langle u \rangle}{\partial y} \right| \tag{3}$$

However, in a plane shear flow, ensemble mean of u does not vary across the channel which implies that D_{t-y} vanishes in disagreement with experimental findings [2]. Equation 3 may be applied near the banks where there is a strong transverse velocity shear but in the thalweg that equation cannot be used. Nevertheless, the first part of Prandtl theory could be still considered because from dimensional reasoning it relates turbulent mixing coefficient to a turbulent length and velocity scales, L_T and U_T, respectively, as:

$$D_{t-y} = L_T U_T \tag{4}$$

In a plane shear flow turbulence is generated by vertical velocity shear which arises as a result of bed friction [2]. Since shear velocity u^* is a measure of bed friction, it could be selected as velocity length scale in Eq. 4. This is also consistent with literature for vertical diffusivity and longitudinal dispersion coefficient [2]. There is indeed something of controversy about the proper length scale to be used in Eq. 4. It is very common to assume flow depth h as length scale for transverse diffusivity too, that is $h = L_T$, since this parameter controls the largest vertical eddies. However, it may be argued that the maximum length scale of transverse eddies, which are responsible for transverse turbulent mixing, cannot be restricted to the water depth [2]. Thus, although it is not clear how vertical eddies generated by bed friction undergo rotation and become transverse eddies, it was also suggested to apply channel width W as characteristic turbulent length scale for D_{t-y} [10].

2.2 Analysis of experimental data in a straight rectangular channel

Gualtieri and Mucherino re-analyzed a 217 literature data set collected in straight rectangular laboratory channels [19]. The data set was obtained bringing together 139 data collected by Rutherford [2] from different investigators and 78 experimental data more recently collected by Chau under different bottom roughness conditions, but always in the same 4.0 m width laboratory channel [18]. This data set is believed to be the largest ever prepared. The analysis of these data confirmed that a linear relationship holds between D_{t-y} and hu^* [19]:

$$D_{t-y} = 0.166 hu^* \tag{5}$$

Note that Fischer et al. [20] from the review of several studies proposed:

$$D_{t-y} = 0.15 hu^* \tag{6}$$

where there is the possibility of an error of ±50%. On the other hand, Rutherford stated that the most likely value of the transverse diffusivity in plane shear flow is for 0.13 [2]. Therefore, Eq. 5 provides a larger value for D_{t-y} than previous literature equations.

Second, the analysis confirmed that D_{t-y}/Wu^* decreased with the increasing aspect ratio W/h approaching a constant value for high aspect ratio [19]. Third, the analysis of experimental data lead to derive for $W/h < 7.5$ [19]:

$$D_{t-y} = 0.0315 Wu^* \tag{7}$$

Finally, they suggested that the use of flow depth h or channel width W as characteristic length scale of transverse turbulent mixing in straight rectangular channels is equivalent [19].

As above outlined, a shallow flow can be defined as having a lateral extent greater than its vertical confinement, as is the case in natural rivers, estuaries, stratified layers in lakes, the upper ocean and even for large scale motions in the oceans [19,21]. Experiments in shallow flow as well as numerical stability analysis demonstrated that these flows are inherently unstable to lateral velocity shear. These instabilities often result in the formation of 2D coherent vertical structures, that are greater than the water depth [9,21,22].

Recently a turbulent diffusion model in a shallow flow was developed by Rummel et al. [21], as it follows. They considered a vertical line-source injection into a uniform shallow flow, where far from any boundaries the turbulence field at a fixed water depth can be considered homogeneous and stationary in the horizontal directions. Since in a continuous injection the downstream concentration gradient is negligible and the slender plume approximation can be invoked, they considered one-dimensional diffusion in the lateral direction only, in analogy with Taylor's theorem [21]. From this approach they evaluated transverse turbulent mixing coefficient for only bottom-generated turbulence and for three cases of grid turbulence highlighting that in grid-turbulence transverse mixing was stronger [21]. This result suggested to compare numerical data with and without a grid, as below described.

3 Modeling turbulent shallow flows. Numerical simulations

3.1 Modeling turbulent shallow flows

As above explained, in shallow flow turbulent field can be considered as homogeneous and stationary in the horizontal plane and mixing is governed by two-dimensional coherent structures [21]. Therefore, as a first approximation two-dimensional or depth-averaged models

may be applied to describe hydrodynamics and mass-transfer processes. The simplest level of modeling the turbulent flow is the Reynolds Averaged Navier–Stokes approach. Since the flow is turbulent, mass conservation equation and the Navier–Stokes equations of motion must be averaged over a small time increment applying Reynolds decomposition, where flow quantities are decomposed in a temporal mean and a fluctuating component. The fluctuations include all the turbulent motions. The application of such decomposition results in the Reynolds-averaged Navier–Stokes equations (RANS), where the effect of turbulence appears as a number of terms representing the interaction between the fluctuating velocities and termed Reynolds stresses. They introduce closure problem, which can be solved, in analogy with the viscous stresses in laminar flow, by using an *eddy viscosity* or turbulent viscosity concept [23]. According to the kinetic theory of gases, the molecular viscosity of a fluid is proportional to the product of the molecular mean free path and the average speed of the molecules. By analogy, the eddy viscosity can also be expressed as a product of the characteristic turbulent length and velocity scales. Hence, dimensional reasoning provides for eddy viscosity an equation like Eq. 4. Different approaches can be used to derive these scales. At the simplest level of complexity, one may expect that eddy viscosity would be determined by large-scale eddies, the size of which is close to the characteristic dimension and velocity of the flow itself. Thus, eddy viscosity would be linked to the overall velocity gradient, as proposed by Prandtl one century ago in his mixing length model [23]. This model nowadays is used only for an initial guess of the flow field [24], since it does not include any effects of the history of the flow and the turbulence transport on the mixing length [25]. In all turbulence models based on the concept of eddy viscosity, an obvious choice for defining U_T is the turbulent kinetic energy k, so that:

$$U_T = k^{1/2} \tag{8}$$

Note that $k^{1/2}$ is also used as a measure of the averaged turbulent intensity [23]. A transport equation for k can easily be derived. On the other hands, defining and providing adequate length scale L_T is more difficult and uncertain [24] and many variants were proposed in the literature. Thus, two basic classes of differential turbulence models based on the concept of eddy viscosity can be distinguished, depending on how many differential equations need to be solved to provide eddy viscosity [24]:

- one-equation models, where only the differential transport equation for k is solved, whereas L_T is defined algebraically, usually in term of flow geometrical parameters;
- two-equations models, where, in addition to k-equation, another differential equation is solved, which provides the characteristic turbulence length scale L_T, either directly, or in combination with k. In these models, the length scale is obtained from a *scale-providing* differential transport equation, in which is present a product involving the length scale and the turbulent kinetic energy [24]:

$$\begin{aligned}\frac{D\left(k^p L_T^q\right)}{Dt} &= \text{Production of } \left(k^p L_T^q\right) - \text{Destruction of } \left(k^p L_T^q\right) \\ &\quad + \text{Diffusion of } \left(k^p L_T^q\right)\end{aligned} \tag{9}$$

  depending on the choice of p and q, the variable L_T can have different physical meaning. The most popular scale-providing variable is the rate of turbulent energy dissipation rate ε, which is the exact sink term in the equation for k. Equations for k and ε, together with the eddy-viscosity stress–strain relationship constitute the k–ε turbulence model. By solving the two transport equations for k and ε, turbulent viscosity can be estimated.

Another widely applied *scale-providing* variable is $\omega = \varepsilon / k$, which can be interpreted as a characteristic frequency and gives rise to the k–ω turbulence model.

The k–ε model, k–ω and their variation are the most widely used turbulence models and this is largely due to their ease in implementation, economy in computation and, most importantly, being able to obtain reasonable accurate solution with the available computer power [25]. However, several shortcomings have been discovered over three decades of use and validation. In open channel flows modeling, it is known that in the case of prismatic channels where there are no geometrical variations along the channel, the k–ε turbulence model fails to predict any evidence of secondary flow. This is because the k–ε model assumes that the turbulence is isotropic, whereas turbulence is known to be anisotropic. In fact, it is the anisotropic behavior of turbulence as it approaches the walls and free surface, namely the imbalance in the normal Reynolds stresses, that creates secondary circulation even if in a straight channel [26]. Also, the assumption of isotropy, can lead the k–ε model to predict large turbulent viscosity and, consequently, high turbulent diffusivities. Nevertheless, recent studies demonstrated that for simplified cases, where mean velocities and bulk mixing properties are needed, RANS-modeling of shallow flows is still appropriate [27]. More generally, the k–ε model may be recommended for a quick preliminary estimation of the flow field, or in situations where modeling other physical phenomena, such as chemical reactions, combustion, radiation, multi-phase interactions, brings in uncertainties that overweigh those inherent in the k–ε model itself [24].

3.2 Numerical simulations

The above discussion supports a preliminary application of k–ε model to predict the major mean-flow features and the bulk mixing properties. For a planar, incompressible flow RANS equations are:

$$\frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{v}}{\partial y} = 0 \tag{10}$$

$$\begin{aligned} \bar{\rho}\left(\frac{\partial \bar{u}}{\partial t} + \bar{u}\frac{\partial \bar{u}}{\partial x} + \bar{v}\frac{\partial \bar{u}}{\partial y}\right) &= \bar{\rho} g_x - \frac{\partial \bar{p}}{\partial x} + \left(\mu + \rho \nu_t\right) \nabla^2 \bar{u} \\ \bar{\rho}\left(\frac{\partial \bar{v}}{\partial t} + \bar{u}\frac{\partial \bar{v}}{\partial x} + \bar{v}\frac{\partial \bar{v}}{\partial y}\right) &= \bar{\rho} g_y - \frac{\partial \bar{p}}{\partial y} + \left(\mu + \rho \nu_t\right) \nabla^2 \bar{v} \end{aligned} \tag{11}$$

where ρ and μ are fluid density and viscosity, p is fluid pressure and u, v are velocity components in the x and y directions, respectively. The overbar indicates time-averaged quantities. Notably, in Eq. 11 there is the eddy kinematic viscosity ν_t, that in the above assumption of isotropic turbulence could be estimated as:

$$\nu_t = \frac{C_\mu k^2}{\varepsilon} \tag{12}$$

where k and ε are turbulent kinetic energy per mass unit and its dissipation rate, respectively, and $C_\mu = 0.09$. To estimate these parameters the two-equations of standard k–ε model are [11]:

$$\rho \frac{\partial k}{\partial t} + \rho \vec{\bar{V}} \cdot \nabla k = \nabla \cdot \left[\left(\frac{\mu + \mu_t}{\sigma_k}\right) \nabla k\right] + \frac{1}{2}\mu_t \left(\nabla \vec{\bar{V}} + \left(\nabla \vec{\bar{V}}\right)^{T^2} - \rho \varepsilon\right) \tag{13}$$

Table 1 Values of the constants of the standard k–ε model

C_μ	σ_k	σ_ε	$C_{1\varepsilon}$	$C_{2\varepsilon}$
0.09	1.00	1.30	1.44	1.92

$$\rho\frac{\partial\varepsilon}{\partial t}+\rho\overline{\vec{V}}\cdot\nabla\varepsilon=\nabla\cdot\left[\left(\frac{\mu+\mu_t}{\sigma_\varepsilon}\right)\nabla\varepsilon\right]+\frac{1}{2}C_{1\varepsilon}\frac{\varepsilon}{k}\mu_t\left(\nabla\overline{\vec{V}}+\left(\nabla\overline{\vec{V}}\right)^T\right)^2-\rho C_{2\varepsilon}\frac{\varepsilon^2}{k} \tag{14}$$

where μ_t is dynamic eddy viscosity, whereas C_μ, σ_k, σ_ε, $C_{1\varepsilon}$ and $C_{2\varepsilon}$ are constants and their values are listed in Table 1.

The transport of solutes in the rectangular geometry could be modelled using the 2D advection-diffusion equation for isotropic turbulence [11]:

$$\frac{\partial\bar{C}}{\partial t}+\bar{u}\frac{\partial\bar{C}}{\partial x}+\bar{v}\frac{\partial\bar{C}}{\partial y}=\frac{\partial}{\partial x}\left(D_t\frac{\partial\bar{C}}{\partial x}\right)+\frac{\partial}{\partial y}\left(D_t\frac{\partial\bar{C}}{\partial y}\right) \tag{15}$$

where molecular diffusion was neglected and only turbulent diffusion was considered with D_t as turbulent diffusivity and $\bar{C}$ as solute concentration.

These equations were solved using Multiphysics 3.4$^{\text{TM}}$ modeling package, which is a commercial multiphysics modeling environment [11]. Multiphysics 3.4$^{\text{TM}}$ can solve for the same flow domain both motion equations and advection-diffusion equation. Particularly, both the k–ε Turbulence Model application mode and the Advection and Diffusion application mode were used. They solve from Eqs. 10 to 15 for the pressure $\bar{p}$, the velocity vector components $\bar{u}$ and $\bar{v}$, the k–ε model parameters and the solute concentration $\bar{C}$ within the domain of the flow [11].

Multiphysics 3.4$^{\text{TM}}$ was applied to the two rectangular 2D geometries presented in Fig. 2. They are both the view from above, i.e. in the x-y plan, of the channel used by Lau and Krishnappan in their experimental works on turbulent transverse mixing [10]. However, in Geometry B there was a grid formed by 6 squares elements, each with dimension 0.02×0.02 m, with a center-to-center spacing $d = 0.06$ m. The grid was located 0.5 m upstream the point of injection. In Geometry A the square grid was absent.

Lau and Krishnappan conducted their experiments in a rectangular flume 30.7 m long and 0.60 m wide. Extra sidewalls were installed to convert the flume to 0.45 or 0.30 m widths. Also flume bed roughness was varied by using sands of different sizes. In addition, some experiments were made with a smooth coverings on the bottom flume. The flow depth varied from 0.013 to 0.050 m, whereas mean flow velocities were kept approximately constant at 0.20 m/s for most of the experiments. Mixing measurements were made using salt solution as the tracer. The solution was continuously discharged from a constant head injection apparatus into the middle of the flume at approximately mid-depth. Injections were made 11 m downstream from the beginning of the flume to ensure fully developed boundary layers for all the experiments. The tracer concentrations were measured at mid-depth at 8 or more stations downstream the injection point by using a single electrode conductivity probe. The probe was moved along channel width to derive cross-stream concentration distributions. These stations were chosen so that vertical mixing was completed before the first station [10].

For the simulations water with density $\rho = 999.05\,\text{kg/m}^3$ and molecular viscosity $\mu = 1.14 \times 10^{-3}\,\text{kg/m}\cdot\text{s}$, was selected as fluid. Inflow velocity was 0.20 m/s. For both Geometry A and B, for the k–ε model, boundary conditions were assigned at the inlet, the outlet and at the walls:

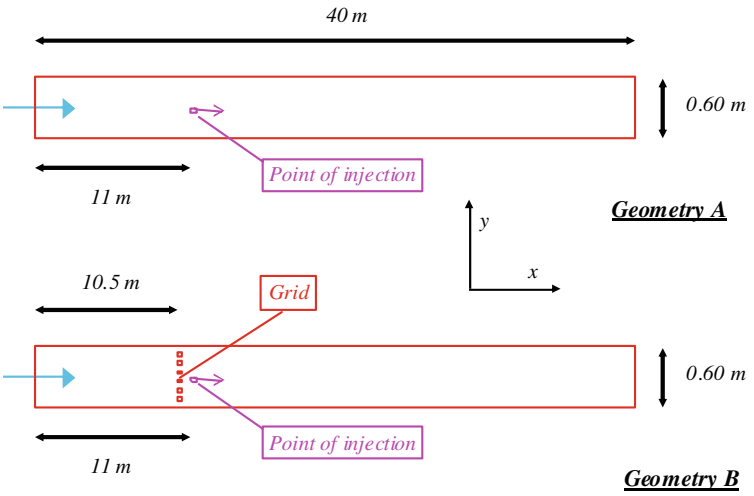


Fig. 2 Sketch of the simulated 2D geometries

- at the inlet, an *inflow* type boundary condition was applied, with uniform velocity profile. Also inlet turbulent intensity and length scale were assigned. Turbulent intensity was set up to 5%, which corresponds to fully turbulent flows. Usually, turbulent intensity can be derived from the Reynolds number. The turbulent length scale is a physical quantity related to the size of the large eddies that contain the energy in turbulent flows. also it is a measure of the size of the turbulent eddies that are not resolved [11]. For fully developed channel flows, this parameter can be approximately derived as $0.07 \times W$, where W is the channel width;
- at the outlet, a *zero pressure* type condition was assigned;
- at the walls and at the square elements, *logarithmic law of the wall* boundary condition was applied. It is well-known that turbulent flows are significantly affected by the presence of walls. First, no-slip condition must be satisfied at the walls for the mean velocity field. Very close to the wall, viscous damping reduces the tangential velocity fluctuations, while kinematic blocking reduces the normal fluctuations. Turbulent eddies are distorted and constrained in size, being compressed in the wall-normal direction and elongated in the streamwise direction [28]. Moreover, as at the wall turbulent energy is damped and dissipated into heat through molecular viscosity, toward the outer part of the near-wall region, however, the turbulence is rapidly increasing by the production of turbulence kinetic energy due to the large gradients in the mean velocity and in the other parameters of the flow field. Also, in this region it should be expected that momentum and scalar transport occur most vigorously. Hence, accurate flow modelling in the near-wall region significantly affects successful prediction of turbulent wall-bounded flows. The classical k–ε model due to its basic hypothesis of isotropy needs to be modified to account for the effect of the walls on the local structure of turbulence [28]. Different approaches were proposed in the literature [28]. Herein to account for solid walls the approach based on the so-called *wall functions* was applied to bridge the viscosity-affected region between the wall and the fully turbulent region. This approach is the most commonly because it eliminates the need for very fine computational meshes that are required for direct integration of the governing equations all the way to the wall and because it reduces numerical stiffness due to the stiff source

terms in the turbulence closure equations and large aspect ratio meshes [28]. Basically a wall function is a semi-empirical relation between the value of velocity and wall friction which replaces the thin boundary layer near the wall [11]. This approach is expected to be accurate for high Reynolds numbers and situations where pressure variations along the walls are not very large. However, it can often be used outside its frame of validity with reasonable success [11]. In particular, logarithmic wall functions applied to finite elements assume that the computational domain begins a distance δ_w from the real wall. They also assume that the flow is parallel to the wall and that the velocity can be described by:

$$u^+ = \frac{u}{u^*} = \frac{1}{\kappa}\ln\left(\frac{\delta_w u^*}{\nu}\right) + C_1 \tag{16}$$

where κ is Von Kármán constant, which is equal to 0.42, and C_1 is universal constant for smooth walls equal to 5.5. Note that the ratio of the kinematic viscosity to the friction velocity, ν/u^* is the viscous length scale. Moreover, the term in the round brackets is δ_w^+ and the logarithmic wall functions are formally valid for values of between 30 and 100. Multiphysics 3.4^{TM} allows to select δ_w^+ value, which was assigned equal to 100.

In both Geometry A and B, for the advection-diffusion equation, boundary conditions were assigned at the inlet, the outlet, at the walls and in the injection point:

- at the inlet, a *concentration* type boundary condition was applied, assuming zero concentration entering the domain;
- at the outlet, an *advective flux* type condition was assigned;
- at the walls, an *insulation* type condition was applied. this condition means that the solute cannot cross the walls;
- at the injection point, a *concentration* type boundary condition was applied, assuming herein a constant concentration equal to 100.

Different mesh characteristics were tested. After that, the mesh generation process was made assuming, among the others, as *element growth rate* and *resolution of narrow regions* 1.2 and 1.00, respectively. The *element growth rate* determines the maximum rate at which the element size can grow from a region with small elements to a region with larger elements [11]. The value must be greater or equal to 1. In the *resolution of narrow regions* field the number of layer of elements created in narrow regions is controlled [11].

For both Geometry A and B, different values for the maximum element size were selected for the geometry upstream and downstream the point of injection. Finer values were selected downstream of this point to better capture transverse mixing. Maximum element sizes were 0.1 and 0.05 m in the upstream region and in the downstream region, respectively. Also, for both Geometry A and B, at the walls, maximum element size was 0.05 m, whereas, at the square elements in Geometry B, maximum element size was 0.01 m. Thus, the mesh for geometry A and B has 35,124 and 37,284 triangular elements, respectively, with a minimum *element quality* of 0.617 and 0.642, respectively. The *element quality* measure is related to its aspect ratio, which means that anisotropic elements can get a low quality measure even though the element shape is reasonable [11]. It is a scalar from 0 to 1. Mesh quality visualization demonstrated a quite uniform quality of the elements of the mesh. Moreover, it is well-known that by using a mesh, the dependent variables are approximated with a function that can be described with a finite number of parameters, the degrees of freedom [11]. In the two simulations, the number of degrees of freedom was 377,718 and 400,846, respectively.

About the solver settings, stationary segregated solver with non-linear system solver was used, where the relative tolerance and the maximum number of segregated iterations were set

to 1.0×10^{-3} and 100, respectively. The segregated solver allows to split the solution steps into substeps. These are defined by grouping solution components together. This procedure can save both memory and assembly time. Three groups were considered, namely velocity components $\bar{u}$ and $\bar{v}$ and pressure $\bar{p}$, *turbulence model parameters* k and ε, and the solute concentration $\bar{C}$. For both the simulations, artificial diffusion methods were not introduced neither in the flow field solution nor in the concentration field solution.

4 Data post-processing: transverse mixing coefficient estimation

Data resulting from numerical simulations were later post-processed to estimate the overall transverse mixing coefficient. Starting point were the simulated cross-stream concentration distributions of the tracer. Two methods were applied to derive D_{t-y} for the whole geometry:

- the method of moments, where the transverse mixing coefficient is derived from the rate of change of spatial variance of the tracer transverse profile [2]:

$$D_{t-y} = \frac{1}{2}\frac{d}{\mathrm{d}t}\left(\sigma_y^2\right) \tag{17}$$

where t is the time of travel and the transverse variance is given by [2]:

$$\sigma_y^2 = \frac{\int_{y=0}^{W}(y-\tilde{y})\bar{C}(x,y)\mathrm{d}y}{\int_{y=0}^{W}\bar{C}(x,y)\mathrm{d}y} \tag{18}$$

and the centroid of tracer distribution is [2]:

$$\tilde{y} = \frac{\int_{y=0}^{W} y\bar{C}(x,y)\mathrm{d}y}{\int_{y=0}^{W}\bar{C}(x,y)\mathrm{d}y} \tag{19}$$

Equations 17–19 can be applied provided that four conditions are accomplished, as explained by Rutherford [2]. First, tracer mixing must obey Fick's law with a constant mixing coefficient. second, the tracer must not impinge on the channel banks. Third, the tracer source must be steady. Fourth, the channel must be uniform so that the plume does not expand and contract along the transverse cross-section and its velocity is constant. Therefore, Eq. 17 can be approximated by the finite difference form [2]:

$$D_{t-y} = \frac{u}{2}\frac{\sigma_y^2(x_2)-\sigma_y^2(x_1)}{x_2-x_1} \tag{20}$$

where u is the average flow velocity, which is assumed to be constant. The cross-sectional values of D_{t-y} were later averaged along the geometry. Notably in a fickian process, transverse turbulent mixing coefficient D_{t-y} and transverse variance σ_y^2 are related as [2]:

$$\sigma_y^2 = 2D_{t-y}t \tag{21}$$

which may allow to derive another estimation of the average transverse mixing coefficient D_{t-y} in each cross-section.

- The method based on the transverse profile of turbulent kinematic viscosity ν_t, which provides a local value of the turbulent diffusivity. note that ν_t values were obtained from Eq. 12. The average value for each transverse profile provides the value of D_{t-y} for that cross-section. In turn, averaging cross-sectional D_{t-y} values along the geometry provides again the value for the whole geometry.

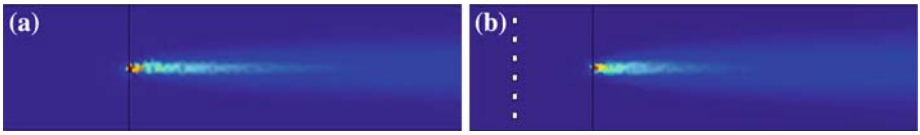


Fig. 3 Concentration field in the Geometry (**a**) n.1 and (**b**) n.2

The average flow velocity u was obtained averaging u values along the geometry centreline.

Other methods were proposed in the literature to estimate the transverse mixing rate. First, to estimate the degree of mixing in each cross-section of the geometry, the mixing index proposed by Rutherford can be calculated [2]:

$$P_{mix} = \frac{\bar{C}_{min}}{\bar{C}_{max}} \tag{22}$$

where $\bar{C}_{min}$ and $\bar{C}_{max}$ were minimum and maximum cross-sectional concentration of the tracer, respectively. P_{mix} values may be used to evaluate D_{t-y} as described by Rutherford [2].

Second, simulated cross-section tracer concentration distributions can be applied to derive a further estimation of the D_{t-y} in each cross-section. In fact, the width of a fickian plume L_{plume} is related to the properties of Gaussian distribution. It is well-known that 95% of the area under the Gaussian curve lies in the interval $-1.96 < (y/\sigma_y) < +1.96$, whereas the 99.74% of this area is in the range $-3.00 < (y/\sigma_y) < +3.00$. At the extremes of this interval, the concentration is equal to $0.011 \times \bar{C}_{max}$. Therefore, from Eq. 21, it holds:

$$L_{plume} = 6\sigma_y = 6\sqrt{2D_{t-y}t} \tag{23}$$

Since the length of the plume in the transverse direction L_{plume} can be obtained from the simulated cross-section tracer concentration distributions, Eq. 23 may provide a further estimation of D_{t-y} in each cross-section.

5 Analysis of numerical results: discussion

Numerical simulations provided field velocity and pressure, k and ε, kinematic viscosity ν_t values throughout the flow domain of both geometries, that is without and with the row of square elements. Figure 3a and b present the concentration field from $x = 10.2$ m and $x = 13.0$ m in the Geometry n.1 and n.2, respectively. Note that the injection point was at $x = 11.0$ m and the grid was located at $x = 10.5$ m. The plume width was larger in the grid geometry.

Table 2 lists the locations of the cross-section where simulated cross-section tracer concentration distributions were collected. For each cross-section 120 points were considered.

Figure 4a and b presents the cross-section tracer concentration distributions at $x = 11.125$ m and $x = 11.25$ m, respectively.

Figure 5a and b shows the cross-section tracer concentration distributions at $x = 11.50$ m and $x = 12.00$ m, respectively. Note that after 1.00 m from the injection point, the maximum cross-section concentration decreased for both the Geometries to about 39% of the maximum value at $x = 11.125$ m.

Table 2 Locations of cross-section tracer concentration distributions

Reach	x (m)	Δx (m)	Geometry n.1—No-grid		Geometry n.2—Grid	
			$\bar{C}_{max}$ (mole/m^3)	P_{mix}	$\bar{C}_{max}$ (mole/m^3)	P_{mix}
1	11.125	–	56.91	0.000	51.05	0.000
2	11.25	0.125	42.44	0.000	36.98	0.000
3	11.50	0.25	31.51	0.000	27.10	0.000
4	11.75	0.25	25.87	0.000	22.43	0.000
5	12.00	0.25	22.66	0.000	19.64	0.000
6	12.25	0.25	20.44	0.000	17.68	0.000
7	12.50	0.25	18.77	0.000	16.26	0.000
8	12.75	0.25	17.50	0.000	15.13	0.000
9	13.00	0.25	16.45	0.000	14.21	0.000
10	13.50	0.50	14.85	0.000	12.82	0.001
11	15.00	1.50	12.01	0.003	10.31	0.014
12	15.50	0.50	11.40	0.007	9.77	0.023
13	17.50	2.00	9.65	0.044	8.28	0.081
14	20.00	2.50	8.24	0.138	7.22	0.185
15	25.00	5.00	6.53	0.398	6.15	0.410
16	30.00	5.00	5.62	0.637	5.58	0.601
17	35.00	5.00	5.17	0.800	5.16	0.751
18	40.00	5.00	4.94	0.895	4.88	0.858

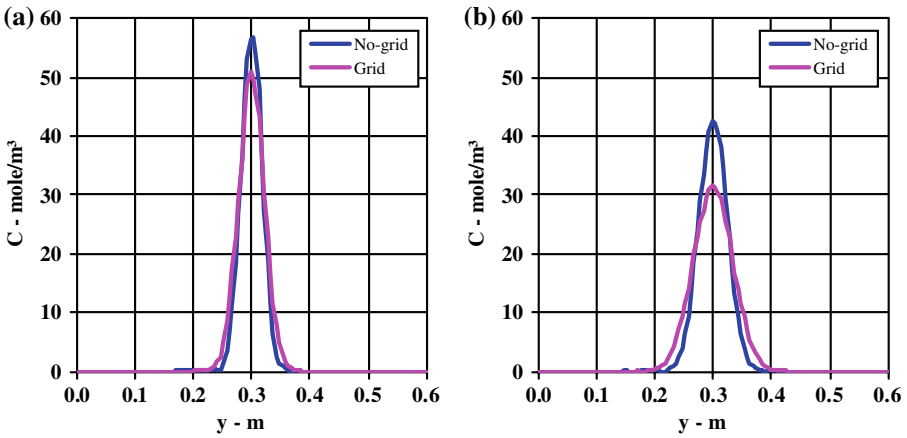


Fig. 4 Concentration transverse distributions at (**a**) $x = 11.125$ m and (**b**) $x = 11.25$ m

Figure 6a and b presents the cross-section tracer concentration distributions at $x = 15.00$ m and $x = 17.50$ m, respectively. Note that at $x = 15.00$ m in the Geometry n.2 the plume reached the walls, whereas in the Geometry n.1 it is still away from them.

Finally, Fig. 7a and b shows the cross-section tracer concentration distributions at $x = 25.00$ m and $x = 40.00$ m, respectively. In this final cross-section, the tracer is in both geometries almost completely mixed in the geometry width and its maximum concentration is $< 10\%$ of the initial maximum concentration.

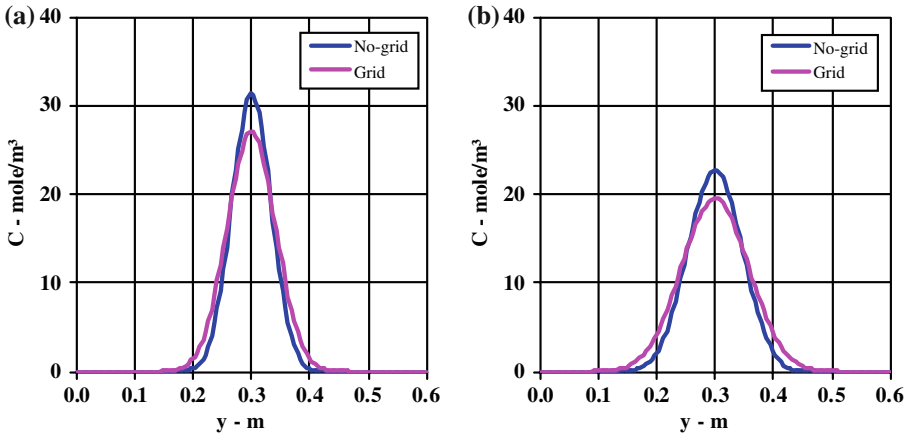


Fig. 5 Concentration transverse distributions at (**a**) x =11.50 m and (**b**) x =12.00 m

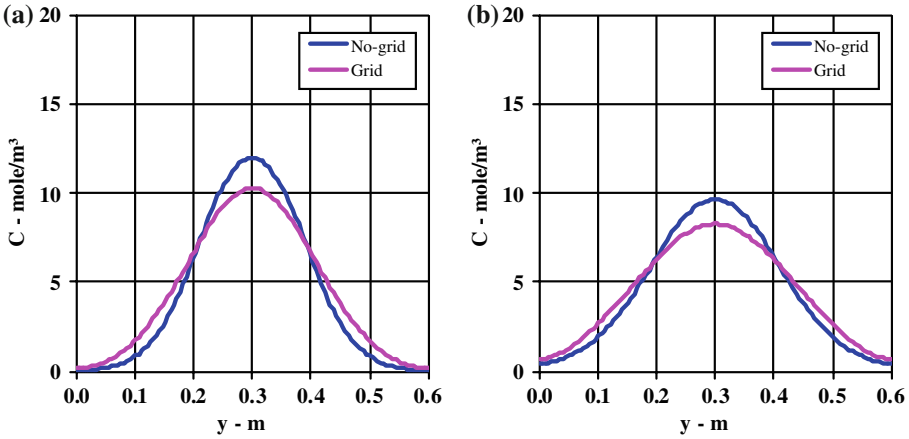


Fig. 6 Concentration transverse distributions at (**a**) x =15.00 m and (**b**) x =17.50 m

In all the above plots with the tracer cross-section distribution, maximum value was lower in the Geometry n.2 and plume width was larger. The values of $\bar{C}_{max}$ and P_{mix} for both the simulated geometries and each location are listed in Table 2. Moreover, in Fig. 8 a plot with the values of the mixing degree P_{mix} along the geometry is presented. The growth of P_{mix} was quite similar and final values at the geometry outlet were 0.89 and 0.85 for Geometry n.1 and n.2, respectively.

Table 3 lists the values for both the modelled geometries of the transverse variance σ_y^2 obtained from Eqs. 18 and 19 and of the plume width L_{plume} calculated by applying Eq. 23. Plume width data in bold are those equal to the geometry width W, which means that the plume reached geometry walls.

In both geometries transverse variance was increasing with the distance from the injection point as expected. Moreover, both σ_y^2 and plume width values in the grid geometry were generally higher of about 25% than those in the Geometry n.1, reflecting an enhanced degree of turbulent mixing due to the grid.

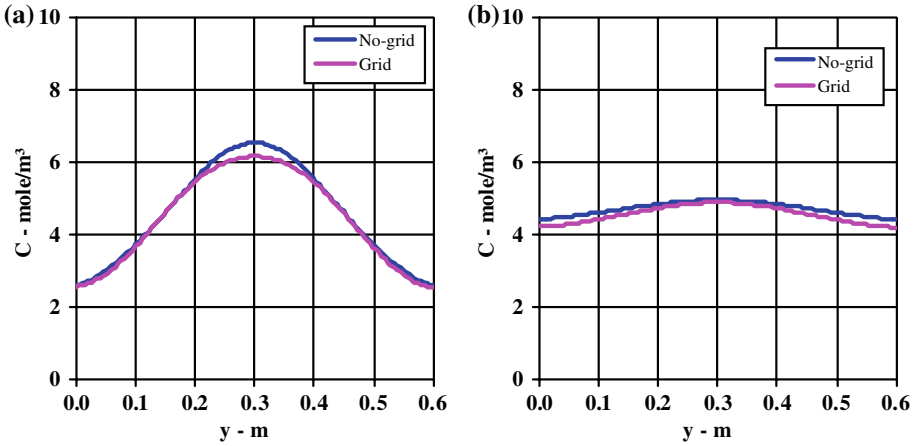


Fig. 7 Concentration transverse distributions at (**a**) x =25.00 m and (**b**) x =40.00 m

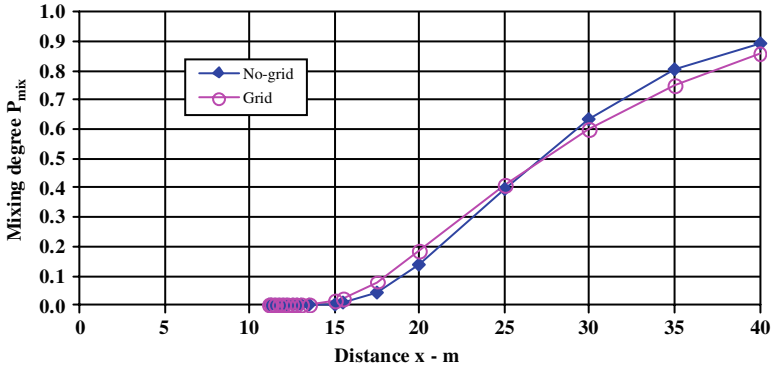


Fig. 8 Growth of mixing degree P_{mix} along the geometries

Table 4 lists the values of the transverse mixing coefficient D_{t-y} for both geometries deriving from the application of the methods of moments, that is Eq. 20, and from the transverse profile of turbulent kinematic viscosity ν_t, that is Eq. 12. Note that from reach n.13 and 11, in Geometry n.1 and n.2, respectively, the method of moments cannot be properly applied because the plume impinged the walls of the geometry.

The data listed in Table 4 were averaged to derive one value for each geometry. The values were weighted using the length of each reach. Also, both for the method of moments and the method based on turbulent kinematic viscosity, the average value was obtained by using only the values up to reach n.13 and n.11 for Geometry n.1 and n.2, respectively. A comparison showed that D_{t-y} values in the geometry with the grid, that is Geometry n.2, were generally larger. Second, the average value derived from method of moments for Geometry n.2 was about 50% larger than that for Geometry n.1. Third, the average D_{t-y} values derived from turbulent kinematic viscosity were lower than those obtained from Eq. 20.

Finally, the data of spatial variance of the tracer transverse profile were again used to derive an overall D_{t-y} for each geometry, as recently done by Rummel et al. [21]. Considering Eq. 22, these data were plotted against the travel time of the plume t. In these plots, that are presented in Fig. 9a and b, the transverse mixing coefficient D_{t-y} was taken

Table 3 Values of transverse variance and plume width

Reach	x (m)	Δx (m)	Geometry n.1—No-grid		Geometry n.2—Grid	
			σ_y^2 (m^2)	L_{plume} (m)	σ_y^2 (m^2)	L_{plume} (m)
1	11.125	–	0.0003	0.1109	0.0005	0.1361
2	11.25	0.125	0.0006	0.1462	0.0009	0.1815
3	11.50	0.25	0.0011	0.1966	0.0017	0.2471
4	11.75	0.25	0.0016	0.2370	0.0025	0.2975
5	12.00	0.25	0.0021	0.2672	0.0032	0.3378
6	12.25	0.25	0.0026	0.2975	0.0040	0.3782
7	12.50	0.25	0.0030	0.3277	0.0046	0.4084
8	12.75	0.25	0.0035	0.3479	0.0053	0.4336
9	13.00	0.25	0.0039	0.3681	0.0060	0.4588
10	13.50	0.50	0.0048	0.4084	0.0072	0.4992
11	15.00	1.50	0.0075	0.5193	0.0105	**0.6000**
12	15.50	0.50	0.0084	0.5597	0.0115	**0.6000**
13	17.50	2.00	0.0120	**0.6000**	0.0149	**0.6000**
14	20.00	2.50	0.0163	**0.6000**	0.0182	**0.6000**
15	25.00	5.00	0.0228	**0.6000**	0.0229	**0.6000**
16	30.00	5.00	0.0265	**0.6000**	0.0260	**0.6000**
17	35.00	5.00	0.0285	**0.6000**	0.0280	**0.6000**
18	40.00	5.00	0.0295	**0.6000**	0.0292	**0.6000**

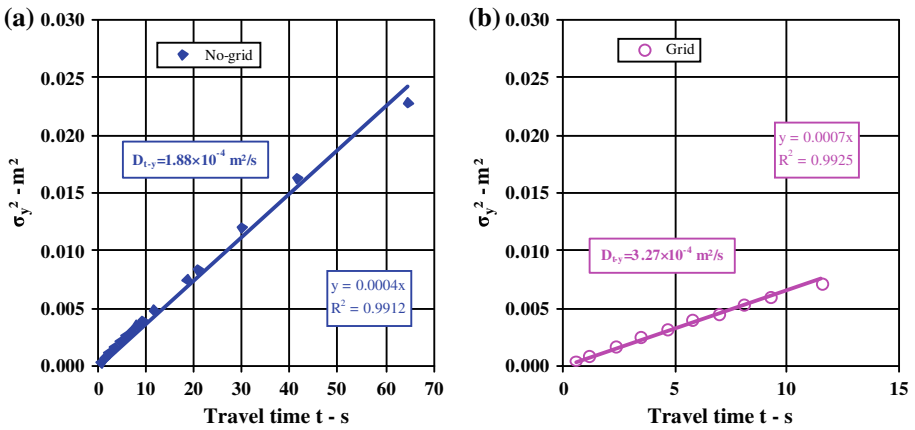


Fig. 9 Variance of the transverse concentration profile versus the travel time

as the fitting parameter and the line was forced to go through zero. The resulting values for D_{t-y} were 1.88×10^{-4} m^2/s and 3.27×10^{-4} m^2/s for the no-grid case and the grid-case, respectively. It should be noted that these values are quite similar to those obtained averaging the data derived from the method of moments, as listed in Table 4. After all, numerical results demonstrated that the introduction of a grid upstream the point of injection resulted in an enhanced mixing of the tracer and increase of about 70% in D_{t-y}.

Table 4 Values of transverse mixing coefficient D_{t-y}

Reach	x (m)	Δx (m)	Geometry n.1—No-grid		Geometry n.2—Grid	
			Eq. 19	Eq. 11	Eq. 19	Eq. 11
1	11.125	–	–	1.69×10^{-4}	–	2.83×10^{-4}
2	11.25	0.125	2.27×10^{-4}	1.69×10^{-4}	3.50×10^{-4}	2.78×10^{-4}
3	11.50	0.25	2.18×10^{-4}	1.68×10^{-4}	3.39×10^{-4}	2.69×10^{-4}
4	11.75	0.25	2.16×10^{-4}	1.68×10^{-4}	3.26×10^{-4}	2.61×10^{-4}
5	12.00	0.25	2.07×10^{-4}	1.68×10^{-4}	3.14×10^{-4}	2.55×10^{-4}
6	12.25	0.25	2.05×10^{-4}	1.68×10^{-4}	3.04×10^{-4}	2.50×10^{-4}
7	12.50	0.25	1.97×10^{-4}	1.68×10^{-4}	2.92×10^{-4}	2.45×10^{-4}
8	12.75	0.25	1.96×10^{-4}	1.68×10^{-4}	2.82×10^{-4}	2.40×10^{-4}
9	13.00	0.25	1.92×10^{-4}	1.69×10^{-4}	2.73×10^{-4}	2.36×10^{-4}
10	13.50	0.50	1.92×10^{-4}	1.69×10^{-4}	2.61×10^{-4}	2.29×10^{-4}
11	15.00	1.50	1.93×10^{-4}	1.71×10^{-4}	$\mathbf{2.33 \times 10^{-4}}$	2.10×10^{-4}
12	15.50	0.50	1.97×10^{-4}	1.72×10^{-4}	$\mathbf{2.08 \times 10^{-4}}$	2.05×10^{-4}
13	17.50	2.00	$\mathbf{1.98 \times 10^{-4}}$	1.78×10^{-4}	$\mathbf{1.80 \times 10^{-4}}$	1.88×10^{-4}
14	20.00	2.50	$\mathbf{1.85 \times 10^{-4}}$	1.88×10^{-4}	$\mathbf{1.40 \times 10^{-4}}$	1.74×10^{-4}
15	25.00	5.00	$\mathbf{1.40 \times 10^{-4}}$	2.10×10^{-4}	$\mathbf{9.88 \times 10^{-5}}$	1.66×10^{-4}
16	30.00	5.00	$\mathbf{8.10 \times 10^{-5}}$	2.24×10^{-4}	$\mathbf{6.49 \times 10^{-5}}$	1.76×10^{-4}
17	35.00	5.00	$\mathbf{4.27 \times 10^{-5}}$	2.31×10^{-4}	$\mathbf{4.17 \times 10^{-5}}$	1.97×10^{-4}
18	40.00	5.00	$\mathbf{2.16 \times 10^{-5}}$	2.32×10^{-4}	$\mathbf{2.51 \times 10^{-5}}$	2.16×10^{-4}
Average	–	–	1.99×10^{-4}	1.83×10^{-4}	2.98×10^{-4}	2.27×10^{-4}

The data in bold are those where the plume width reached geometry walls

6 Comparison between numerical results and experimental data: discussion

Numerical results were compared with the experimental data obtained by Lau and Krishnappan [10], that are summarized in Table 5. The data refer to different bed roughness. The values of transverse mixing coefficient were in the range from 0.335×10^{-4} m^2/s and 1.41×10^{-4} m^2/s. In the average, D_{t-y} was 0.815×10^{-4} m^2/s. Note that in the experimental data the aspect ratio W/h was not the same, ranging from 12.10 to 42.86. At larger values, the role of secondary currents should has been less important.

Comparison between the simulated values for only Geometry n.1 and the experimental data showed that numerical results tended to overestimate the transverse turbulent mixing coefficient. In particular, the numerical value, that is 1.88×10^{-4} m^2/s, exceeded for about 30% the maximum experimental value, that is 1.41×10^{-4} m^2/s. This is not unexpected since as above stated the assumption of isotropy can lead the k–ε model to predict large turbulent viscosity and, consequently, high turbulent rates of mixing. The first objective of this paper was just to identify how large would be this overestimation inherent to the hypothesis underlying k–ε model. On the other hand, also literature predictive equations for D_{t-y} are affected by large uncertainties. As already stated, in the classical equation proposed by Fischer et al. [20], i.e. Eq. 6, there is the possibility of an error of ±50%, which is larger than that associated with the above numerical simulation. After all, if only average velocities and

Table 5 Experimental data of Lau and Krishnappan [10]

Data	W (m)	h (m)	W/h	R_h (m)	u (m/s)	D_{t-y} (m^2/s)
1	0.60	0.042	14.46	0.036	0.337	1.40×10^{-4}
2	0.60	0.039	15.38	0.035	0.308	1.14×10^{-4}
3	0.60	0.042	14.29	0.037	0.179	7.40×10^{-5}
4	0.60	0.049	12.37	0.042	0.155	9.07×10^{-5}
5	0.60	0.050	12.10	0.043	0.302	1.41×10^{-4}
6	0.60	0.039	15.31	0.035	0.204	1.11×10^{-4}
7	0.60	0.033	18.07	0.030	0.196	8.78×10^{-5}
8	0.60	0.017	34.48	0.016	0.201	5.92×10^{-5}
9	0.60	0.014	41.96	0.014	0.208	6.01×10^{-5}
10	0.60	0.031	19.61	0.028	0.197	6.15×10^{-5}
11	0.60	0.023	26.09	0.021	0.204	5.02×10^{-5}
12	0.60	0.018	33.52	0.017	0.200	3.63×10^{-5}
13	0.60	0.014	42.86	0.013	0.201	3.35×10^{-5}
Average	–	–	–	–	–	0.815×10^{-4}

bulk mixing properties are needed, RANS-modeling of transverse turbulent mixing appears to be acceptable also for its easy implementation and economic computational effort.

7 Conclusions

Nowadays there is no comprehensive and accepted theory of transverse turbulent mixing and the prediction of its rate is mainly based upon the results of experimental works carried on in laboratory channels or in streams and rivers.

The first objective of this paper was to present the preliminary results of a numerical study undertaken to simulate the transverse mixing of a steady-state point source of a tracer in a two-dimensional rectangular geometry, which is expected to reproduce a shallow flow. An approach based on the Reynolds Averaged Navier–Stokes equations (RANS) was applied, where the closure problem was solved by using a turbulent viscosity concept. Particularly, the classical k–ε model was used since it is the easiest to implement and the most economical in computation. Also, recent studies on shallow flows have demonstrated that for simplified cases, where mean velocities and bulk mixing properties are needed, RANS-modeling of shallow flows is still appropriate. On the other hand, there was awareness about the limitations inherent to the hypothesis of isotropic turbulence assumed by the k–ε model.

Numerical values of the transverse turbulent rate of mixing were obtained by using the method of moments and the values of kinematic viscosity predicted by the model. Comparison with the experimental data collected by Lau and Krishnappan for a shallow flow in the same 2D geometry confirmed that k–ε model tended to overestimate the rate of transverse turbulent mixing since this model produces large turbulent viscosity. However, numerical value was of about 30% above the maximum experimental data. This may still be an acceptable result providing the larger uncertainties associated with literature predictive equations. So,

if only bulk mixing properties are needed, RANS-modeling of transverse turbulent mixing appears to be an acceptable tool.

The second objective was to assess the effect on transverse turbulent mixing of a grid formed by square elements located upstream the point of injection. A comparison between the numerical results with and without the grid demonstrated that the introduction of the grid caused an enhanced mixing of the tracer and an increase of about 70% in D_{t-y}.

Further research will be addressed to extend to 3D case this analysis based on the RANS approach.

Acknowledgements The author acknowledges the preliminary work carried out by the students G. Mazzeo, M. A. Maffei, L. Fiorentino and S. Orlando within the *Environmental Hydraulics* course at the University of Napoli *Federico II*.

References

1. Chapra SC (1997) Surface water quality modeling. McGraw-Hill, New-York, NY, USA, p 846. ISBN 0-07-011364-5
2. Rutherford JC (1994) River mixing. Wiley, Chichester, UK, p 348. ISBN 0-471-94282-0, 1–347
3. Nokes RI (1986) Problems in turbulent dispersion. PhD Thesis, University of Canterbury, Christchurch
4. Boxall JB, Guymer I (2003) Analysis and prediction of transverse mixing coefficients in natural channels. J Hydraul Eng 129(2):129–139. doi:10.1061/(ASCE)0733-9429(2003)129:2(129)
5. Ayyoubzadeh SA, Faramarz M, Mohammadi K (2004) Estimating longitudinal dispersion coefficient in rivers. In: Proceedings of the 6th international conference on hydroinformatics 2004, Singapore, June 21–24, 2004
6. Gualtieri C, Ciaravino G, Pulci Doria G (2006) Analysis of longitudinal dispersion equations in streams and rivers. In: Proceedings of the 7th international conference on hydroscience & engineering (ICHE 2006), Philadelphia (USA), September 10–13, 2006. ISBN 0977447405
7. Gualtieri C (2006) Environmental hydraulics. Lecture notes (in Italian), CUEN Editore, p 410. ISBN 88-7146-717-5
8. Cushman-Roisin B (2007) Environmental fate and transport. Lecture Notes, Thayer School of Engineering, Dartmouth College, NH, USA
9. Jirka GH (2001) Large scale flow structures and mixing processes in shallow flows. J Hydraul Res 39(6):567–573
10. Lau YL, Krishnappan BG (1977) Transverse dispersion in rectangular channel. J Hydraul Div 103(HY10):1173–1189
11. Multiphysics 3.4 User's Guide. ComSol AB, Sweden, 2008
12. Webel G, Schatzmann M (1984) Transverse mixing in open channel flow. J Hydraul Eng 110(4):423–435. doi:10.1061/(ASCE)0733-9429(1984)110:4(423)
13. Nokes RI, Wood IR (1988) Vertical and lateral turbulent dispersion: some experimental results. J Fluid Mech 187:373–394. doi:10.1017/S0022112088000473
14. Pope SB (2000) Turbulent flows. Cambridge University Press, Cambridge, UK
15. Schlichting H (1979) Boundary layer theory. McGraw Hill Book, New York, NY, USA, p 818. ISBN 0-07-055334-3
16. Henderson FM (1966) Open channel flow. Macmillan Series in Civil Engineering, New York, NY, USA, p 522
17. Chanson H (2004) Environmental hydraulics of open channel flows. Elsevier Butterworth-Heinemann, Burlington, MA, USA, p 430. ISBN 0-7506-6165-8
18. Chau KW (2000) Transverse mixing coefficient measurements in an open rectangular channel. Adv Environ Res 4:287–294. doi:10.1016/S1093-0191(00)00028-9
19. Gualtieri C, Mucherino C (2007) Transverse turbulent diffusion in straight rectangular channels. In: 5th international symposium on environmental hydraulics (ISEH 2007), Tempe, USA, December 4–7, 2007
20. Fischer HB, List EJ, Koh RCY, Imberger J, Brook NH (1979) Mixing in inland and coastal waters. Academic Press, New York, NY, USA, p 484
21. Rummel AC, Socolofsky SA, Carmer CF, Jirka GH (2005) Enhanced diffusion from a continuous point source in shallow free-surface flow with grid turbulence. Phys Fluids 17:075105-1-12

22. Socolofsky SA, Jirka GH (2004) Large scale flow structures and stability in shallow flows. J Environ Eng Sci 3(5):451–462. doi:10.1139/s04-032
23. Kundu PK, Cohen IM (2004) Fluid mechanics. Elsevier Academic Press, San Diego, CA, USA, p 760. ISBN 978-0-12-178253-5
24. Hanjalić K (2004) Closure models for incompressible turbulent flows. Lecture Notes at Von Kármán Institute, p 75
25. Ingham DB, Ma L (2005) Fundamental equations for CFD river flow simulations. In: Bates PD, Lane SN, Ferguson RI (eds) Computational fluid dynamics. Applications in environmental hydraulics. Wiley, Chichester, England, p 534. ISBN 978-0-470-84359-8
26. Knight DW, Wright NG, Morvan HP (2005) Guidelines for Applying Commercial CFD Software to Open Channel Flow. Report based on research work conducted under EPSRC Grants GR/R43716/01 and GR/R43723/01, p 31
27. Van Prooijen BC, Uijttewaal WSJ (2005) Horizontal mixing in shallow flows. In: Czernuszenko W, Rowinski P (eds) Water quality hazards and dispersion of pollutants. Springer Science+Business Inc., New York, NY, USA, p 250. ISBN 0-387-23321-0
28. Sotiropoulos F (2005) Introduction to statistical turbulence modeling for hydraulic engineering flows. In: Bates PD, Lane SN, Ferguson RI (eds) Computational fluid dynamics. applications in environmental hydraulics. Wiley, Chichester, England, p 534. ISBN 978-0-470-84359-8

Environ Fluid Mech (2010) 10:157–175
DOI 10.1007/s10652-009-9151-6

ORIGINAL ARTICLE

Assessing the parameterisation of the settling flux in a depth-integrated model of the fate of decaying and sinking particles, with application to fecal bacteria in the Scheldt Estuary

Anouk de Brauwere · Eric Deleersnijder

Received: 2 March 2009 / Accepted: 2 September 2009 / Published online: 18 September 2009

Abstract The fate of reactive tracers is often modelled by depth-averaged equations. When integrating the depth-resolved equations, it appears that the term describing the settling of particles is dependent on the concentration just above the bottom. Because in a depth-averaged framework this quantity is not available, the settling term needs to be parameterised. The most natural choice is to make the settling flux dependent on the average concentration. This approximation is acceptable if the water column is well mixed, but these conditions are not necessarily met in real applications. Therefore, this study aims at assessing and understanding the error made by using a depth-averaged model in a range of realistic conditions. For the definition of these conditions, typical values for the Scheldt Estuary and the Dutch-Belgian coast were taken. The realistic inspiration for the reactive tracer in this study is the fecal bacterium *Escherichia coli*, whose own dynamics are characterised by settling and gradual decay by mortality. In an attempt to understand the relative importance of several factors like settling, mortality, mixing and stratification on the error made by a depth-averaged approach, a number of simplified test cases were investigated. It follows that, as expected, the error is acceptable if the situation is mixing-dominated. However, the effect of mortality and stratification was less obvious in advance. For instance, it appeared that errors can also be significant if settling and mortality have the same characteristic timescales. Stratification often has the effect to increase the error made by the depth-averaged model.

Keywords Settling · Mixing · Mortality · Decay · Pycnocline · Residence time · Depth-integrated model · Reactive tracer

A. de Brauwere · E. Deleersnijder
Centre for Systems Engineering and Applied Mechanics(CESAME), Université catholique de Louvain, 4 Avenue G.Lemaître, 1348 Louvain-la-Neuve, Belgium

A. de Brauwere (✉)
Analytical and Environmental Chemistry, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
e-mail: adebrauw@vub.ac.be

1 Introduction

In the present study, the error made by vertically averaging the advection-diffusion-reaction equations for decaying and sinking particles is investigated. In particular, we are interested in modelling the fate of fecal bacteria in the Scheldt Estuary by means of a two-dimensional depth-averaged model and would like to estimate the potential error introduced by the depth-integration. In order to enable statements about the error's dependence on factors like mixing, settling, decay and stratification, a number of simplified case studies are studied.

Vertically integrated models have been regularly used in the past to describe estuaries (e.g. [15,24]), including the Scheldt Estuary ([23] (2D); [14] (1D)). Usually the depth-integration is motivated by the fact that the water column is rather well mixed. This assumption is generally satisfied under conditions of strong currents, shallow waters and/or negligible stratification, but these conditions are not necessarily verified for the whole domain. Locally different conditions may be experienced and thus the assumption may be locally violated.

Fecal bacteria contamination in natural waters is generally monitored to assess the sanitary risk associated with various water utilisations (e.g. bathing, shellfish harvesting, production of drinking water). Although the measured indicator organisms may not be harmful themselves, they indicate the potential presence of other fecal pathogens. Vertically integrated models have already been used to simulate fecal bacteria in a few studies, e.g. considering a coastal lagoon [21] or natural streams [17]. Fecal bacteria enter the water column solely through external sources. Depending on the area (often determined by the population density), these are primarily point sources (e.g. waste water treatment plant outlets) or diffuse sources (e.g. water runoff from manure covered land). By comparison with similar basins, it is expected that in the Scheldt point sources are the most important input type of fecal bacteria into the domain [13,20]. After entering the domain, the bacteria are transported by turbulent diffusion and advection, and in addition they gradually disappear from the water column due to mortality and sedimentation. Mortality can be modelled by a first order decay process [16,21], which remains appropriate for vertically integrated variables.

The representation of the settling is less straightforward. First there is the complexity of the in situ reality: the bacteria are generally thought to be either free in suspension or attached to particles [12,17]. The strength (or reversibility) of the linkage between bacteria and a particle can vary but is difficult to predict. It is clear that free or attached bacteria will behave differently. For instance, thanks to the protective effect of the particles, attached bacteria are reported to survive longer, and, not surprisingly, they sink faster than free bacteria [16]. However, the two (or more) classes of bacteria are often not modelled explicitly, because the individual behaviours are not quantitatively known and/or the interaction between free and attached bacteria has proven to be even more difficult to describe quantitatively [16,17]. Instead, "average" bacteria are considered, exhibiting a mortality and sedimentation behaviour representing an average between 100% free and 100% attached bacteria. Besides this natural complexity, representation of the settling also presents a more formal difficulty. In a depth-resolved model the settling flux to the sea floor is naturally proportional to the concentration just above the bottom. However, with depth-averaged equations the settling flux cannot be represented like this anymore and must be parameterised. The most popular and intuitive parameterisation is proportional to the average concentration, but the magnitude of the error introduced by this approximation is not clear. If vertical mixing is much more efficient than settling, the water column will be virtually homogeneous and hence the depth-averaged approximation will be perfectly adequate.

To enable more quantitative statements, it is useful to introduce the ratio of the mixing to the settling timescales, which is a dimensionless parameter known as the Peclet number, $Pe = T_{\text{mixing}}/T_{\text{settling}} = WH/K$, with W, H and K typical values of respectively the sinking velocity, the water height and the vertical eddy diffusivity. So, the error made by vertical integration will be small if $Pe \ll 1$. Yet, when considering typical ranges of $W(10^{-6}$–10^{-5}m/s), H(1–100 m) and K (0.001–1m^2/s) in the Scheldt Estuary and Belgian coastal zone, it appears that the Peclet number may range from 10^{-6} to values as high as 10, i.e. a value >1. At this point, one might object that H and K are not independent of eachother, and therefore the range of possible values of Pe is actually narrower than that just suggested. Nonetheless, values of the order of unity or even larger cannot be ruled out a priori.

Apart from the relative importance of mixing vs. settling, there is a second factor that may influence the reliability of the depth-integrated approximation: the magnitude of the mortality rate. For instance, if the mortality rate is very high, the bacteria die almost instantly and consequently the average concentration tends to zero independently of other factors—and so depth-integration again becomes acceptable—although admittedly there is not much relevance in studying fecal bacteria when they are virtually inexistent. In order to assess the simultaneous effect of mortality, settling and mixing, the outputs of a simple numerical 1D model will be compared to those of its depth-integrated counterpart.

A last issue has been considered: the presence or absence of a pycnocline. Due to intense warming/cooling, or precipitation/evaporation density gradients may appear in the water column. This gradient represents a physical barrier impeding vertical mixing, and thus significantly modifies the vertical diffusivity profile. The presence or absence of this feature may thus affect the acceptability of a depth-averaged approach. It will also be investigated how this stratification effect is further influenced by the vertical position of the point source injecting bacteria in the domain.

Finally, we should probably also admit that a number of potential factors are not taken into account in the analysis. This is a deliberate choice to enable to focus on the selected key factors (mortality, settling, mixing and stratification).

For example, processes of bottom sediment resuspension may also have an impact on the reliability of a depth-averaged model. As these processes are not explicitly taken into account, the results are valid if

(i) resuspension is negligible, i.e. their frequency of occurrence is much smaller than the typical mortality constant. This situation has been reported in some studies (e.g. [21]);
(ii) settling in fact represents the *net* vertical movements, i.e. the balance between real settling and resuspension (only for stationary situations);
(iii) the modelled concentration of fecal bacteria is not interpreted as the total concentration but as the concentration of bacteria which haven't touched the bottom yet. This last interpretation is always valid and does not affect the relevance of this study, because its focus is not on the resuspension but on the parameterisation of the settling flux.

Another simplification is the negligence of horizontal processes; in other words horizontal homogeneity is assumed, in order to concentrate on the vertical processes which are the most influenced by depth-averaging, and reduce the dimensionality of the test cases.

The remainder of this study is structured as follows: in the next section the models are presented and the difference between the depth-resolved and depth-integrated equations are clearly shown. In Section 3 the concept of residence time is introduced. Section 4 contains the results of a number of case studies, focusing on the error made by the depth-averaged model compared to the output of the depth-resolved approach. The final section summarises these results and gives some concluding remarks.

Table 1 Natural ranges of the relevant dimensional and dimensionless variables

	Range of values	References
Dimensional variables		
w (m/s)	10^{-6}–10^{-5}	[1,20,21]
h (m)	1–100	Bathymetric data of Scheldt and Belgian coastal zone
$\lambda(s^{-1})$	10^{-6}–10^{-4}	[1,11,20]
$\bar{\kappa}(m^2/s)$	10^{-4}–1	[10]; Appendix 1
Dimensionless variables		
$\tilde{z} = z/h$	0–1	
$Pe = wh/\bar{\kappa}$	10^{-6}–10	
$\tilde{\lambda} = \lambda/(w/h)$	0.1–10^4	

2 Models

As introduced above, only the vertical processes will be considered by assuming horizontal homogeneity. Taking into account diffusion, mortality, settling and a source term, the concentration $C(t, z)$ of fecal bacteria (bacteria/m^3) is the solution of the following partial differential problem:

$$\frac{\partial C}{\partial t} = q - \lambda C + \frac{\partial}{\partial z}\left(wC + \kappa \frac{\partial C}{\partial z}\right) \tag{1}$$

with initial and boundary conditions:

$$C(0, z) = C_0(z); \quad \left[wC + \kappa \frac{\partial C}{\partial z}\right]_{z=h} = 0; \quad \left[\kappa \frac{\partial C}{\partial z}\right]_{z=0} = 0. \tag{2}$$

z is the vertical coordinate (m), which is zero at the bottom and equal to h at the surface; $\kappa(z)(\geq 0)$ is the eddy diffusivity (m^2/s); $q(t, z)$ is the source of bacteria (bacteria/m^3s); λ is the mortality rate (s^{-1}); and $w(\geq 0)$ is the settling velocity (m/s). Table 1 summarises the natural ranges of values of the variables. To remain as general as possible, the model equations in this section are given for general source terms $q(t, z)$, but for the later assessments only point sources will be considered.

This study will focus on the depth-averaged concentration

$$\bar{C}(t) = \frac{1}{h}\int_0^h C(t, z)\mathrm{d}z. \tag{3}$$

Its associated model, the depth-integrated counterpart of model (1)–(2), is found by first integrating (1)–(2) and assuming a constant sinking velocity w:

$$\frac{\mathrm{d}\bar{C}}{\mathrm{d}t} = \bar{q} - \lambda\bar{C} - \frac{w}{h}C(t, 0), \tag{4}$$

with

$$\bar{q}(t) = \frac{1}{h}\int_0^h q(t, z)\mathrm{d}z \tag{5}$$

and initial condition

$$\bar{C}(0) = \bar{C}_0 = \frac{1}{h}\int_0^h C_0(z)\mathrm{d}z. \tag{6}$$

Equation 4 is the correct depth-integrated equation, but because it contains the term $wC(t,0)/h$, it cannot be solved when working only with depth-integrated variables. Therefore, one needs to find a workable parameterisation for this sedimentation term. Assuming that the water column is well-mixed, it is often posed that $C(t,0) \approx \bar{C}(t)$, giving the actual depth-integrated (0D) model equation:

$$\frac{\partial \bar{C}'}{\partial t} = \bar{q} - \left(\lambda + \frac{w}{h}\right)\bar{C}' \tag{7}$$

with initial conditions

$$\bar{C}'(0) = \bar{C}_0; \tag{8}$$

and the prime is associated with variables of the depth-averaged model. Expressions (4) and (7) look very similar; indeed the first two terms (related to the source and mortality) are identical. The only difference lies in the expression of the last term which represents the settling flux. Consistent with intuition, the settling flux in the 1D model depends on the concentration at the bottom, but this information is not available in a depth-averaged model. Therefore, the settling flux in the 0D model is parameterised so that it depends on the average concentration. This seemingly minor discrepancy may significantly impact the 0D solutions, especially if the Peclet number is high, i.e. when settling proceeds faster than vertical mixing.

For further simplification of the equations, it is convenient to introduce dimensionless variables, normalised with typical values. For the one-dimensional model, we take

$$\tilde{t} = \frac{t}{h/w}, \quad \tilde{z} = \frac{z}{h}, \quad \tilde{C} = \frac{C}{\hat{C}}, \quad \tilde{q} = \frac{q}{Q}, \quad \tilde{\lambda} = \frac{\lambda}{w/h}, \quad \tilde{\kappa} = \frac{\kappa}{\bar{\kappa}}, \tag{9}$$

with

$$\bar{\kappa} = \frac{1}{h}\int_0^h \kappa(z)\ \mathrm{d}z, \tag{10}$$

and the Peclet number

$$Pe = \frac{wh}{\bar{\kappa}}. \tag{11}$$

The characteristic concentration $\hat{C}$ may be taken to be Qh/w, where Q is the typical value of the source term q. If there is no source, then it is probably natural to set $\hat{C} = \bar{C}_0$. For this normalisation, it is (again) assumed that the settling velocity w is constant. We can now redefine the ranges within which the dimensionless variables can vary in natural riverine to coastal conditions (fecal contamination is not expected in open sea). For clarity, the ranges are summarised in Table 1. The two quantities that will be varied in particular, are Pe and $\tilde{\lambda}$, the first expressing the relative importance of the settling versus mixing, while the second balances the importance of the mortality process against the sedimentation.

Unless stated otherwise, from now on only dimensionless variables will be used. Therefore, in order to simplify the notations, the tildes will be dropped. Accordingly, the one-dimensional partial differential problem to be solved reads

$$\begin{cases} \frac{\partial C}{\partial t} = q - \lambda C + \frac{\partial}{\partial z}\left(C + \frac{\kappa}{Pe}\frac{\partial C}{\partial z}\right) \\ C(0,z) = C_0(z),\ \left[C + \frac{\kappa}{Pe}\frac{\partial C}{\partial z}\right]_{z=1} = 0,\ \left[\frac{\kappa}{Pe}\frac{\partial C}{\partial z}\right]_{z=0} = 0 \end{cases}, \tag{12}$$

and the depth-averaged concentration can simply be computed as

$$\bar{C}(t) = \int_0^1 C(t,z)\,\mathrm{d}z. \tag{13}$$

In dimensionless variables, the depth-integrated problem reads

$$\begin{cases} \frac{\mathrm{d}\bar{C}'}{\mathrm{d}t} = \bar{q} - (\lambda+1)\bar{C}' \\ \bar{C}'(0) = \bar{C}(0) \end{cases}. \tag{14}$$

One of the objectives of the current study is to assess the difference between $\bar{C}(t)$ and $\bar{C}'(t)$. For the depth-averaged model approach to be acceptable, the difference should obviously be small, but it is not straightforward to make a general statement about the magnitude of this difference, as there is no general closed form solution of (12). Nevertheless, as explicitly shown above for the dimensional equations (4) and (7), we can compare the mass balance for the depth-averaged concentrations, indicating that the difference between the two models lies in the settling term. The actual, quantitative error made by the 0D model compared to the 1D model will be assessed in the Results section for a number of particular cases.

3 Residence time

3.1 Definition

In order to compare the depth-resolved and depth-averaged models' performances some "diagnostic variables" will be analysed and compared. Natural choices are concentrations and settling fluxes, but in addition to those we propose to examine the mean residence time as a diagnosis of the 1D and 0D models. The residence time of a particle in a water body is defined as the time taken by this particle to leave the water body [2,8–10,22]. Every particle injected into the domain has his own residence time, i.e. time it needs to leave the water column, either by dying or by settling. Therefore, each setup is actually associated with a distribution of elemental residence times, which are generally averaged to find the actual (mean) residence time.

A special feature of the mean residence time is that is applies to both an impulse source and the stationary case (exactly the two special cases that will be considered in more detail), because the stationary case is in essence a repetition of identical impulse injections at every time step. In the next section, some expressions will be derived for a special case without mortality.

3.2 Special case: no mortality

In this section mortality processes are neglected. While they surely represent an extreme case, these further simplified expressions present the advantage of allowing a general statement about the residence times computed by the 0D model versus 1D model. Neglecting mortality terms, the depth-resolved model equations in dimensional variables (1)–(2) and the depth-integrated model (7)–(8) are simplified. Note indeed that in this section we will again

use dimensional variables, but this does not affect the interpretation of the results as these are stated in relative terms—and thus equally hold for the dimensionless variables.

The mean time τ spent by the initially present fecal bacteria in the water column—until they settle on the floor of the water column—is

$$\tau^{(\prime)} = \frac{1}{\bar{C}_0^{(\prime)}} \int_0^\infty \bar{C}^{(\prime)}(t)\, \mathrm{d}t, \tag{15}$$

where the prime between brackets signifies that the expression holds both for the depth-resolved and depth-integrated variables. This timescale is generally referred to as the residence time. It is important to note that this residence time only considers the bacteria present in the water column at $t = 0$. Therefore, it is not influenced by any bacteria entering the domain at a later time.

To compute the residence time according to the 0D model, Eqs. 7 and 8 without the mortality and source terms are solved, giving

$$\bar{C}'(t) = \bar{C}'_0 e^{-wt/h}. \tag{16}$$

By substituting (16) into (15), the residence time according to the depth-integrated approach is readily obtained:

$$\tau' = \frac{h}{w}. \tag{17}$$

To derive the depth-resolved timescale τ—without explicitly knowing $\bar{C}(t)$—we must have recourse to the theory developed by [10]. Accordingly, we have

$$\tau = \frac{\int_0^h C_0(z)\theta(z)\, \mathrm{d}z}{\int_0^h C_0(z)\, \mathrm{d}z}, \tag{18}$$

where

$$\theta(z) = \frac{z}{w} + \frac{1}{w} \int_z^h \exp\left[-w \int_z^\xi [\kappa(\zeta)]^{-1} \mathrm{d}\zeta \right] \mathrm{d}\xi \tag{19}$$

is the residence time associated to bacteria having an initial concentration $C(0, z) = \delta(z)$, where $\delta(z)$ is a Dirac impulse located at height z. From (19) it follows that

$$\frac{z}{w} \leq \theta(z) \leq \frac{h}{w}. \tag{20}$$

This implies that

$$\alpha\tau' \leq \tau \leq \tau' \tag{21}$$

with

$$\alpha = \frac{1}{h^2 \bar{C}_0} \int_0^h z C_0(z)\, \mathrm{d}z \leq 1. \tag{22}$$

Thus, having recourse to the depth-integrated model always causes an overestimation of the time spent by bacteria in the water column, at least in this simplified case where mortality can be neglected. Notice that this result was obtained without knowing the actual evolution

Table 2 Overview of the equations for the two considered test cases

		Single initial pulse ("Dirac case")	Stationary situation
1D	Equation	$\frac{\partial C}{\partial t} = -\lambda C + \frac{\partial}{\partial z}\left(C + \frac{\kappa}{Pe}\frac{\partial C}{\partial z}\right)$	$\frac{d}{dz}\left(\frac{\kappa}{Pe}\frac{\mathrm{d}C}{\mathrm{d}z} + C\right) - \lambda C + q = 0$ with $q(z) = \delta(z-1)$
	Initial conditions	$C(0, z) = \delta(z-1)$	Not applicable
	Boundary conditions	$\left[\frac{\kappa}{Pe}\frac{\partial C}{\partial z}\right]_{z=0} = 0$ $\left[C + \frac{\kappa}{Pe}\frac{\partial C}{\partial z}\right]_{z=1} = 0$	$\left[\frac{\kappa}{Pe}\frac{\partial C}{\partial z}\right]_{z=0} = 0$ $\left[C + \frac{\kappa}{Pe}\frac{\partial C}{\partial z}\right]_{z=1} = 0$
0D	Equation	$\frac{\mathrm{d}\bar{C}}{\mathrm{d}t} = -(\lambda+1)\bar{C}$ $\Rightarrow \bar{C}(t) = \bar{C}(0)e^{-(\lambda+1)t}$	$\bar{q} - (\lambda+1)\bar{C}' = 0$ with $\bar{q} = \int_0^1 q(z)\,\mathrm{d}z$ $= 1 \Rightarrow \bar{C}' = (\lambda+1)^{-1}$
	Initial conditions	$\bar{C}(0) = 1$	Not applicable
	Boundary conditions	Not applicable	Not applicable

of the concentration profile, which is just another illustration of the power of the residence time theory. The price to pay is that, although we know that τ' is larger than τ, we don't know how much larger. In the next section, the 1D and 0D models will be compared in a more quantitative manner.

4 Results

4.1 Tested cases

The purpose of this study is to assess the validity of a depth-averaged model for advection-diffusion-decay-settling of fecal bacteria in a wide range of conditions potentially encountered in a real application (the Scheldt Estuary and Belgian coastal zone are taken as focus). To reduce the number of factors and thus keep this assessment feasible, the "full" and integrated model are simplified in several aspects. First, horizontal homogeneity is assumed such that only the vertical processes are of interest and we can speak of "1D" for the depth-resolved and "0D" for the depth-integrated model. Further, two particular cases considering the injection of the bacteria will be considered: (i) one initial Dirac pulse, located at the surface unless stated otherwise; and (ii) a source continuously discharging into the domain at a constant rate $\bar{q}$, and again located at the surface. The first case offers some advantages in terms of interpretation: the behaviour of a single concentration peak is easier to understand than a situation where bacteria are injected continuously. Moreover, this situation is the basis for the residence time definition. The second case is still a simplification of reality, but already much closer to it. Indeed, point sources of fecal bacteria are often modelled as constant-flux inputs. Furthermore, the situation further simplifies to a stationary situation, with the advantage that no temporal changes must be considered. The two tested cases are summarised in Table 2.

These equations are solved numerically using a Patankar-like discretisation scheme for the 0D model [3,4,18] and a scheme inspired by [6] for the 1D model. More details on these schemes are given in Appendix 2.

So, the 0D model results will be compared to the averaged 1D outcomes for both the initial Dirac pulse and the stationary situation, and for Pe and λ values varying within the ranges given in Table 1. In addition, one last factor has been investigated: the presence or

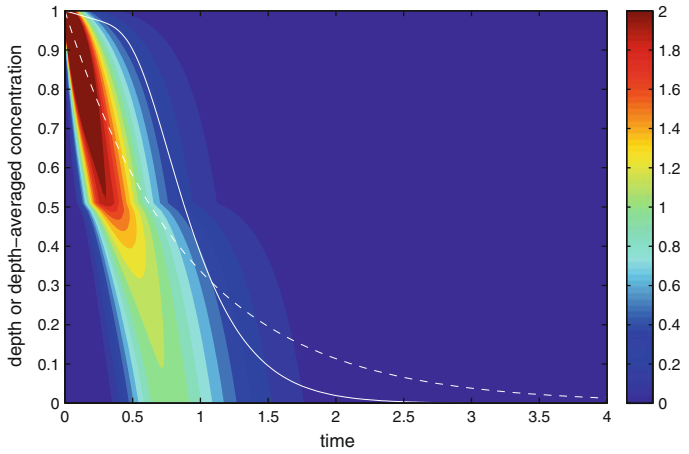


Fig. 1 Example of the time evolution of depth-resolved (contours) and depth-averaged concentrations [*full line* $\bar{C}$ (1D model), *dotted line* $\bar{C}'$ (0D model)] for the Dirac source case. As both depth and average concentration range from 0 to 1, both can be shown on the same vertical axis. Simulated conditions: $Pe = 10$, $\lambda = 0.1$ and there is a pycnocline at mid-depth

absence of stratification. Without stratification, the vertical diffusivity $\kappa(z)$ is parameterised as a parabolic function of z, being zero at the bottom ($z = 0$) and surface ($z = 1$) and according to (9)–(10) it should integrate to one:

$$\kappa(z) = 6z(1-z) \quad \text{in case of no stratification.} \tag{23}$$

In the presence of a pycnocline, vertical mixing is highly hampered in its vicinity. Therefore, the vertical diffusivity profile should exhibit a significant decrease at the height of the pycnocline. For these test cases, a pycnocline at mid-height was considered and the associated κ profile is composed of two parabolic curves joining at the level of the pycnocline such that $\kappa(h = 0.5) = 0$:

$$\kappa(z) = \begin{cases} 24z(0.5-z) & \text{if} z \leq 0.5 \\ 24(z-0.5)(1-z) & \text{if } z > 0.5 \end{cases} \quad \text{in case of stratification.} \tag{24}$$

4.2 Concentrations

To illustrate the difference between the 0D and 1D modelled concentrations, an example for the Dirac source is shown in Fig. 1. For the 1D model both the depth-resolved and the depth-averaged concentration profiles (contours) are shown as a function of time. The latter can be compared to the depth-averaged concentration from the 0D model. The profiles represent the following conditions: $Pe = 10$, $\lambda = 0.1$ and a stratified water column (see (24)). The stratification is not really visible in the vertically integrated concentration profiles, whereas it is clear in the distorted contour lines representing the depth-resolved concentration evolution. Further, it can be seen that (due to the high Pe) the two depth-averaged curves do not overlap. The 0D model will first underestimate the true (1D) average concentration and then overestimate it. While the 0D model can only represent the concentration as an exponential decay, the 1D model nicely shows that the concentration first decreases mainly due to mortality, as it takes some time before the bulk of the bacteria reach the bottom. When this happens the 1D concentration starts decreasing even faster.

With $Pe = 10$, the settling is rather efficient in this example, implying that the injected bacteria are hardly diffused and consequently the concentration spot remains relatively compact. For decreasing Pe, the importance of diffusion increases, and thus the second decrease in $\bar{C}$, coinciding with the particles reaching the floor, starts earlier. Before that moment, mortality is the main disappearance process and the slope of the initial decrease in $\bar{C}$ is thus only related to the mortality rate. For higher values of λ, this initial slope becomes steeper.

To summarise the error made by the 0D model compared to the depth-averaged 1D model, the root mean squared relative errors are computed for values of λ and Pe varying within realistic ranges (Table 1). Very small values for Pe and very large values for λ are not shown in the results, as these represent trivial situations. Indeed, for very small Pe values (<0.1), the hypothesis of a well-mixed water column is satisfied and thus the depth-averaged model is perfectly valid (errors tending to zero). On the other hand, very large mortality rates ($\lambda >$ 10) imply an almost instant disappearance of fecal bacteria, where again both models will provide identical results (average concentrations tending to zero). The resulting error maps are shown in Fig. 2 for the Dirac source and the stationary situation, and for the unstratified and stratified situations. The relative error can be as high as 29% for the Dirac source and 19% for the stationary situation. The maximal error is higher in the case of the Dirac source, probably because this situation is by its nature less homogeneous than the stationary situation, and therefore further from the assumption needed for the 0D model. For the stationary situation always $\bar{C}' < \bar{C}$ (not visible in Fig. 2 only showing rms errors).

Both Dirac and stationary cases show errors increasing with Pe, i.e. increasing importance of settling compared to vertical mixing. This is rather intuitive since in this case indeed, the water column is the least homogeneous, and thus a depth-averaged view is the least acceptable. Curiously however, the errors for the Dirac and stationary cases are not maximal for the same λ. For the Dirac source, the error increases with decreasing mortality, being in line with the interpretation of the Pe dependence: both high Pe and low λ result in least homogeneous water columns. On the other hand, in the stationary situation, the error is maximal for a given, intermediate mortality rate of about 1.6. In other words, both models produce most different average concentrations when the characteristic mortality time is of the same order as the typical sedimentation time (cf. (9)). If the mortality process becomes more efficient than settling (increasing λ), the average concentrations from both models simply tend to zero and their difference too. For the other extreme, mortality becoming less efficient than settling (decreasing λ), the average concentrations of both models will again tend to the same value, namely one, the initial concentration. It is only for intermediate values of λ (of the order of 1) that the 1D and 0D models provide significantly different results.

Figure 2 also illustrates the impact of a pycnocline. Although the overall patterns and maximal values are not influenced by the stratification, it seems that the errors decrease more slowly for decreasing Pe. This is not surprising because the stratifcation has the effect to locally decrease the mixing efficiency.

4.3 Settling fluxes

Remember that in case of a Dirac source the difference between the 1D-depth averaged and 0D concentration lies solely in the settling term (Sect. 2.1). In other words, the plots in Fig. 2 are very similar to plots of absolute root mean square errors for the settling flux (not shown). However, when plotting the relative errors (Fig. 3), the pattern is quite different. The errors are still maximal for high Pe, but also for high λ. The relative errors become huge for very large mortality rates (notice the exponential color scale), because in these conditions the 1D

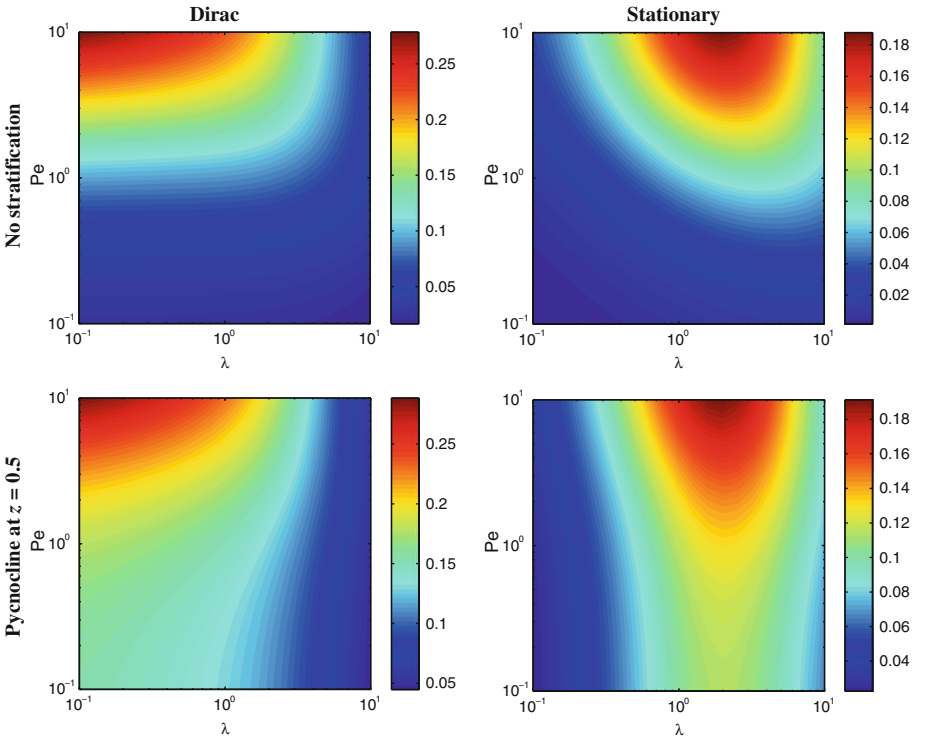


Fig. 2 Relative rms errors between the depth-averaged concentrations as modelled by the 0D model and 1D model. *Left column* Dirac source, *right column* stationary situation, *top row* no stratification, *bottom row* pycnocline at mid-depth. Discretisation step used for the Dirac pulse case was $\Delta t = \Delta z = 1/50$ (Appendix 2)

fluxes are virtually zero. Indeed, for the 1D model most bacteria are already dead before they can actually reach the bottom, while the 0D model still computes a finite flux because it is based on the average concentration.

In Fig. 4 the modelled time evolutions of the flux are shown for some values of *Pe* and λ in the case of a Dirac source at the surface and no stratification. These examples confirm that when the mortality rate increases, the 1D settling flux amplitude decreases, and eventually the settling peak will disappear totally (not shown). Conversely, the initial 0D fluxes are always 1, resulting in a finite 0D flux in all conditions. Examining the *Pe* dependency, Fig. 4 illustrates that the 0D and 1D flux are very similar for small *Pe*. In those cases the time-integrated error made by the 0D model will be small, although there will always be an error at the initial time where the 1D flux is zero and the 0D flux is one.

4.4 Residence time

In this section, the residence time computed from the 1D and 0D models will be compared. As introduced above, the residence time provides an additional diagnosis to interpret the outputs of a model. Furthermore, the definition of the residence time allowed to derive that in the simplified case with no mortality the 0D model will always overestimate the time spent

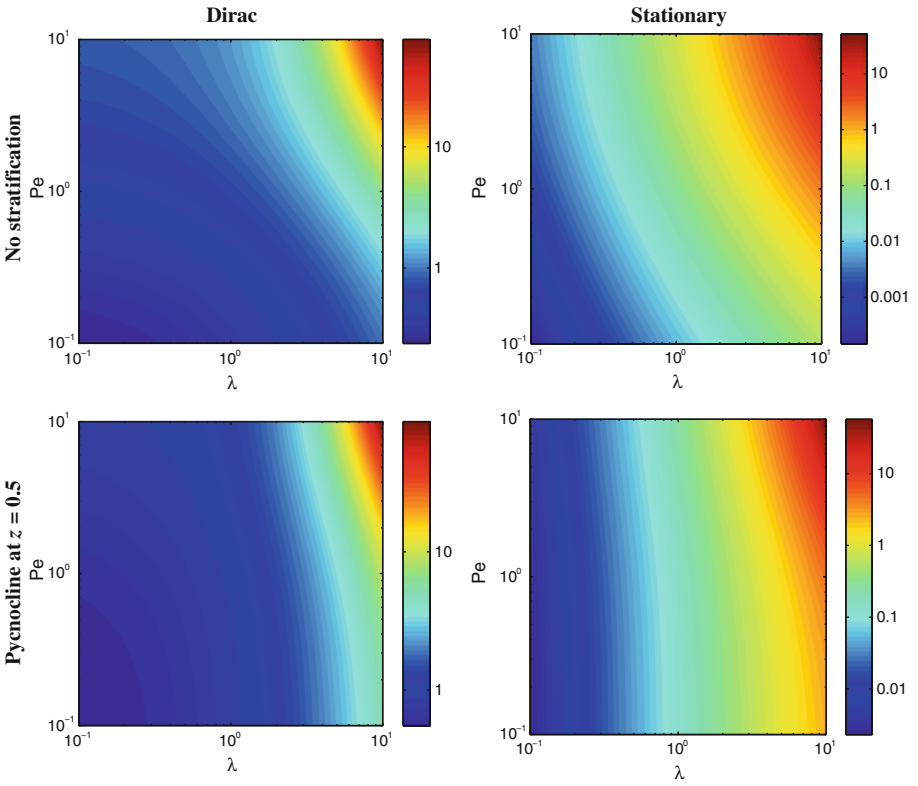


Fig. 3 Relative rms errors between the depth-averaged fluxes as modelled by the 0D model and 1D model. Notice the logarithmic error scale. *Left column* Dirac source, *right column* stationary situation, *top row* no stratification, *bottom row* pycnocline at mid-depth

in the water column by the bacteria (Section 3.2). This result can now be numerically verified and the existence of a similar relation can be saught for the more general case with mortality.

Figure 5 shows the evolution of the relative error as a function of *Pe*, in the special case that mortality is negligible (see Section 3.2). First, note that all relative errors (computed as $(\theta_{0D} - \theta_{1D})/\theta_{1D}$) are always positive, confirming the theoretical conclusion of Section 3.2 that the 0D model would always overestimate the residence time. Next, remember that the magnitude of the error depends on the location of the initial pulse of bacteria (Eqs. 19–22). Indeed, for the case where bacteria are injected at the surface, $\tau = \tau'$. In Fig. 5 it can be seen that for decreasing heights of the initial pulse, the difference between the two models increases. This behaviour is similar for the case with and without stratification. These two cases also converge to the same error for increasing *Pe*, confirming that stratification does not introduce a notable effect if diffusion is negligible. This error appears to be finite and equal to $1/z - 1$, which corresponds to what is expected from (17) and (20).

Even for small *Pe* stratification does not largely affect the residence time error, at least not when the source is located at $z > 0.5$ (top panel of Fig. 5). For these conditions the errors tend to zero, which reiterates the previous observations that depth-averaging is acceptable when the setup is diffusion-dominated. The effect of stratification can only be noticed by a slightly different slope of the error curve for intermediate *Pe* values, but it converges to the same

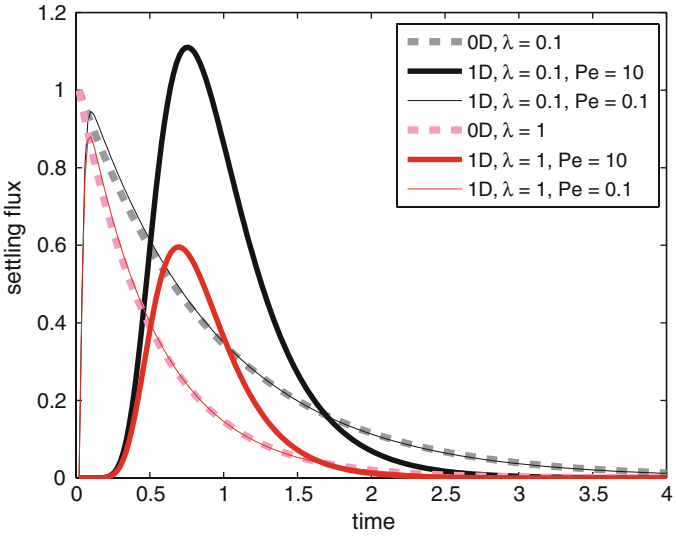


Fig. 4 Time evolution of settling fluxes for some values of *Pe* and λ in the case of a Dirac source at the surface and no stratification. *Black curves* λ = 0.1, *red curves* λ = 1. *Dashed lines* 0D model (independent of *Pe*). *Full lines* 1D model with *Pe* = 0.1 (*thin line*) and *Pe* = 10 (*thick line*)

errors for small and large *Pe*. On the other hand, if the source is located at $z \geq 0.5$, the error for small *Pe* converges to 1 in case of stratification, while it still tends to zero for the unstratified case (lower panel of Fig. 5). Stratification does not affect the residence time computed by the 0D model. Instead, these results show that when the bacteria are injected below the pycnocline, the 1D residence is significantly lowered, especially in the cases dominated by diffusive mixing (low *Pe*). This observation confirms the common sense that the pycnocline will avoid the bacteria to diffuse upward through it, and thus the bacteria are confined to a smaller space and will more rapidly hit the bottom.

To summarize, the *maximal* errors are obtained for large *Pe* and increase with decreasing height of the source location, but are not influenced by the presence of a pycnocline. However, the *minimal* errors (associated with low *Pe*) are significantly affected if stratification is present, but only if the bacteria are injected below this pycnocline: in that case the error amounts to 100% whereas in the other cases, the errors tend to zero.

After investigating this special case without mortality, the effect of mortality is now included too. For this case, no theoretical results are available, so the behaviour of the errors is difficult to predict. However, it will be interesting to compare these results with those obtained without mortality. The results are shown in Fig. 6, illustrating again the importance of the source position. Indeed, the errors are relatively small (although not zero this time) when the bacteria are injected at the surface. This time, however, the errors are always negative, meaning that the 0D model systematically underestimates the residence time. The overall contour patterns are very similar to those of stationary concentration errors (Fig. 2, right column), with a maximal error for $\lambda \approx 1.6$.

When the bacteria are injected at $z = 0.75$, both positive and negative errors are observed, with a maximal positive error for high *Pe* and low λ, and largest negative error again for high *Pe* but for a specific value of λ of about 4. In the absence of stratification there is a large area where the error is negligible (basically for all cases where $Pe < 1$ and also if $\lambda \approx 1$).

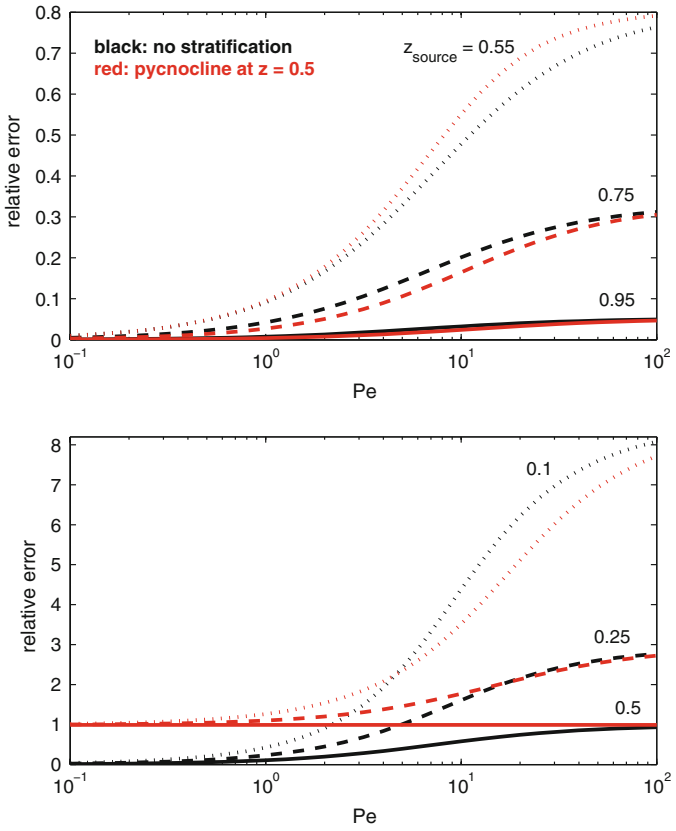


Fig. 5 Relative errors for the mean residence time in the special case without mortality. *Black curves* indicate no stratification, *red curves* the cases with stratification. *Top panel* shows the results for the source pulse located at heights $z > 0.5$ (i.e. above the pycnocline); *lower panel* for source positions $z \leq 0.5$. Results are shown for *Pe* values up to 100, in order to clearly show the convergence of the *black* and *red curves*

In the case with stratification, this area is much smaller, probably because the stratification further increases the real residence time (the bacteria are injected above the pycnocline), thus increasing the negative errors. But even then the relative errors do not exceed 16%.

The errors become much worse for a source located at $z = 0.25$. Indeed, the maximal errors, found for the expected high *Pe*-low λ conditions, are ten times larger. For higher λ values, the errors decrease to almost zero, and never become negative. For decreasing *Pe* values the errors also decrease but much less in the presence of a pycnocline, probably being the same effect observed in the situation without mortality.

Summarising the observed trends in the residence time errors, these are largest and always positive for a source closer to the bottom. For a source located higher (above the pycnocline if present), the errors are both positive and negative and have much smaller values. However, in all these cases, stratification has the effect to increase the errors in the area where *Pe* and $\lambda < 1$, conditions where the errors are negligible in the absence of stratification. Finally, if the source is located at the surface, the errors are consistently negative and they exhibit a pattern very similar to the stationary concentration errors.

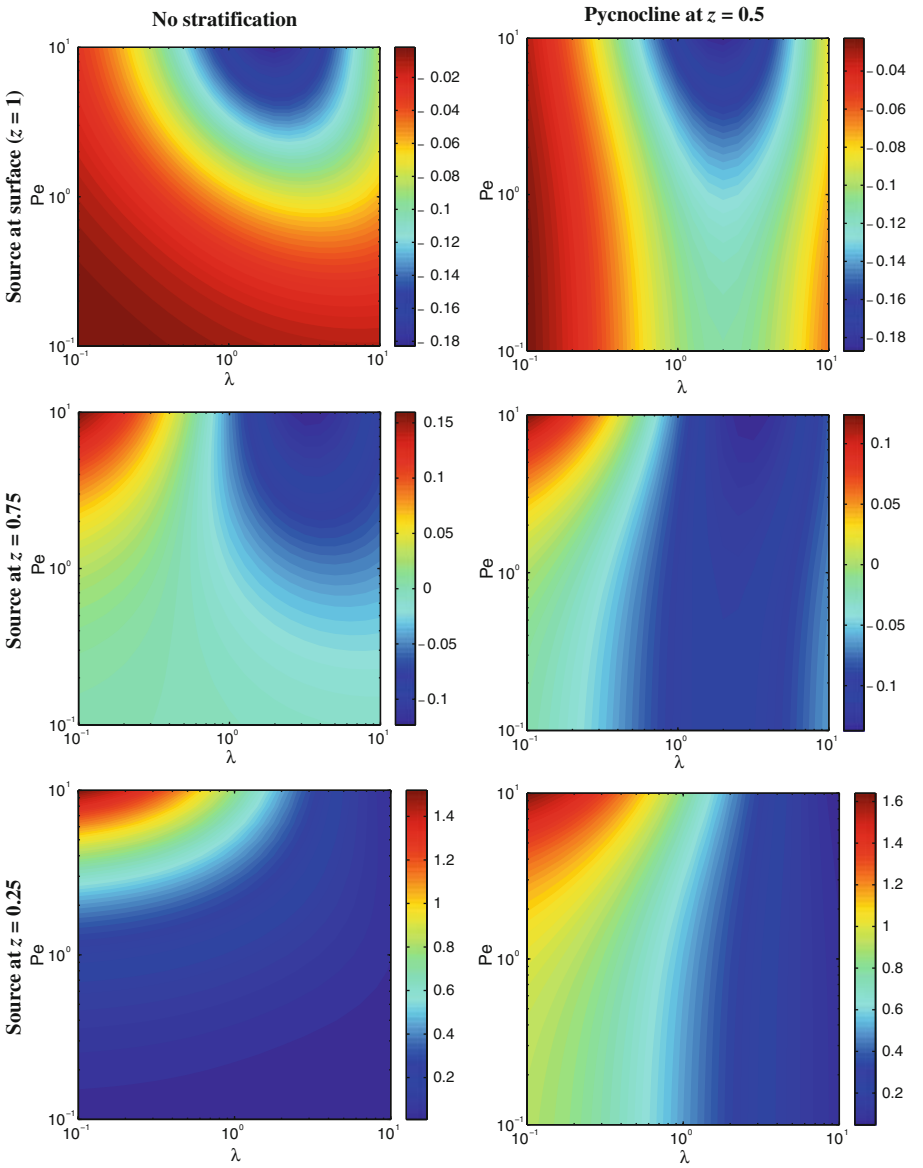


Fig. 6 Relative errors for mean residence time, in the absence (*left*) and presence (*right*) of stratification with a pycnocline at mid-depth. Three different source locations are considered: at the surface (*top row*), at $z = 0.75$ (*middle row*) and $z = 0.25$ (*bottom row*). Notice the different color scales

5 Conclusions

Using a depth-averaged model is an approximation and in this study an attempt was made to assess the impact of this approximation on the computed depth-averaged concentrations, the settling fluxes and the residence times. It is true that the equations under study are relatively simple compared to what may be used in real applications. However, this allowed to reduce

the number of potential factors (mortality, settling, mixing and stratification) and to more clearly identify their effects, sometimes even analytically. Notwithstanding these simplifications, some results are unexpected and therefore interesting to keep in mind for more realistic applications.

The main conclusions are

(1) In all cases the error increases with Pe and errors are typically negligible for $Pe < 1$, which confirms the general expectation that depth-averaging is only acceptable in situations dominated by mixing.
(2) The presence of a pycnocline has the effect to further narrow the region of negligible errors. This is attributable to the fact that stratification restrains the bacteria from crossing the pycnocline, which increases (if the source releases bacteria above the pycnocline) or decreases (if the source is located below the pycnocline) the residence time of the bacteria. As such, the relative effect of mixing (versus settling) is reduced.
(3) In general, low mortality rates increase the errors, because this results in a higher level of heterogeneity.
(4) However, for the stationary test case with bacteria injected at the surface, the concentration errors peak when mortality and settling have equivalent characteristic timescales. This was confirmed by the associated (i.e. for the bacteria injected at the surface) residence time errors. When the point source is located elsewhere, the errors again increase with decreasing λ.
(5) When mortality is negligible, it was demonstrated analytically (and confirmed numerically) that the 0D model overestimates the time spent in the water column by the bacteria.

Admittedly, it is impossible to make quantitative statements about the reliability of more complex depth-averaged models, only based on the presented results. However, we believe that they provide an indication of the error's order of magnitude, and general trends. For instance, in the considered conditions the concentration errors can amount to 29%, implying that in some situations significant errors may occur. At the same time, for many field conditions the errors are expected to be acceptable, namely as long as $Pe < 1$ (as expected), or $\lambda >> 1$. In other words, depth-averaging is allowed if the typical time associated with either mixing or mortality are significantly smaller than the time needed for settling. Only when the bacteria are injected close to the bottom, the errors really become huge. Fortunately, this is not a typical situation encountered in reality, but could happen e.g. if submarine outlets are used to move the actual discharge point away from the coast to avoid contamination of beaches, or if significant resuspension processes of contaminated bottom sediment occurs. These conditions can relatively easily be checked for any modelling exercise to assess the risk of errors due to the vertical integration of the equations.

Acknowledgements Anouk de Brauwere is a post-doctoral researcher with the Research Foundation Flanders, and was honoured by a post-doctoral intercommunity fellowship from the Francqui Foundation during this work. Eric Deleersnijder is a Research Associate with the Belgian National Fund for Scientific Research (FRS-FNRS). This work was conducted within the framework of the Interuniversity Attraction Pole TIMOTHY (IAP VI.13), funded by the Belgian Science Policy (BELSPO).

Appendix 1

In this Appendix it is explained how the range of the depth-averaged eddy diffusivity $\bar{\kappa}$ was determined (Table 1). Let u_*^b and u_*^s denote the friction velocity associated with the bottom

and surface stress, respectively. The eddy diffusivity should behave as $\gamma u_*^b z$ or $\gamma u_*^b (h-z)$, where $\gamma \approx 0.4$ is the von Karman constant, as the bottom or the surface is approached. The eddy diffusivity should reach a maximum at a point located away from the upper and lower boundaries of the domain. Therefore, a simple expression of the eddy diffusivity satisfying the above-mentioned constraints is

$$\kappa(z) = \gamma h \sigma (1-\sigma)[u_*^b + \sigma(u_*^s - u_*^b)], \tag{A1-1}$$

with $\sigma = z/h$. Then, the depth-average of the eddy diffusivity is

$$\bar{\kappa} = \frac{\gamma h (u_*^b + u_*^s)}{12}. \tag{A1-2}$$

Clearly, the ranges of the possible values of u_*^b and u_*^s are rather similar, i.e. $u_*^b \approx u_* \approx u_*^s$, yielding

$$\bar{\kappa} \approx \frac{\gamma h u_*}{6} \approx 0.1 h u_* \tag{A1-3}$$

where u_* is the order of magnitude of the bottom friction velocity and the surface one. Finally, with $u_* \approx 10^{-3}$–10^{-1}m/s and h ranging between 1–100 m, we find the range for $\bar{\kappa}$ used in this study (cf. Table 1).

Appendix 2

In this Appendix the numerical schemes used to discretise the equations in Table 2—or more generally for Eqs. 12 and 14—are outlined in more detail. The vertically integrated model Eq. 14 may be solved analytically, giving the following solution:

$$\bar{C}'(t) = \bar{C}'_0 e^{-(\lambda+1)t} + \int_0^t \bar{q}(\zeta) e^{-(\lambda+1)(t-\zeta)}\, \mathrm{d}\zeta. \tag{A2-1}$$

The solution to (A2-1) may be obtained by evaluating—analytically or numerically—the integral appearing in the right-hand side of (A2-1). Alternatively, (14) may be discretized using a scheme based on Patankar [3,4,18]. If $\bar{C}'_n = \bar{C}'(n\Delta t)$ and $\bar{q}_n = \bar{q}(n\Delta t)$ with $n =$ 0,1,2..., a simple discretisation of the depth-integrated Eq. 14 is

$$\frac{\bar{C}'_{n+1} - \bar{C}'_n}{\Delta t} = \bar{q}_n - (\lambda+1)\bar{C}'_{n+1}. \tag{A2-2}$$

This Patankar-like scheme is readily seen to be positive-definite, for (A2-2) may be rewritten as

$$\bar{C}'_{n+1} = \frac{\bar{C}'_n + \bar{q}_n \Delta t}{1 + (\lambda+1)\Delta t}, \tag{A2-3}$$

an expression which is positive for any value of the time increment Δt—provided $\bar{C}'_n$ is positive.

To discretize the one-dimensional problem, it is appropriate to introduce first the following notations

$$\begin{cases} C_k^n = C[n\Delta t, (k-1/2)\Delta z], \quad q_k^n = q[n\Delta t, (k-1/2)\Delta z], \quad \kappa_{k-1/2} = \kappa[(k-1)\Delta z] \\ n = 0, 1, 2, \ldots \quad k = 1, 2, \ldots K \quad \Delta z = 1/K \end{cases} \tag{A2-4}$$

Then, using a first-order upwind discretisation of the advective term, the Patankar-like discretisation of the governing equation is

$$\frac{C_k^{n+1} - C_k^n}{\Delta t} = q_k^n - \lambda C_k^{n+1} + \frac{\left(C_{\mathrm{k}+1}^n + \frac{\kappa_{\mathrm{k}+1/2}}{Pe}\frac{C_{\mathrm{k}+1}^{n+1} - C_k^{n+1}}{\Delta z}\right) - \left(C_k^n + \frac{\kappa_{\mathrm{k}-1/2}}{Pe}\frac{C_k^{n+1} - C_{\mathrm{k}-1}^{n+1}}{\Delta z}\right)}{\Delta z} \tag{A2-5}$$

while that of the boundary conditions is

$$\frac{\kappa_{1/2}}{Pe}\frac{C_1^{n+1} - C_0^{n+1}}{\Delta z} = 0, \quad C_{\mathrm{K}+1}^n + \frac{\kappa_{\mathrm{K}+1/2}}{Pe}\frac{C_{\mathrm{K}+1}^{n+1} - C_K^{n+1}}{\Delta z} = 0. \tag{A2-6}$$

This scheme is inspired by [6], suggesting that it is stable and positive definite for any value of the parameters. (A2-5)–(A2-6) form a tridiagonal system which was solved by means of an algorithm (Example 5.4.3) from [5]. The disadvantage of the first-order upwind scheme is that it introduces numerical diffusion (also called artificial viscosity, [19]), having magnitude $\frac{\Delta z}{2}\left(1 - \frac{\Delta t}{\Delta z}\right)$. Therefore, $\Delta t = \Delta z$ was taken, thus eliminating the unwanted numerical diffusion and delivering a reliable solution for the advective (here only vertical sinking) term.

References

1. Alvarez-Vázquez LJ, Martínez A, Vázquez-Méndez ME, Vilar MA (2006) Optimal location of sampling points for river pollution control. Math Comput Simul 71:149–160
2. Bolin B, Rodhe H (1973) A note on the concepts of age distribution and transit time in natural reservoirs. Tellus 25:58–62
3. Burchard H, Deleersnijder E, Meister A (2003) A high order conservative Patankar-type discretisation for stiff systems of production-destruction equations. Appl Numer Math 47:1–30
4. 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
5. Dahlquist G, Bjorck A (1974) Numerical methods. Prentice-Hall, City
6. Deleersnijder E, Beckers J-M, Campin J-M, El Mohajir M, Fichefet T, Luyten P (1997) Some mathematical problems associated with the development and use of marine models. In: Diaz JI (ed) The mathematics of models for climatology and environment, NATO ASI Series, vol 148. Springer-Verlag, Berlin, pp 39–86
7. Deleersnijder E, Beckers J-M, Delhez EJM (2006) The residence time of settling particles in the surface mixed layer. Environ Fluid Mech 6:25–42
8. Delhez EJM, Heemink AW, Deleersnijder E (2004) Residence time in a semi-enclosed domain from the solution of an adjoint problem. Estuar Coast Shelf Sci 61:691–702
9. Delhez EJM (2006) Transient residence and exposure times. Ocean Sci 2:1–9
10. Delhez EJM, Deleersnijder E (2006) The boundary layer of the residence time field. Ocean Dyn 56:25–42
11. Garcia-Armisen T, Thouvenin B, Servais P (2006) Modelling faecal coliforms dynamics in the Seine estuary, France. Water Sci Technol 54((3):177–184
12. Garcia-Armisen T, Servais P (2008) Partitioning and fate of particle-associated *E. coli* in river waters. Water Environ Res 80(3):282–288
13. George I, Anzil A, Servais P (2004) Quantification of fecal coliform inputs to aquatic systems through soil leaching. Water Res 38:611–618
14. Hofmann AF, Soetaert K, Middelburg JJ (2008) Present nitrogen and carbon dynamics in the Scheldt estuary using a novel 1-D model. Biogeosciences 5:981–1006

15. Inoue M, Wiseman WJ Jr (2000) Transport, mixing and stirring processes in a Louisiana estuary: a model study. Estuar Coast Shelf Sci 50:449–466
16. Jamieson R, Gordon R, Joy D, Lee H (2004) Assessing microbial pollution of rural surface waters. A review of current watershed scale modeling approaches. Agric Water Manage 70:1–17
17. Jamieson R, Joy DM, Lee H, Kostaschuk R, Gordon R (2005) Transport and deposition of sediment-associated *Escherichia coli* in natural streams. Water Res 39:2655–2675
18. Patankar SV (1980) Numerical heat transfer and fluid flow. McGraw-Hill, New York
19. Roache PJ (1972) On artificial viscosity. J Comput Phys 10:169–184
20. Servais P, Garcia-Armisen T, George I, Billen G (2007) Fecal bacteria in the rivers of the Seine drainage network (France): Sources, fate and modelling. Sci Total Environ 375(1–3):152–167
21. Steets BM, Holden PA (2003) A mechanistic model of runoff-associated fecal coliform fate and transport through a coastal lagoon. Water Res 37:589–608
22. Takeoka H (1984) Fundamental concepts of exchange and transport time scales in a coastal sea. Cont Shelf Res 3:311–326
23. Vanderborght JP, Folmer IM, Aguilera DR, Uhrenholdt T, Regnier P (2007) Reactive-transport modelling of C, N and O_2 in a river-estuarine-coastal zone system: application to the Scheldt estuary. Mar Chem 106:92–110
24. Yuan D, Lin B, Falconer RA (2007) A modeling study of residence time in a macro-tidal estuary. Estuar Coast Shelf Sci 71:401–411

Environ Fluid Mech (2010) 10:177–196
DOI 10.1007/s10652-009-9144-5

ORIGINAL ARTICLE

A model diagnostic study of age of river-borne sediment transport in the tidal York River Estuary

Wenping Gong · Jian Shen

Received: 25 February 2009 / Accepted: 10 July 2009 / Published online: 8 August 2009

Abstract As nutrients and organic matters are transported preferentially in an adsorbed state and tend to bind to the sediments, sediment transport plays an important role on eutrophication processes in the estuaries. The timescale of sediment transport is of significance for studying the retention of pollutants and eutrophication processes in the estuaries. Unlike transport of dissolved substances that is mainly controlled by advection and diffusion processes, the sediment transport is significantly affected by the intermittent settling and resuspension processes. A three-dimensional model with suspended sediment transport was utilized to investigate the transport timescale of river-borne sediment in the tidal York River Estuary. The results indicate that river discharge dominantly determines the age of river-borne sediment in the estuary. High river discharge results in a low sediment age compared to that under mean flow. The intermittent effects of settling and resuspension events greatly affect the river-borne sediment age. Both settling velocity and critical shear stress are shown to be key parameters in determining the sediment transport timescale. The sediment age decreases as settling velocity and/or critical shear stress decrease, while it increases with the increase of settling velocity that prevents the sediment to be transported out of the estuary.

Keywords Transport of suspended sediment · Age of sediment · York River Estuary · Chesapeake Bay · Sediment modeling

W. Gong (✉)
School of Marine Science, Sun Yat-Sen University, 135 Xingangxi Road,
Guangzhou 510275, China
e-mail: gongwp@mail.sysu.edu.cn

J. Shen
Virginia Institute of Marine Science, 1208 Greate Road,
Gloucester Point, VA 23062, USA

1 Introduction

Many estuaries and coastal seas have become more eutrophic over the past decade as anthropogenic inputs of nutrients increase [2,29]. The amount of nutrients discharged into the estuary and the transport time required for these nutrients to be exported to the concerned area along with other biogeochemical processes are vital elements leading to the eutrophication. Because of the cohesive properties of fine sediments, nutrients and organic matters tend to bind to the sediments and are transported preferentially in an adsorbed state [19]. When sediments are stirred up, adsorbed nutrients and organic matters are released into the water column and can impact significantly the ecological balance of the estuaries [23]. Moreover, the amount of sediment in suspension influences the turbidity, which controls the light penetration in the water column and influences the primary production [27]. Therefore, sediment transport and its timescale play important roles in affecting eutrophication processes in estuaries.

The age concept of a water parcel has been used to quantify the transport timescale in lagoons, estuaries and oceans successfully (e.g., [1,3–6,31,36,38]). Delhez et al. [5] introduced the age theory based on the advection–diffusion of a tracer and provided a general methodology to compute age using an Eulerian modeling approach. The theory has been widely applied to conduct diagnosis study on the water renewal [15], mixing [16], vertical transport [37], and estuarine dynamics [32,33].

Recently, Mercier and Delhez [26] extended the age concept to conduct diagnosis study of the sediment transport in the Belgian Coastal Zone. They made a distinction between two complementary aspects of resuspension and horizontal transport of the sediment, namely "resuspension age" and "transport age". The "resuspension age" is introduced to study the duration of the resuspension–deposition events, and the age of a particle is defined as the time elapsed since that particle is eroded from the bottom. The "transport age" is introduced to measure the mean age of the suspended sediment particles to travel from a source region to a concerned location. In this study, we applied this age concept to study the transport time of river-borne sediment discharged into the estuary. The age of a particle is the time elapsed since the particle is discharged into the estuary. Therefore, the age at any location in the estuary measures the average time taken by the suspended particles to travel from the headwater to the observation location. In essence, the age of river-borne sediment in this study is similar to the "transport age" proposed by Mercier and Delhez [26].

In this effort, a three dimensional Environmental Fluid Dynamic Code (EFDC) model developed by Hamrick [18] was used to compute the age of sediment in the tidal York River, a partially-mixed estuary (Fig. 1). A sediment module described in Lin and Kuo [22] was utilized. The suspended sediments were continuously discharged into the estuary at the headwaters of two tributaries, Pamunkey and Mattaponi Rivers, under different river flow conditions. Apart from the sediment concentration, a variable of age concentration was introduced. The sediment concentration and age concentration were calculated based on the corresponding advection–diffusion equations [26]. As the suspended sediment in the water column exchanges with that at the seabed all the time, the sediment and age concentrations at the seabed were also calculated. The effects of river discharge, inflowing sediment concentrations at the headwaters, sediment settling velocity and critical shear stress for erosion on sediment age were examined. Through this study, the spatially varying timescales of cohesive sediment transport under different hydrodynamic conditions in the estuary are determined quantitatively. It is hoped that this study would provide an approach to understand the river-borne sediment transport timescale in the York River, which endures both intermittent settling and resuspension events, and advection and diffusion processes that are experienced by other dissolved substances in the estuary.

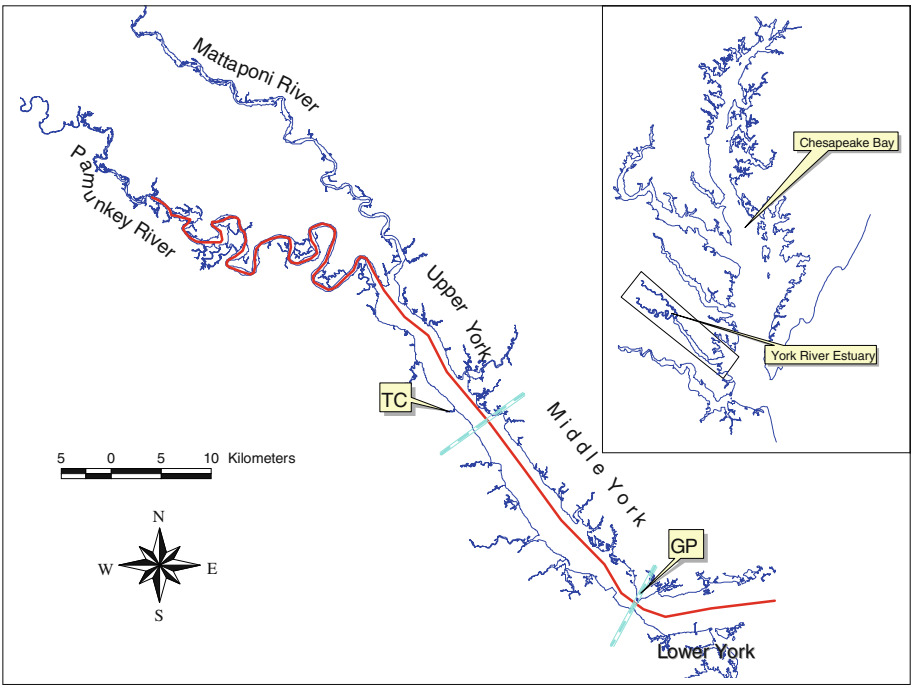


Fig. 1 A map of tidal York River Estuary (*double line* in the York River separated lower York, middle York and upper York)

2 Study site

The York River Estuary is a western tributary of the Chesapeake Bay (Fig. 1), which formed through inundation during the Holocene sea level rise [28]. It is a partially mixed estuary with a fortnightly stratification–destratification cycle [17] and can be subdivided into lower, middle, and upper reaches [9]. The lower York is the region below Gloucester Point (GP), where the river is relatively wide (3–6 km) with a broad shoal on the northern side and a narrow shoal on the southern side; The middle York begins at GP, where a constriction and bend in the river occur approximately 10 km from the mouth. In the middle York, there are two channels: a main channel to the north with a mean depth of approximately 10 m and a secondary channel with a mean depth of 5 m to the south. The upper York extends just above the secondary channel to the confluence of the Pamunkey and Mattaponi Rivers, which is about 53 km from the river mouth. The mean tidal range at the mouth is 0.7 m and increases to 0.9 m at the heads of the Pamunkey and Mattaponi Rivers, where the long-term mean freshwater discharges are approximately 28.7 and 14.4 $m^3\ s^{-1}$, respectively, based on flow data from United States Geological Survey (USGS) at upstream stations of the Pamunkey and Mattaponi Rivers from 1942 to 2001. Tidal influences reach upstream towards the fall-lines, which are approximately 97 km upriver in the Mattaponi and 60 km upriver in the Pamunkey, respectively.

The percentage of mud of the bottom sediment is more than 80% in the channel areas, whereas the mud percentage becomes less than 40% along the shorelines [28]. The mud percentage of the bottom sediment is quite high from the middle portion to the upstream

end of the York River while much sandier beds are located at its upstream and downstream portions of the River. This characteristic of bottom sediment distribution was concluded to contribute to the existence of the Secondary Turbidity Maximum (STM) in the estuary [21].

Sediment dynamics in the York River Estuary have been studied intensively. The bottom sediment texture or size distribution in the estuary was investigated by Nichols et al. [28]. The mass properties of the sediment, such as porosity, density, shear stress, etc., were studied by Faas [11]. Dellapenna et al. [8,9] investigated the sea-bed mixing, ephemeral deposition, and fine-scale strata formation in the estuary. Kniskern and Kuehl [20] studied the spatial and temporal variabilities of seabed disturbance in the estuary. Lin and Kuo [21,22] explored the mechanism of the evolution of the STM aside from the Primary estuarine Turbidity Maximum (PTM). Dickhudt et al. [10] investigated the variability and controls on seabed erodibility in the estuary. Recently Gong and Shen [14] investigated the sediment transport in the York River during Tropical Cyclone Isabel of 2003.

3 Methodology

3.1 Cohesive sediment transport model

The EFDC model was used in this study [18]. The model solves the Navier–Stokes equations for a domain with a free surface assumption. Mellor and Yamada's level 2.5 turbulence closure model [25] modified by Galperin et al. [12] is utilized in the model to obtain the vertical eddy viscosity and diffusivity. Sigma vertical coordinates and Cartesian or curvilinear-orthogonal horizontal coordinates are used in the model. In this study, throughout the mainstem of the York River, the Cartesian grid was utilized with a horizontal resolution of 250 m. Curvilinear grid with varying grid sizes was used in the Pamunkey and Mattoponi Rivers. The model had eight equally vertical layers.

The sediment transport model is coupled with the hydrodynamic model with the same temporal and spatial resolutions. The governing equation for the total suspended solids (TSS) concentration is:

$$\frac{\partial C}{\partial t}+\frac{\partial Cu}{\partial x}+\frac{\partial Cv}{\partial y}+\frac{\partial C(w-w_s)}{\partial z}=\frac{\partial}{\partial x}\left(k_h\frac{\partial C}{\partial x}\right)+\frac{\partial}{\partial y}\left(k_h\frac{\partial C}{\partial y}\right)+\frac{\partial}{\partial z}\left(k_z\frac{\partial C}{\partial z}\right) \quad (1)$$

where C is the sediment concentration in the water column; x and y are the horizontal coordinates and z is the vertical coordinate positive upward; u, v, w are the water velocity components in x, y, z directions, respectively; w_s is the sediment settling velocity; and k_h and k_z are the horizontal and vertical turbulent diffusion coefficients, respectively.

At the water surface, no sediment flux is allowed, and the boundary condition is

$$w_s C+k_z\frac{\partial C}{\partial z}=0 \quad (2)$$

At the bottom, the boundary condition for sediment flux is

$$w_s C+k_z\frac{\partial C}{\partial z}=D-E \quad (3)$$

where E is the mass of sediment eroded from the seabed per unit area per unit time, also known as the erosion or resuspension rate; D is the mass of sediment deposited to the seabed per unit area per unit time, or deposition rate. As the bottom sediments are mainly composed of silty clay [7,28], only the mud is simulated in this study. The erosion rate for cohesive sediment is simulated as:

$$E = \begin{cases} M\left(\frac{\tau_b}{\tau_{ec}} - 1\right) & \text{for} \quad \tau_b > \tau_{ec} \\ 0 & \tau_b \leq \tau_{ec} \end{cases} \tag{4}$$

where τ_b is the bed shear stress, M is an empirical constant with the same unit as E, and τ_{ec} is the critical shear stress for erosion.

The deposition rate D is calculated as:

$$D = \begin{cases} w_s C_b\left(1 - \frac{\tau_b}{\tau_{dc}}\right) & \text{for} \quad \tau_{dc} > \tau_b \\ 0 & \tau_{dc} \leq \tau_b \end{cases} \tag{5}$$

where C_b is the sediment concentration at the bottom layer of the water column. τ_{dc} is the critical shear stress for deposition.

In this study, constant values of empirical parameters of $M = 0.0005\,\mathrm{gm^{-2}\,s^{-1}}$ and $\tau_{dc} = 0.035\,\mathrm{Pa}$ were used within the model domain. These values are the same as those used in Lin and Kuo [22], which were obtained through model calibration of TSS simulations. τ_{ec} and w_s were varied for examining the effects of critical shear stress for erosion and settling velocity on the sediment transport timescale. As the effect of M can be expected to be opposite to that of τ_{ec}, it was not tested in this effort.

Apart from the suspended sediment in the water column, the sediment concentration at the seabed C_s is computed as:

$$\frac{\partial C_s}{\partial t} = D - E \tag{6}$$

where the seabed has been assumed as well mixed. Note that the unit of C_s is $\mathrm{gm^{-2}}$, if a unit depth at the seabed is assumed, then C_s has the same unit as C.

3.2 Sediment age

The sediment age can be computed based on the sediment and age concentrations [26]. The sediment concentration is obtained by Eqs. 1–5. The age concentration in the water column is computed as:

$$\frac{\partial \alpha}{\partial t} + \frac{\partial \alpha u}{\partial x} + \frac{\partial \alpha v}{\partial y} + \frac{\partial \alpha (w - w_s)}{\partial z} = \frac{\partial}{\partial x}\left(k_h \frac{\partial \alpha}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_h \frac{\partial \alpha}{\partial y}\right) + \frac{\partial}{\partial z}\left(k_z \frac{\partial \alpha}{\partial z}\right) + C \tag{7}$$

where C is the sediment concentration, α is the age concentration.

The age concentration at the seabed is described as:

$$\frac{\partial \alpha_s}{\partial t} = C_s + D_\alpha - E_\alpha \tag{8}$$

where

$$D_\alpha = \frac{\alpha_b}{C_b} D \tag{9}$$

$$E_\alpha = \frac{\alpha_s}{C_s} E \tag{10}$$

in which α_b, α_s are the age concentration at the bottom layer in the water column and at the seabed, respectively. The boundary condition at the interface of water column and seabed is:

$$w_s \alpha + k_z \frac{\partial \alpha}{\partial z} = D_\alpha - E_\alpha \tag{11}$$

The sediment mean age "a" at both the water column and seabed can then be calculated as follows:

$$a = \frac{\alpha}{C}. \tag{12}$$

3.3 Simulation set-up

For the model simulations, the suspended sediments were discharged continuously from the headwaters of Pamunkey and Mattaponi Rivers. Constant or time series of sediment concentrations for the tributaries were specified for different model simulations, the age concentrations were set as zero at the headwaters. For tracking the sediment age in the estuary, we assumed that initially there was no sediment in the water column and at the seabed for the computation purpose, as we are mainly interested in the timescale of river-borne sediment transport. The clock started as soon as the sediment was discharged into the estuary. The erosion rate E was limited by the available sediment on the seabed at each time step. When the total sediment deposited on the seabed divided by the timestep was less than the calculated erosion rate E, the erosion rate was adjusted to be equivalent to the available sediment on the seabed divided by timestep.

The transport processes of dissolved substances, depends primarily on the low frequency and mean motions of the water in the estuary [24]. For a given estuary, the variation of low-frequency residual flow is determined by the density field, river flow, wind and the nonlinear rectification of the periodic tides. As revealed by Shen and Haas [31], the water age in an estuary is mostly controlled by the river discharge. In this study, the response of sediment transport timescale to the low frequency and mean motions of the water was studied through a series of experiments under the mean and high flow conditions. The flow data obtained from USGS at stations upstream of Pamunkey and Mattaponi Rivers from 1942 to 2001 were analyzed. The mean flow and the 90th percentile of high flow were selected for the model experiments, which are representative of mean and wet hydrologic conditions. The freshwater discharges used for model experiments are listed in Table 1.

Though wind is revealed to be important in influencing the mixing, stratification and the estuarine circulation [13,30], wind forcing is not included in the model simulations, which will be studied in the future.

The difference of the sediment transport from those of dissolved substances lies in the fact that apart from the effect of low-frequency residual flow, it is also affected by the occurrence of intermittent sedimentation/resuspension events. The selection of parameters for settling velocity, critical shear stress for erosion etc, would impose impacts on the sediment transport timescale. If the settling velocity and erosion rates are set to zero, the corresponding sediment age will resemble the water age for a dissolved conservative tracer. In regard to this aspect, a suite of experiments were conducted to test the effects of settling velocity. The effect of critical shear stress for erosion was examined in another experiment. Because in our model simulations, sediment resuspension is controlled by the availability of river-borne sediment

Table 1 Flow conditions for model experiments

River	High flow ($m^3\,s^{-1}$)	Mean flow ($m^3\,s^{-1}$)
Mattaponi	38.5	14.4
Pamunkey	69.0	28.7

at the seabed, which is affected by the inflowing sediment concentration at the headwaters of the tributaries, a suite of experiments were also conducted to investigate the effect of inflowing sediment loading on the age distribution.

In the model simulations, seven tidal constituents (M_2, S_2, N_2, K_1, O_1, M_4 and M_6) were used to specify the tidal open boundary for the model [31]. The open boundary condition for salinity was based on the specification of inflowing salinity during the flood tide. Outflowing salinities were calculated using upwind salinities immediately inside the open boundary. When the flow at the open boundary changed from outflow to inflow, the model provided a linear interpolation of inflowing salinity based on the last outflowing salinities and the specified incoming salinity within a specified time interval. A 1.5 h time interval was used for the simulation based on previous model calibration [34,35]. Once the flooding time was longer than this time interval, the prescribed incoming salinity was used as the boundary condition. A constant salinity concentration of 24 and 19 psu was specified at the bottom and the surface, respectively. Model linearly interpolated salinity for other middle layers for incoming salinity. The incoming sediment and age concentrations at the open boundary were set to be zero, as we were concerned with the transport of sediments from the headwaters of the tributaries. The outflowing sediment and age concentrations were calculated using the same way as the salinity. The temperature was not simulated during the model experiments. The model was initially run for 3 months for each flow condition to obtain a dynamic equilibrium condition. The flow field and salinity distribution under this equilibrium condition were used as the initial condition for the model experiments so that the experiments could be 'hot' started.

The simulation duration for each experiment varied with river flow and depended on the model parameters (e.g., settling velocity). The simulation experiments for different set-ups are summarized in Table 2. The sediment mean age of each vertical layer was calculated based on Eq. 12. The resultant at each vertical layer was averaged to obtain the vertically averaged sediment mean age. Since the model was forced with seven tidal constituents at the mouth, only dynamic equilibrium could be reached. Under both the high flow and mean flow conditions, the sediment and age concentrations attained dynamic equilibrium in the late simulation periods. Therefore, an average of vertically averaged sediment mean age during the last 15 days was computed, which is referred as Average Sediment Mean Age (ASMA) in the following sections.

4 Model results

4.1 Sediment age under the high and mean flow conditions

The distribution of ASMA under the high flow condition (Experiment E1) is shown as contour plots in Fig. 2. Figure 2a, b depict the ASMA in the water column and at the seabed respectively. The results show that a substantial time is required for the sediment discharged at the headwaters to be transported downstream in the estuary. It takes approximately 250 days for the sediment to transport from the headwaters to the river mouth (Fig. 2a). The ASMA at the seabed can reach up to 300 days at the river mouth (Fig. 2b). Compared to the water age of 55 days at the river mouth for the conservative dissolved substance under high flow condition [31], a considerable increase of transport timescale is observed for the river-borne sediment. Apart from the increase of magnitude of transport timescale, the spatial distribution is also fundamentally altered. The most striking feature of the spatial distribution of ASMA in the water column is that there emerges a low age valley (represented by 100-day contour in Fig. 2a) between the upper and middle portions of the York River, which is lower than

Table 2 A summary of the model experiments

Experiment	River flow ($cm\,s^{-1}$)	River inflow sediment concentration ($mg\,l^{-1}$)	Settling velocity ($m\,s^{-1}$)	Critical shear stress for erosion (Pa)	Simulation duration (days)	Comment
E1	High	20	5.0×10^{-5}	0.1	730	See Table 1 for flow
E1-1	High	20	5.0×10^{-5}	0.1	730	The river discharge and sediment form the Mattaponi River are blocked
E1-2	High	20	5.0×10^{-5}	0.1	730	The sediment discharge from the Mattaponi River are excluded with the same river discharge as E1
E2	Mean	20	5.0×10^{-5}	0.1	4,380	
E3	Time series of flow	Time series of sediment concentration	5.0×10^{-5}	0.1	2,920	
E4	High	50	5.0×10^{-5}	0.1	730	
E5	High	10	5.0×10^{-5}	0.1	730	
E6	High	20	1.0×10^{-5}	0.1	730	
E7	High flow	20	1.0×10^{-4}	0.1	3,650	
E8	High	20	5.0×10^{-5}	0.05	730	

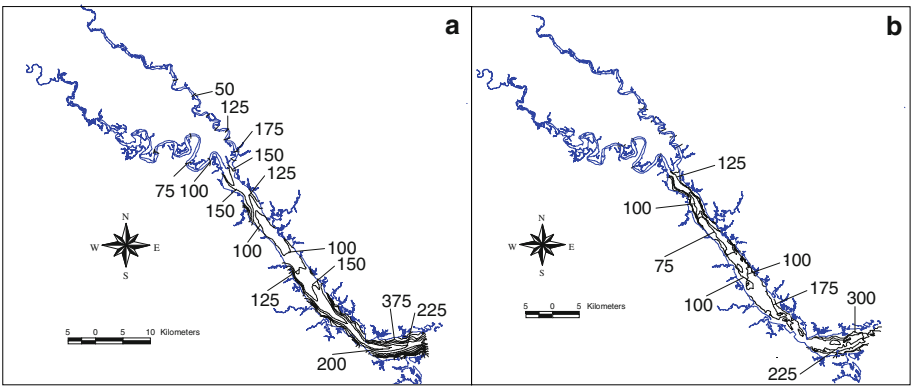


Fig. 2 The distribution of ASMA: **a** in the water column; **b** at the seabed (in days) under high flow condition

both the upstream and downstream. Downstream of this age valley, the tongue-shaped age contours direct downstream in the channel. The low age is located in the channel, whereas the high age is located along the side shoal areas. Upstream of the valley, the tongue-shaped age contours direct upstream with the low age located in the channel. Consistent with the spatial distribution of the ASMA in the water column, the ASMA at the seabed presents the similar pattern. Corresponding to the low age valley in the water column, the ASMA at the seabed is 75–100 days, which is lower than both the downstream and upstream areas. Compared to the transverse distribution of water age [31], a distinguished difference can be observed. For the transport of conservative tracer without settling and resuspension, the water age is lower near the southern bank, which is attributed to the influence of bathymetry, Coriolis force, and tidal asymmetries. However, the ASMA is generally higher in the side shoal areas along both southern and northern banks than that in the channel, suggesting that the cohesive sediment is easier to be trapped along the shallow areas, while it is transported more quickly in the channel.

A further examination of the longitudinal distribution of salinity, sediment concentration, age concentration and sediment age (Fig. 3) indicates that the age valley between the middle and upper York (Fig. 3d) is around the null point where the limit of salt intrusion resides, approximately 40 km from the river mouth (Fig. 3a). This area is also a transition zone between the upstream well-mixed and downstream partially-mixed portions of the estuary. A peak of sediment concentration is developed in this zone that merits the mechanisms similar to the formation of the Estuary Turbidity Maximum (ETM) (Fig. 3c). The formation of the ETM in the York River Estuary has been attributed to tidal asymmetry, estuarine circulation, turbulence suppression by stratification and bottom resuspension [21]. For this study, due to the limit of sediment availability at the seabed as we assumed that there were no sediments in the water column and at the seabed initially, the peak sediment concentration (about 1.5 mg l^{-1}) is not as distinguished as the normal ETM. The higher age in the upstream of the ETM zone could be caused by the transport of sediment particles from the Mattaponi River, whose sediment ages are much older than those from the Pamunkey River. A combination of these two sediment sources could result in an increased sediment age there. To test this hypothesis, two supplementary experiments (E1-1 and E1-2) were conducted. In E1-1, the cells connecting the Mattaponi and York Rivers were set as land cells to block both the water and sediment discharges from the Mattaponi River down to the York River. In E1-2, the model specifications were the same as those in E1 except that no sediment loading was introduced to the headwater of the Mattaponi River.

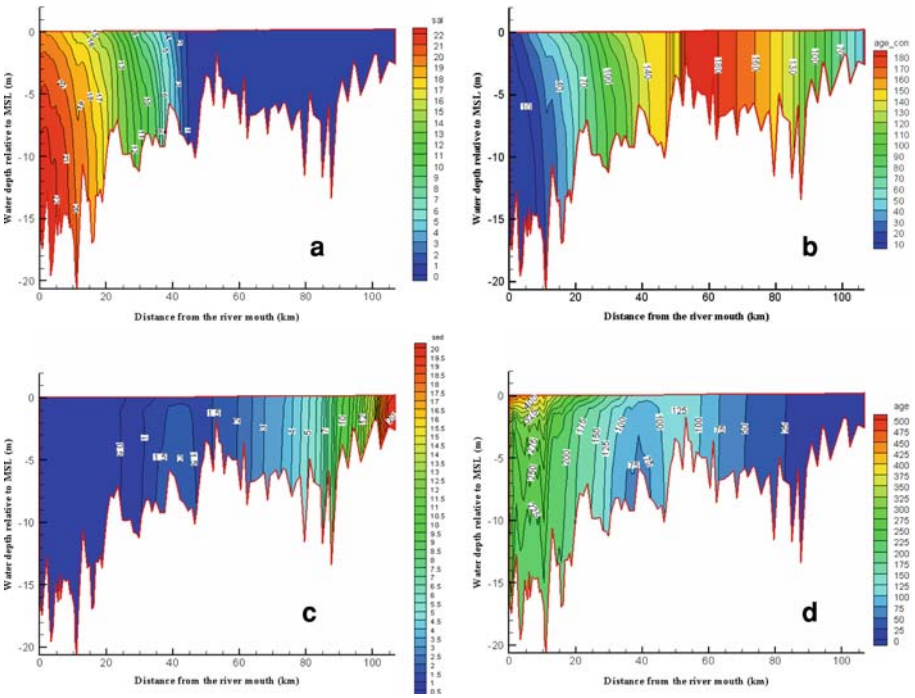


Fig. 3 The distribution of **a** salinity, **b** age concentration, **c** sediment concentration, **d** age along the longitudinal transect under high flow condition

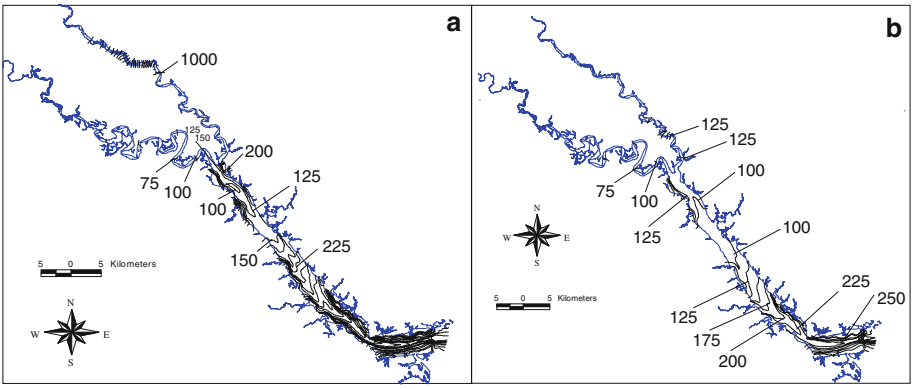


Fig. 4 Model results of ASMA for: **a** without both flow and sediment discharges from the Mattaponi River (exp. E1-1); **b** without sediment loading from the Mattaponi River (exp. E1-2)

The result (Fig. 4a) of E1-1 shows that the age valley around the salt intrusion limit (the ETM zone) is apparently existent, but not as distinct as that of Experiment E1. However, the result (Fig. 4b) of E1-2 indicates that the age valley is obviously existent in the York River. This suggests that the old sediment from the Mattaponi River is not a necessity for the development of the age valley, but it helps to augment the sediment age upstream of the age valley.

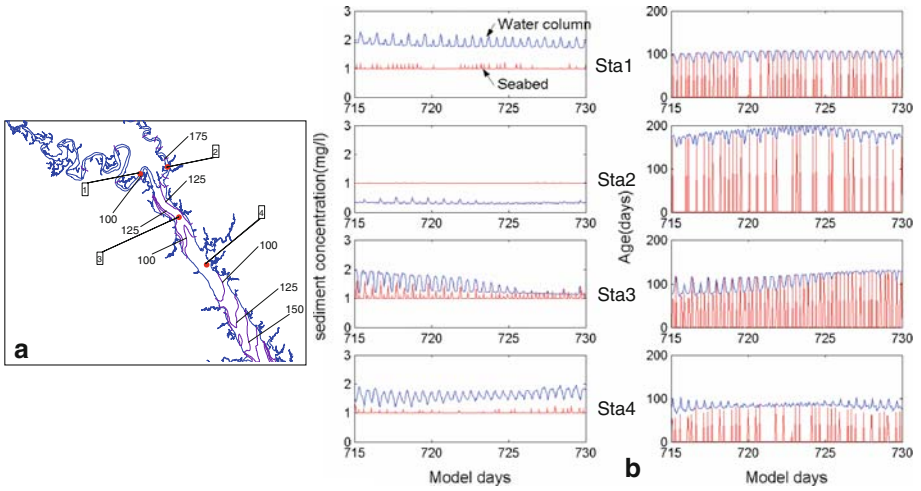


Fig. 5 Time series of vertically averaged sediment concentration and age in the water column and at the seabed at four selected stations: **a** locations of four stations; **b** variations of sediment concentration and age (*coarse lines* signify those in the water column, *thin lines* depict those at the seabed)

This indicates that the interaction of flow and sediment loading from the two tributaries plays a role in the formation of the low age valley. The sediment from the Pamunkey River is much easier to be transported downstream to the York River, while the sediment from the Mattaponi River has the tendency to be deposited near the junction area of the Mattaponi River and York River. Figure 5 shows time series of vertically averaged sediment concentration and age in the water column and at the seabed for four selected stations near the age valley. Stations 1 and 2 are located in the Pamunkey and Mataponi, respectively. Stations 3 and 4 are located upstream and inside the age valley, respectively (Fig. 5a). It can be seen (Fig. 5b) that more sediment is deposited and resuspended at the station upstream of the age valley in a tidal cycle, and the exchange of sediment between the water column and seabed is more active. Sediment age increase as the deposition period increases, especially during the neap tide. As there is more sediment on the seabed available for resuspension, which is generally older than that in the water column, the vertically averaged sediment age in the water column increases. Inside the age valley, however, the exchange between the water column and seabed is relatively inactive, and there is less sediment suspended from the seabed, resulting in a lower age in the water column.

Another important feature of the sediment age that is different from the water age lies in the vertical distribution. In contrast to the water age stratification pattern with lower age at the surface and higher age at the bottom [32], the higher sediment age is located at the surface while the lower sediment age is at the bottom in the lower and middle York (Fig. 3d). The probable reason for this pattern could be attributed to the intermittent sediment settling and resuspension processes. Note that the water age in the upper portion of the York River is about 25–45 days under high flow condition [31], suggesting that suspended sediment discharged into the headwaters of the tributaries will not be transported to the York River immediately even without experiencing periodical settling and resuspension. Under periodical settling and resuspension, the sediments from Pamunkey and Mattaponi Rivers need for approximately 100 and 175 days to be transported to the York River, respectively. Large portion of the suspended sediment in the York River is from the seabed resuspension

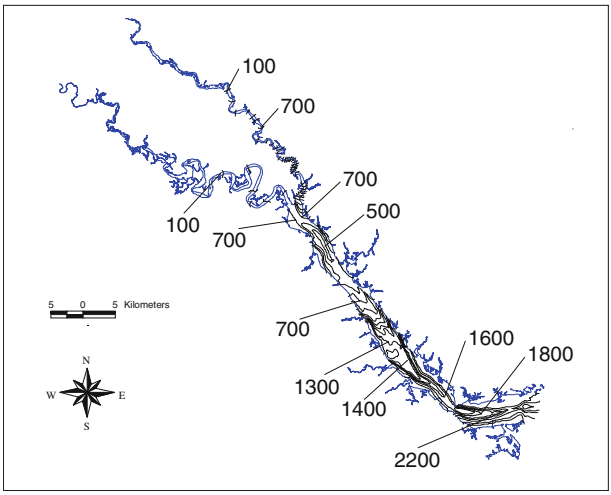


Fig. 6 The distribution of ASMA (in days) under mean flow condition

that deposited previously on the seabed when current is weak around the slack water period. Therefore, a typical vertical sediment concentration profile in the water column is developed with a high sediment concentration near the bottom and low sediment concentration near the surface mainly through vertical diffusion and settling. A much longer time is required for the suspended sediment to be transported from the bottom to the surface, resulting in a higher sediment age at the surface.

Under the mean flow condition (Experiment E2), the ASMA increases considerably (Fig. 6). The sediment particles discharged from the headwaters of the tributaries need approximately 2,200 days (about 6 years) to transport to the river mouth. The sediment from the headwater of the Mattaponi River is almost entrapped in its river course. The ASMA there can reach 2,400 days. The result suggests that the estuary serves as a trap for the sediment coming from the river discharge under mean flow condition. This is consistent with the fact that the York River Estuary has been undergoing long-term gradual infilling [28]. Due to the nature of intermittent settling and resuspension, the behavior of the sediment is much different from that of a conservative tracer, which escapes from the estuary within 110 days under mean flow condition [31]. With salt intrusion farther upstream than under high flow condition, the ASMA valley is shifted upstream, while the lateral distribution of the sediment age demonstrates the same pattern as that under high flow condition, with the higher age located in the side shoals and the lower age in the channel.

4.2 Sediment age of a real-time simulation

A real time simulation (Experiment E3) was conducted for examining the effects of varying hydrodynamic conditions on the sediment transport timescale. The simulation initially started from 1 January 2000 and simulated for 4 years. During the simulation period, there existed several intervals with extremely low river discharge (Fig. 7), suggesting a much long time was needed to transport the sediment downstream. The simulation was thus iterated one more time, extending the simulation time to be 8 years. At the headwaters of the tributaries, two USGS gauging stations at Hanover (ID: 01673000) and Beulahville (ID: 1674500)

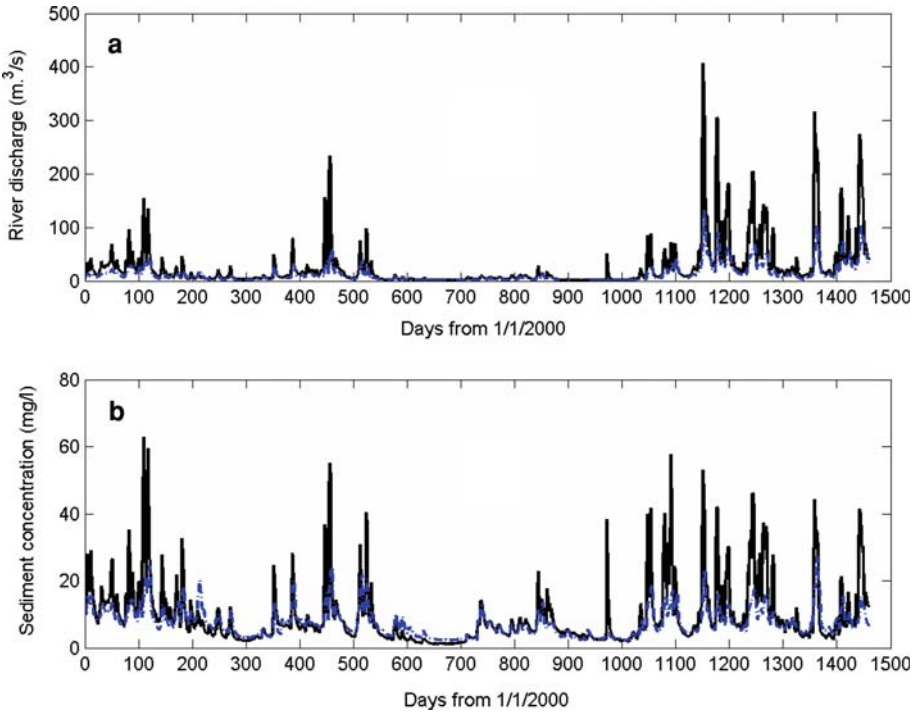


Fig. 7 Time series of river discharge (**a**) and sediment concentration (**b**) from 1 January 2000 (*solid line* means the Pamunkey River, *dashed line* signifies the Mattaponi River)

provided the daily freshwater discharges. The suspended sediment concentrations at those two stations were obtained from a seven-parameter log-linear regression model [22]. The parameters of the regression model were determined from historical data when TSS was monitored bi-weekly (1974–1994) and were used for the prediction of TSS based on river discharge and the time of the year. The time series of river discharges and TSS concentrations for the Pamunkey and Mattaponi Rivers are shown in Fig. 7.

A snapshot of the model result of ASMA averaged over the last 15 days is shown in Fig. 8. It can be seen that the sediment age is lower than that under high flow condition. The sediment takes approximately 200 days to escape from the estuary. Along the longitudinal transect of the estuary (Fig. 9a), the sediment age increases from the headwaters of the tributaries to the river mouth continuously, and the low age valley is not visible. As the river discharge was much high for the late part of the simulation period, the sediment discharged into the estuary was quickly flushed out.

Two locations near Gloucester Point and Taskinas Creek (Fig. 1, GP and TC) were selected for examining the time series variations of the sediment age. The vertically averaged sediment mean ages are shown in Fig. 9b. The ages at both stations vary with the river discharge. A low discharge results in high ages, and vice verse. Two periods of age decrease are much obvious in the simulation results, which are around Day 500 and 1,100, respectively. In 2003 (the last year in the simulation period), the river flow was exceptionally high, and the ages at both stations decreased significantly. This result further confirms that the sediment transport timescale in the estuary is controlled highly by the river discharge.

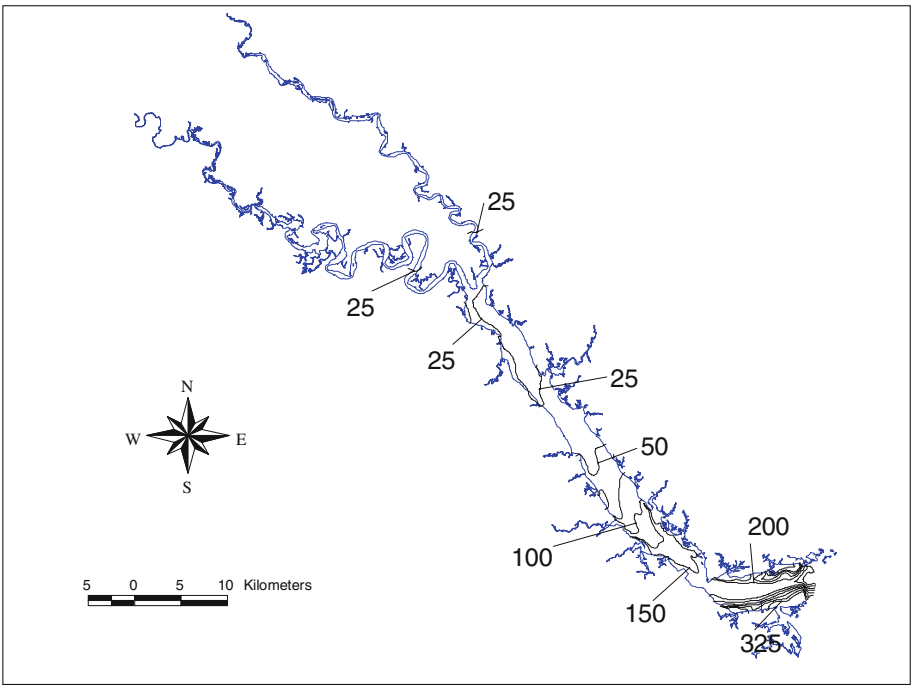


Fig. 8 A snapshot of distribution of ASMA (in days) averaged over the last 15 days in the real time simulation

5 Discussion

5.1 The effect of inflowing sediment concentration

In the above simulations, a constant sediment inflow concentration of $20\,\mathrm{mg\,l^{-1}}$ was specified at the headwaters of both the Pamunkey and Mattaponi Rivers. Because sediment deposition depends on the sediment concentration in the water column which relies on sediment supply, the sediment age calculation can vary with different sediment loading. We conducted experiments (E4 and E5) by specifying a higher and lower sediment concentration at the river discharge locations, respectively. A constant concentration of 50 and $10\,\mathrm{mg\,l^{-1}}$ was specified continuously for 730 days. Other conditions were set as the same as E1.

A comparison between the model results of E1 and E4 is shown in Fig. 10a. It illustrates that the pattern of sediment age distribution does not change much with the higher inflow sediment concentration. An ASMA valley is visible around the salt intrusion limit with a value of 100 days. In the upstream and downstream of the low age valley, the ASMA increases. However, with high sediment loading, the ASMA contours recede upstream in the lower and middle York, representing an increase of ASMA with the increase of sediment loading. With an increased sediment loading, more sediment will be transported to the middle and lower York, which are subject to deposition, resuspension, advection and diffusion. Therefore, large percentages of sediment in the water column come from the seabed. The ASMA thus increases as the sediments are deposited on the seabed and transported gradually to the downstream.

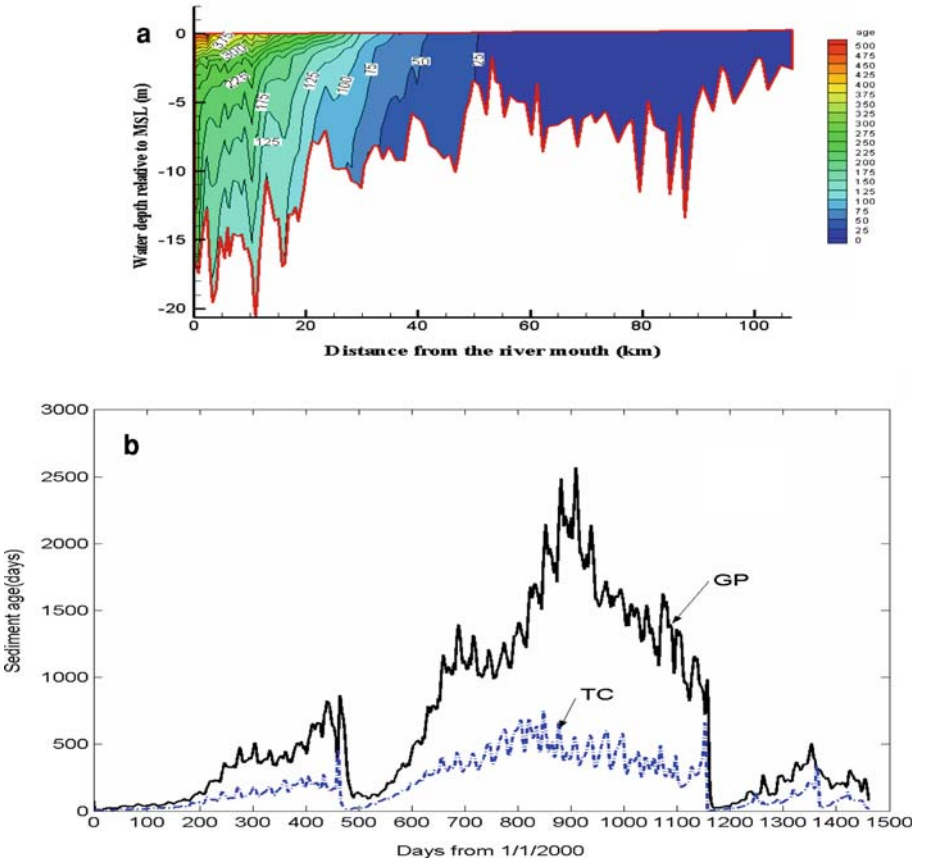


Fig. 9 **a** The distribution of sediment mean age (in days) over the last 15 days along the longitudinal transect under the real time situation; **b** time series of vertically averaged sediment mean ages at GP and TC stations

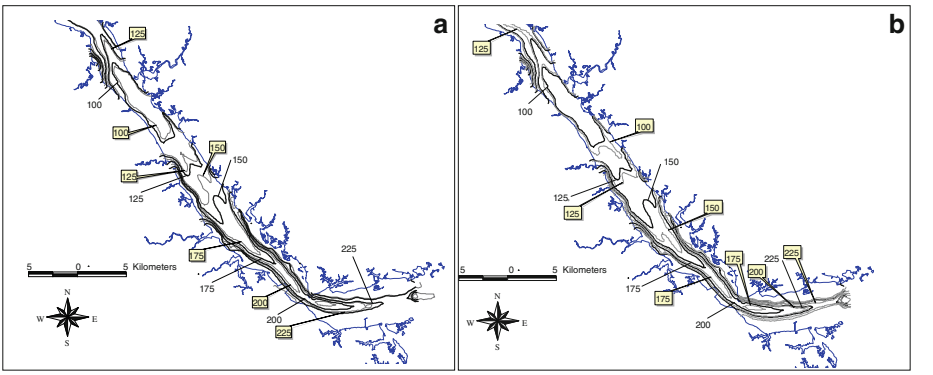


Fig. 10 **a** Comparison of ASMA between experiments E4 and E1 (*coarse solid lines* are the results of E4 and *thin solid lines* of E1); **b** comparison of ASMA between experiment E5 and E1 (*coarse solid lines* the results of experiment E5 and *thin solid lines* are the results of experiment E1)

A comparison between the model results of E5 and E1 is shown in Fig. 10b. In contrast to the result with a higher inflowing sediment concentration (E4), the age contours protrude more downstream in the lower and middle York, and the ASMA decreases compared to that of E1. This suggests less sediment is available at the seabed for a prolonged period.

5.2 The effect of settling velocity

The settling velocity of the sediment plays an important role in the sediment transport and therefore it can significantly impact the sediment age. We did another suite of experiments (E6 and E7) to test this effect. The settling velocity was specified as 1.0×10^{-5} and 1.0×10^{-4} m s^{-1}, respectively. These settling velocities represent those of single silt particle and sediment flocs, respectively.

With the use of smaller settling velocity, the sediment escapes from the estuary much quickly (Fig. 11). It takes approximately 80 days for the sediment particles to transport from the headwaters of the tributaries to the river mouth. The ASMA becomes close to the water age of about 55 days at the river mouth [31]. The longitudinal sediment mean age distribution (not shown here) indicates that the sediment age increases continuously from the upstream to the river mouth and the low age valley disappears. The lateral distribution with higher age in the side shoals and lower age at the channel still prevails.

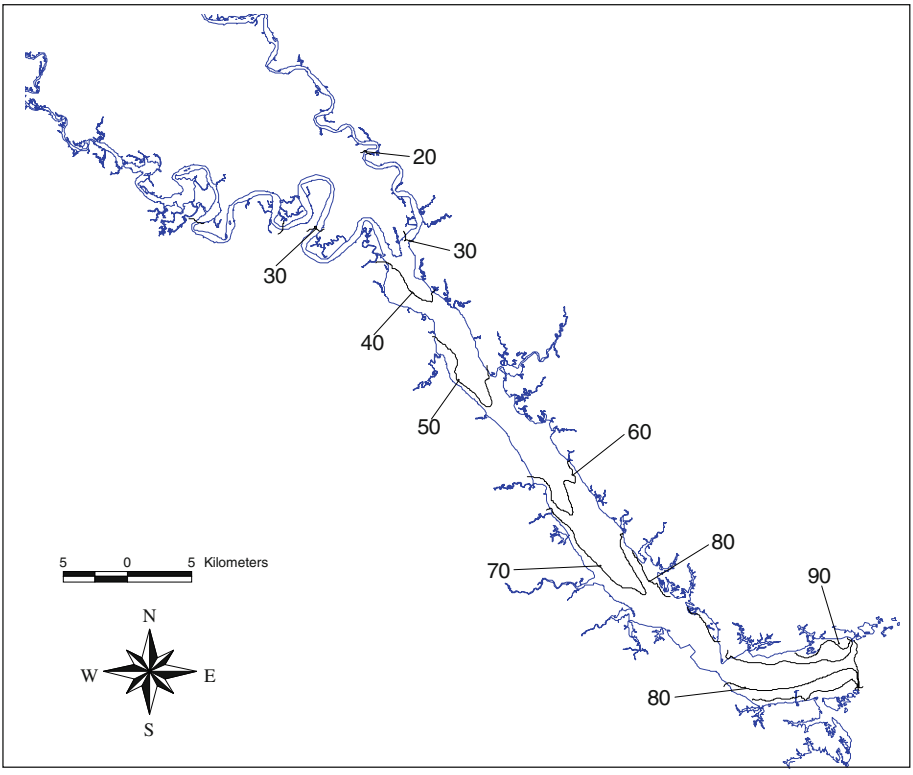


Fig. 11 The distribution of ASMA (in days) with decrease of settling velocity (exp. E6)

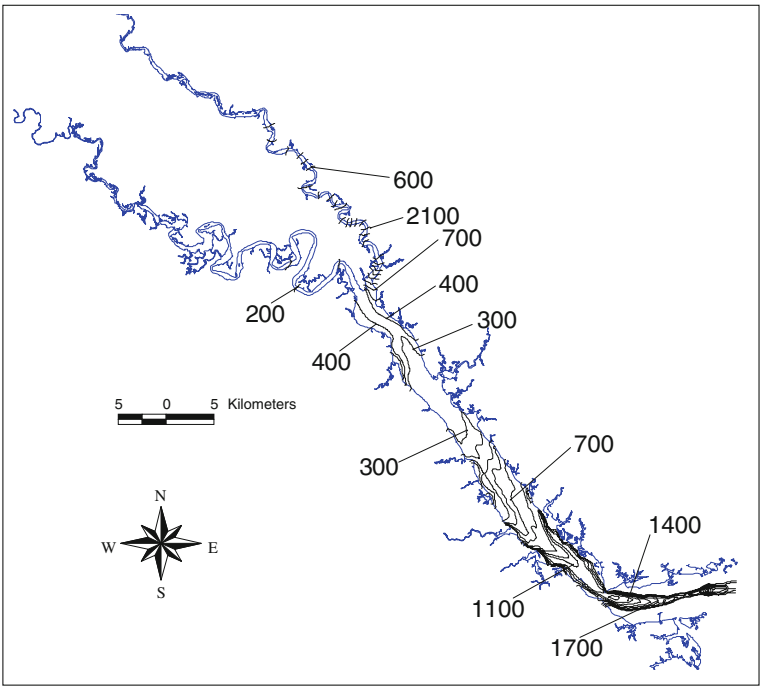


Fig. 12 The distribution of ASMA (in days) with increase of settling velocity (exp. E7)

With a larger settling velocity, the sediment particles need much longer time to transport to the river mouth as sediments experience more frequent deposition and resuspension. The sediment age at the river mouth increases tremendously to 1,700 days, which is an approximately 6 folder increase compared to that of E1 (Fig. 12). This signifies that the sediment flocs would be unlikely to escape from the estuary. The longitudinal distribution of the sediment mean age reveals that the low age valley with a value of 300 days appears again (Fig. 12).

5.3 The effect of critical shear stress for erosion

Similar to the settling velocity, the critical shear stress for erosion is important in determining the sediment transport timescale. Another model test (Experiment E8) was conducted to examine this effect. The critical shear stress for erosion was reduced to be 0.05 Pa and the other parameters were unchanged. The model result is shown in Fig. 13.

With smaller critical shear stress for erosion, sediment deposited at the seabed becomes easier to be eroded, thus the deposition period at the seabed decreases. Therefore, the ASMA in the estuary decreases in general. The sediment particles take approximately 200 days to transport from the headwaters to the river mouth. The spatial distribution pattern of the ASMA is much similar to that of E1. The longitudinal distribution (not shown here) shows the same pattern as that of E1, the age valley of 50 days is located around the salinity intrusion limit area.

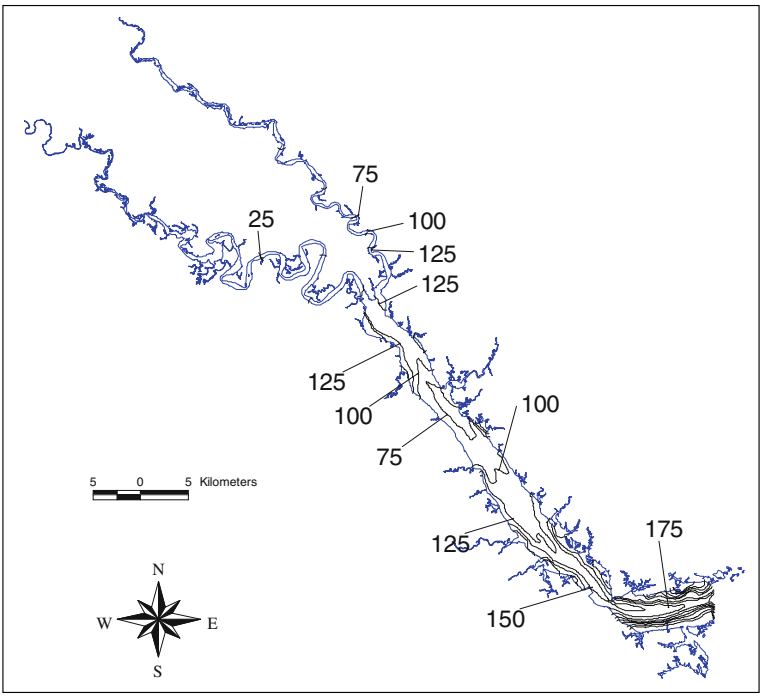


Fig. 13 The distribution of ASMA (in days) with reduction of critical shear stress (exp. E8)

6 Summary and conclusion

This study was designed to quantitatively investigate the river-borne sediment transport timescale in an estuary using the general concept of age. The EFDC was utilized for the numerical study. A sediment transport module [22] was used for modeling the sediment concentration, and the age concentration was simulated after the work of Mercier and Delhez [26]. The sediment age in this study resembles the "transport age" proposed by Mercier and Delhez [26]. As indicated by Lin and Kuo [22], there are some limitations of the sediment model. Constant values of critical shear stresses for erosion and deposition were applied over the entire model domain while varying bed and suspended sediment size distributions were observed in the York River [7,28], and in-situ measurements indicated that the critical shear stress for erosion vary with space and time [10]. The processes of flocculation and bed sediment consolidation were not considered, which may play important roles in the sediment transport in the estuary. Therefore, this study can not be considered as an accurate simulation of the sediment transport in the York River Estuary, rather it should be considered as a diagnostic study of the river-borne sediment transport timescale under ideal conditions.

The diagnostic study suggests that the river-borne sediment age is largely dependent on the river discharge. A high river flow leads to drastic decrease of river-borne sediment age as demonstrated in our real-time sediment simulation. A constant high flow condition corresponds to a relatively low ASMA of approximately 250 days in the estuary, while a mean flow condition results in a highly increased ASMA on the order of 2,000 days. The longitudinal distribution of sediment age shows the presence of an age valley near the salt intrusion limit.

The lower sediment age is distributed in the channel, while the higher age is located in the side shoal areas. The vertical distribution shows a pattern with higher sediment age at the surface and a lower sediment age at the bottom which results from sediment deposition and resuspension processes.

The river-borne sediment age is greatly influenced by the sediment settling velocity. A smaller settling velocity can lead to the sediment particles escaping from the estuary quickly, while a larger settling velocity will entrap the sediment particles in the estuary for a long period of time. With a low settling velocity, the sediment age becomes more similar to the water age for a conservative dissolved tracer. In another aspect, the critical shear stress for erosion also makes a difference for the sediment transport timescale. With a low critical shear stress, the sediment deposited on the bottom will be eroded quickly, and the sediment age at the seabed decrease. Consequently, the sediment age in the water column decreases.

Acknowledgements The first author was supported by the "one-hundred-scholar" program of Sun Yat-Sen University, China during the manuscript's completion. Part of this work was done during the first author's visit in Virginia Institute of Marine Science, The College of William and Mary. The authors would like to appreciate Mr. Mac Sisson for his help in improving the manuscript. Dr. Daoru Wang of Hainan Marine Development Plan and Design Research Institute, China is also thanked for providing some computing facilities during the first author's staying in Hainan, China.

References

1. Beckers JM, Delhez E, Deleersnijder E (2001) Some properties of generalized age-distribution equations in fluid dynamics. J Appl Math 61(5):1526–1544
2. Carpenter SR, Caraco NF, Correll DL, Howarth RW, Sharpley AN, Smith VH (1998) Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol Appl 8: 559–568
3. Deleersnijder E, Campin JM, Delhez EJM (2001) The concept of age in marine modeling. I. Theory and preliminary model results. J Mar Syst 28: 229–267
4. Delhez EJM, Deleersnijder E (2002) The concept of age in marine modeling. II. Concentration distribution function in the English Channel and the North Sea. J Mar Syst 31: 279–297
5. Delhez EJM, Campin J-M, Hirst AC, Deleersnijder E (1999) Toward a general theory of the age in ocean modeling. Ocean Model 1: 17–27
6. Delhez EJM, Lacroix G, Deleersnijder E (2004) The age as a diagnostic of the dynamics of marine ecosystem models. Ocean Dyn 54: 221–231
7. Dellapenna TM (1999) Fine-scale strata formation in biologically and physically dominated estuarine systems within the lower Chesapeake and York River subestuary. Ph.D dissertation, Virginia Institute of Marine Science, the College of William and Mary, Gloucester Point, Virginia
8. Dellapenna TM, Kuehl SA, Schaffner LC (1998) Sea-bed mixing and particle residence times in biologically and physically dominated estuarine systems: a comparison of lower Chesapeake Bay and the York River subestuary. Estuar Coast Shelf Sci 46: 777–795
9. Dellapenna TM, Kuehl SA, Schaffner LC (2003) Ephemeral deposition, seabed mixing and fine-scale strata formation in the York River Estuary, Chesapeake Bay. Estuar Coast Shelf Sci 58: 621–643
10. Dickhudt P, Friedrichs CT, Schaffner LC (2007) Seasonal variability and controls on erosion in a partially mixed estuary. In: The 9th INTERCOH conference, Brest, France, 25–28 September 2007
11. Faas RW (1973) Sedimentational regimes of the York River, Southeastern Virginia, as shown by mass properties. Chesap Sci 14(3): 181–187
12. Galperin B, Kantha LH, Hassid S, Rosati A (1988) A quasi-equilibrium turbulent energy model for geophysical flows. J Atmos Sci 45: 55–62
13. Geyer WR (1997) Influence of wind on dynamics and flushing of shallow estuaries. Estuar Coast Shelf Sci 44: 713–722
14. Gong W, Shen J (2009) Response of sediment dynamics in the York River Estuary, USA to tropical cyclone Isabel of 2003. Estuar Coast Shelf Sci. doi:10.1016/j.ecss.2009.06.004
15. Gourgue O, Deleersnijder E, White L (2007) Toward a generic method for studying water renewal, with application to the epilimnion of Lake Tanganyika. Estuar Coast Shelf Sci 74: 628–640
16. Gustafsson KE, Bendtsen J (2007) Elucidating the dynamics and mixing agents of a shallow fjord through age tracer modeling. Estuar Coast Shelf Sci 74(4): 641–654

17. Haas LW (1977) The effect of spring-neap tidal cycle on the vertical salinity structure of the James, York, and Rappahannock Rivers, Virginia, USA. Estuar Coast Shelf Sci 5: 485–496
18. Hamrick JM (1992) A three-dimensional environmental fluid dynamics computer code: theoretical and computational aspects. Special Report in Applied Marine Science and Ocean Engineering, No. 317. College of William and Mary, VIMS, 63 pp
19. Holt JT, James ID (1999) A simulation of the southern North Sea in comparison with measurements from the North Sea project. Part 2: suspended particulate matter. Cont Shelf Res 19: 1617–1642
20. Kniskern TA, Kuehl SA (2003) Spatial and temporal variability of seabed disturbance in the York River subestuary. Estuar Coast Shelf Sci 58: 37–55
21. Lin J, Kuo AY (2001) Secondary turbidity maximum in a partially mixed microtidal estuary. Estuaries 24: 707–720
22. Lin J, Kuo AY (2003) A model study of turbidity maxima in the York River Estuary, Virginia. Estuaries 26(5): 1269–1280
23. Liu JT, Chao S-Y, Hsu RT (2002) Numerical modelling study of sediment dispersal by a river plume. Cont Shelf Res 22: 1745–1773
24. McCarthy RK (1993) Residual currents in tidally dominated, well-mixed estuaries. Tellus 45A: 325–340
25. Mellor GL, Yamada T (1982) Development of a turbulence closure model for geophysical fluid problems. Rev Geophys Space Phys 20: 851–875
26. Mercier C, Delhez EJM (2007) Diagnosis of the sediment transport in the Belgian Coastal Zone. Estuar Coast Shelf Sci 74: 670–683
27. Morris AW, Howarth MJ (1998) Bed stress induced sediment resuspension (SERE 88/89). Cont Shelf Res 18: 1203–1213
28. Nichols MM, Kim S-C, Brouwer CM (1991) Sediment characterization of the Chesapeake Bay and its tributaries. NOAA National Estuarine Inventory, Virginia Institute of Marine Science, the College of William and Mary, Gloucester Point
29. Nixon SW (1995) Coastal marine eutrophication: a definition, socialcauses, and future consequences. Ophelia 41: 199–219
30. Scully ME, Friedrichs C, Brubaker J (2005) Control of estuarine stratification and mixing by wind-induced straining of the estuarine density field. Estuaries 28(3): 321–326
31. Shen J, Haas L (2004) Calculating age and residence time in the tidal York River using three-dimensional model experiments. Estuar Coast Shelf Sci 61: 449–461
32. Shen J, Lin J (2006) Modeling study of the influences of tide and stratification on age of water in the tidal James River. Estuar Coast Shelf Sci 68(1–2): 101–112
33. Shen J, Wang HV (2007) Determining the age of water and long-term transport timescale of the Chesapeake Bay. Estuar Coast Shelf Sci 74: 585–598
34. Shen J, Sisson M, Kuo AY, Boon J, Kim S (1997) Three-dimensional numerical modeling of the tidal York River system, Virginia. In: Spaulding ML, Blumberg AF (eds) Estuarine and coastal modeling. Proceedings of the fifth international conference, Alexandria, Virginia, USA, pp 495–510
35. Sisson GM, Shen J, Kim S-C, Boon J, Kuo AY (1997) VIMS Three dimensional hydrodynamic-eutrophication Model (HEM-3D): application of the hydrodynamic model to York River system. Special report in Applied Marine Science and Ocean Engineering, No. 341. Virginia Institute of Marine Science, The College of William and Mary, Gloucester Point, Virginia
36. Takeoka H (1984) Fundamental concepts of exchange and transport time scales in a coastal sea. Cont Shelf Res 3(3): 322–326
37. White L, Deleersnijder E (2007) Diagnoses of vertical transport in a three dimensional finite-element model of the tidal circulation around an island. Estuar Coast Shelf Sci 74(4): 735–749
38. Zimmerman JTF (1976) Mixing and flushing of tidal embayments in the Western Dutch Wadden Sea, part I: distribution of salinity and calculation of mixing time scales. Neth J Sea Res 10: 149–191

Environ Fluid Mech (2010) 10:197–211
DOI 10.1007/s10652-009-9143-6

ORIGINAL ARTICLE

Estimating river flow effects on water ages by hydrodynamic modeling in Little Manatee River estuary, Florida, USA

Wenrui Huang · Xiaohai Liu · Xinjian Chen · Michael S. Flannery

Received: 1 February 2009 / Accepted: 6 July 2009 / Published online: 28 July 2009

Abstract The water age in a tidal river in Florida, Little Manatee River, has been investigated in this study by the application of a three-dimensional hydrodynamic model. In response to a pulse dye release in the upper end of the river boundary, the hydrodynamic model determines the water age for a given location by recording the time for the dye to reach the given river location. The hydrodynamic model uses horizontal curvilinear orthogonal grids to represent the complex river system that includes several bayous and tributaries. The model has was calibrated and verified in previous study by using two continuous data sets for a 6 month period. Satisfactory model verifications indicate that the hydrodynamic model is capable of quantifying the mixing and transport process for calculating the water age in the tidal river. For 17 freshwater inflow scenarios in the Little Manatee River, the hydrodynamic model was applied to simulate water ages along the main channel of the river at 2-km interval. Flow rates in the 17 scenarios varying from 0.26 to 76.56 m^3/s cover the range of the observed flows in the Little Manatee River. Water ages from model predictions range from the minimum 1.2 days under the maximum 76.56 m^3/s inflow condition to the 50 days under the minimum 0.26 m^3/s inflow condition. Empirical regression equations at three selected stations, with the correlation coefficient R^2 above 0.96, were derived from numerical model simulations to correlate water ages to freshwater inflows. The empirical water-age equation derived from hydrodynamic model simulations can be used to provide quick and low-cost estimations of water ages in response to various inflow scenarios for studying physical–chemical and biological processes in the river.

Keywords Water age · Hydrodynamic model · Estuary · Tidal river · Mixing time scale

W. Huang (✉) · X. Liu
Department of Civil and Environmental Engineering, Florida State University, Tallahassee, FL 32310, USA
e-mail: whuang@eng.fsu.edu

X. Chen · M. S. Flannery
Southwest Florida Water Management District, Tampa, FL 33637-6759, USA

1 Introduction

Water age is one of mixing and transport time scales, which is important for the assessment of estuarine mixing and water quality [3–5,18,26]. Zimmerman [26] defines water age as the time a water parcel has spent since entering the estuary through one of the boundaries. Deleersnijder et al. [3] explained more details regarding the definition of water age in marine environment. Transport time scales are useful tools to quantify the importance of hydrodynamic processes in the transport and fate of pollutants in coastal and estuarine water systems, because most of the living biomass and masses of nutrients, contaminants, and suspended particles are carried in a fluid medium in aquatic systems. For an estuarine system, retention of pollutants in the system and time required to remove pollutants out of the system are critical to the water quality. Correlation between retention time scales and water quality has been described by Monsen et al. [18] and Takeoka [24]. Guyondet et al. [8] investigated the effects of water renewal on the oyster aquaculture potential of the Richibucto estuary, New Brunswick, Canada, indicating that the region could sustain a high oyster biomass in the estuary. Andrejev et al. [1] studied the age and renewal time of water masses in a semi-enclosed basin in the Gulf of Finland. Kratzer and Biagtan [15] conducted dye-tracer studies during 1994–1995 to determine travel times in the lower San Joaquin River Basin of California.

Transport time scales and water renewal in estuaries usually have a complex spatial distribution because of the effects of complex hydrodynamic processes and irregular topographic features. Based on the definition of water age, particles at different locations within a waterbody have different ages. Consider a pulse dye release study in a one-dimensional river. As the pulse of dye travels downstream, dye is observed at different stations at different times. Therefore, the average dye mass age at a downstream location will be greater than the average dye mass age at an upstream station. In addition, dispersion causes the initial scalar spike to spread. Therefore, at each site some mass arrives having a greater than average age while other mass arrives having a less than average age. The development of hydrodynamic models in the past two decades has fortunately provided effective tools to quantify the transport time scales. Three dimensional estuarine hydrodynamic models, if appropriately calibrated and verified, are capable of simulating the mixing and transport time scales in estuarine and marine systems. Successful examples include Huang and Spaulding's [12] modeling study of residence time in Apalachicola Bay, and Huang's [11] study of flushing time modeling in North Bay of Florida. Successful applications on modeling water ages have been demonstrated by some researchers in recent years. Shen and Haas [21] used three-dimensional model experiments to estimate water ages and residence times in the tidal York River. Shen and Wang [22] determining the age of water and long-term transport timescale of the Chesapeake Bay by conducting three dimensional modeling study. Ribbe et al. [20] assessed water renewal time scales for marine environments from three-dimensional modeling in a case study for Hervey Bay, Australia. England [5] discussed water age and ventilation time scale in a global ocean model. Gourguea et al. [7] described a generic method for studying water renewal, with application to the epilimnion of Lake Tanganyika.

In this paper, an application of a three dimensional hydrodynamic model to estimate water age distributions for a range of inflow conditions in Little Manatee River in Florida is presented. The hydrodynamic model was previously calibrated and verified by field observations to quantify the mixing and transport processes in the estuarine system under wind, tides, and freshwater inflows. To support water resources management, seventeen inflow scenarios were selected to characterize low, normal, and high flow conditions in the river system. Model simulations of water ages, for a pulse of dye entering into the river from the most upstream end, were conducted for stations at 2 km interval along the main channel of the

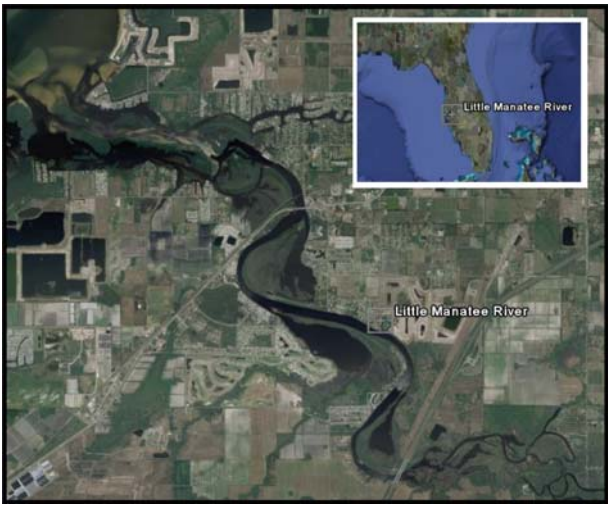


Fig. 1 Little Manatee River in Florida, USA

river. Empirical regression equations were obtained to correlate water ages at three stations (lower, mid, and upper river stations). Results from this study are intended to support water resources management in determining the minimum flow condition, at which further water withdrawals from the river would be significantly harmful to the water resources or ecology in the estuarine system.

2 Study site: Little manatee river, Florida, USA

The Little Manatee River (LMR) is a tidal river estuarine system located in the southwestern portion of Tampa Bay on the gulf coast of west-central Florida (Fig. 1). The river, which has a watershed of about 575 km^2, flows generally westward for about 65 km toward its discharge point into Tampa Bay near Ruskin. The Little Manatee River is a shallow water system, including several branches and bayous. The primary land use in the watershed is agriculture, including pasture and rangeland, row crop, citrus and fish farming. The tidal river estuary receives freshwater inflows from multiple input locations, including main upstream channel of the river, small ungaged streams, and other rainfall runoff from watershed. Although tidal effects on water levels are discernable at the USGS flow gage at Highway 301 (about 24 km from the river mouth), brackish waters are typically restricted to the lowermost 16 km in the river during dry seasons [19].

Freshwater inflow is a dominant factor affects the physical, chemical, and biological characteristics of estuaries. The Southwest Florida Water Management District (SWFWMD) manages water resources to balance the needs of current and future water users while protecting the ecological integrity of natural water systems, including Little Manatee River system. Florida Statutes require the state's water management districts establish rules for minimum flows and water levels for priority water bodies within their respective jurisdictional boundaries. Minimum flows for a given watercourse are defined as "the limit at which further withdrawals would be significantly harmful to the water resources or ecology of the area". The Little Manatee River is one of the priority water bodies for which minimum flow rules are

to be adopted. Accordingly, the Southwest Florida Water Management District (SWFWMD) evaluates the freshwater inflow requirements of receiving estuaries when determining minimum flows for coastal rivers. Many fish species are estuarine dependent, spending part of their development in an estuarine environment. When a river or stream drops below its minimum flow, the aquatic life of that ecosystem can be harmed. The inflows from rivers and streams affect the mixing and transport process, and the time required to remove potential pollutant spills in estuaries. Therefore, adequate river inflows are important for maintaining good water quality in estuaries. The water age study will provide necessary information for biologists to determine the minimum and critical river inflow that may cause adverse impact on estuarine aquatic ecosystem.

3 Description of hydrodynamic model for Little Manatee River

The Environmental Fluid Dynamics Code (EFDC) developed by Hamrick [9] was applied to Little Manatee River. The EFDC model is a general-purpose modeling package for simulating one- or multi-dimensional flow, transport, and bio-geochemical processes in surface water systems including rivers, lakes, estuaries, reservoirs, wetlands, and coastal regions. This model is supported by the U.S. Environmental Protection Agency (EPA) and has been used extensively throughout the USA [25]. In addition to hydrodynamic, salinity, and temperature transport simulation capabilities, EFDC is capable of simulating near field and far field discharge dilution from multiple sources, eutrophication processes, the transport and fate of toxic contaminants in the water and sediment phases. The computational schemes in the EFDC model are equivalent to the widely used Princeton Ocean Model (POM) by Blumberg and Mellor [2] in many aspects. The EFDC model uses sigma vertical coordinate and curvilinear orthogonal horizontal coordinates. It employs second order accurate spatial finite differencing on a staggered or C grid to solve the equations of momentum, while time integration is implemented using a second order accurate three-time level, finite difference scheme with an internal–external mode splitting procedure to separate the internal shear or baroclinic mode from the external free surface gravity wave or barotropic mode. The external mode solution is semi-implicit and simultaneously computes the two-dimensional (2-D) surface elevation field by a preconditioned conjugate gradient procedure. The external solution is completed by the calculation of the depth-averaged barotropic velocities using the new surface elevation field. The model's semi-implicit external solution allows large time steps that are constrained only by the stability criteria of the explicit central difference or higher order upwind advection scheme. The EFDC model's internal momentum equation solution, at the same time step as the external solution, is implicit with respect to vertical diffusion. The internal solution of the momentum equations is in terms of the vertical profile of shear stress and velocity shear. Time splitting inherent in the three-time-level scheme is controlled by periodic insertion of a second-order accurate two-time-level trapezoidal step. To describe the estuarine mixing process, the hydrodynamic model dynamically couples the Mellor–Yamada level 2.5 turbulence closure [6,16] to solve turbulent kinetic energy and turbulent length scale for estimating the vertical turbulent eddy viscosity and diffusivity. To account for the larger grid sizes in horizontal directions, the EFDC model employs the [23] formulation to explicitly calculate horizontal viscosity and diffusivity coefficients.

The Little Manatee River is a shallow water system. The average water depth varies from approximately 3 m near the river mouth to 1 m in upstream channel of the river. To approximate the meandering channel of the river system, an orthogonal horizontal grid system was developed (Fig. 2). The average grid size is approximately 80 m. The orthogonal curvilinear

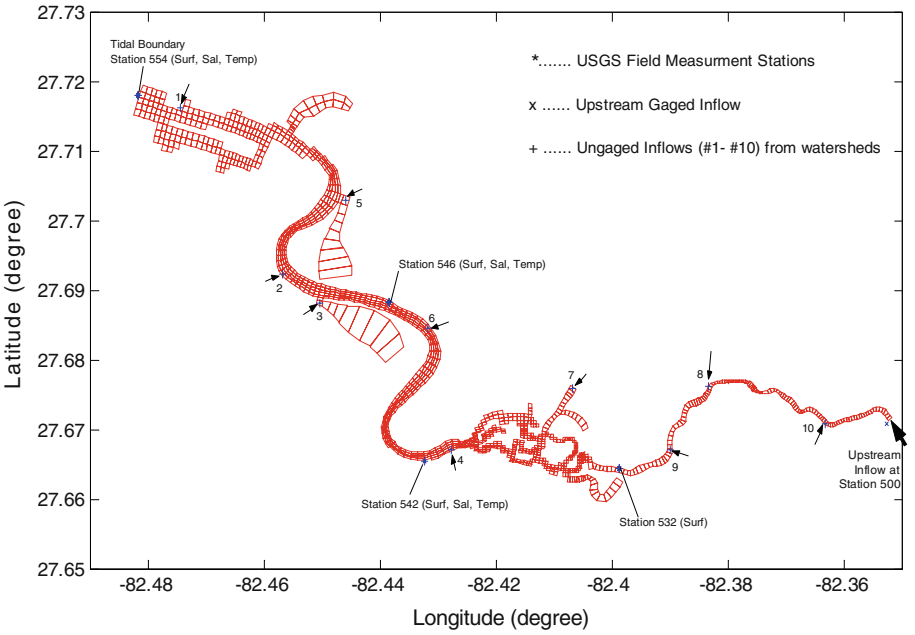


Fig. 2 Model grid system, freshwater inflow locations, and field data monitoring stations in Little Manatee River. The distance from the river mouth to upstream boundary is approximately 19 km

model grid system adequately characterizes the meandering channel and river boundaries including three bayous. Multiple grids were employed along the river cross sections in the river main stem to account for the variations of the river's bathymetry. Three sigma layers were adopted in the vertical dimension to account for the density stratification and bidirectional flows in the river.

A field data collection program was conducted to support model calibration and verification data collections included hourly measurements of water levels, salinity, and water temperature at four stations in the Little Manatee River (Fig. 2). Wind data were measured at the SWFWMD meteorological station near Dover in eastern Hillsborough County, located approximately 16 miles northeast of the mouth of the Little Manatee River. Gaged freshwater inflows from the upstream of the Little Manatee River were obtained from USGS Station #500 in Little Manatee River near Wimauma (site #500), which is located on the main river channel about 24 km above the mouth of the river. Ungaged freshwater flows from basins downstream of the gage Station #500 were produced by HSPF (Hydrological Simulation Program—Fortran) surface water model simulations [14]. The HSPF simulations use inputs of rainfall and other hydrologic information such as land use, evapotransportation, and infiltration to estimate the runoff from the ungaged basins. Freshwater inflows at the upstream USGS gaged station and ungaged flows from two sub-basins from hydrological model simulations are shown in Fig. 3.

During the field data collection period, some data gaps existed due to technical problems in measurement instruments. After careful process and analysis of the observation data, continuous hourly data without gap for the 1/1/2005–6/30/2005 were selected for model calibration and verification. The data were divided into two data sets. Data for the period of January 1–February 28 of 2005 were used in model calibration, while data for the period of

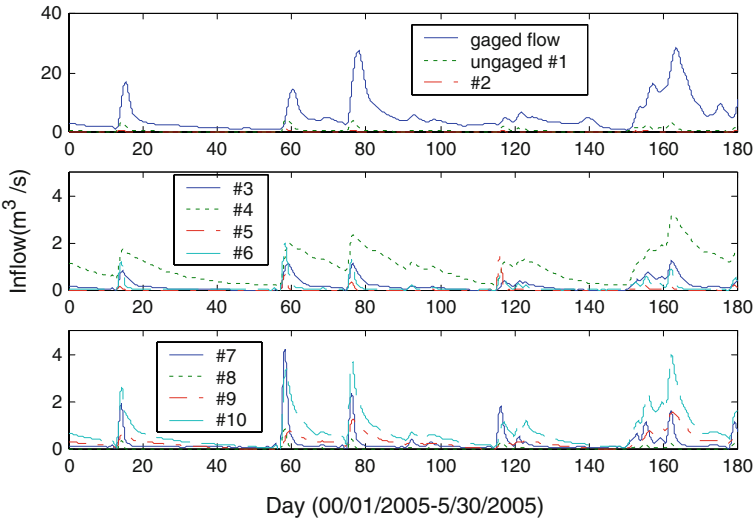


Fig. 3 Gaged inflow from upstream flow station, and HSPF moded ungaged inflows from sub-basins

3/1/2005–6/30/2005 were applied in model verification. For water level verification, coefficients of determination values (R^2) are above 0.94 at all four stations, with root-mean-square errors (RMSE) below 0.003 m. For salinity verification, R^2 values are above 0.94, and the maximum RMSE is 1.9 ppt at all three-observation stations. For water temperature verification, R^2 values are 0.99, 0.96, and 0.93; and the RMSE are 0.37, 1.14, 1.39 degrees, at Station 554, Station 546, and Station 542, respectively. Details of model calibrations and verifications have given in Huang et al. [13]. Results indicate that the calibrated EFDC model provides satisfactory descriptions water level variations, and the mixing and transport processes between fresh and salt waters in the Little Manatee River system. Hourly salinity variations due to tides and the low-frequency trend of salinity change induced by freshwater inputs are reasonably reproduced by the model simulations. Figures 4–5 show that salinity in both Station 546 and Station 542 reduced to 0 ppt during the high flow conditions, indicating the model's capability to predict the mixing in fresh and saline water fronts. In Station 542, salinity was generally below 10 ppt during the model verification period, and declining to near 0 ppt in March and June of 2005. Therefore, the model calibrated with salinity can be used to simulate the mixing and transport of a similar conservative constituent in the Little Manatee River.

Salinity at low tide on the 9th February 2005 resulted from hydrodynamic model simulation is given in Fig. 6. Fresh water moves to downstream at low tide. From the most upstream segment to the area halfway between the upstream and downstream boundary, salinity is below 2 ppt and freshwater is dominant. The freshwater with 5 ppt reaches the lower LMR river area about 1/3 of river length from the lower LMR rover mouth. The salinity at the lower river tidal boundary is about 18 ppt. Salinity at bayous is higher than that in the river main channel at low tide. Figure 6 also demonstrate the hydrodynamic model's capability in calculating freshwater and saline water mixing and transport process in the estuary. Dye is generally considered as a conservative constituent, and its mixing and transport process is similar to the freshwater release in the estuary. Therefore, the movement of dye along the main channel of the river in response to the pulse of initial dye release can be calculated and tracked by the hydrodynamic model simulations.

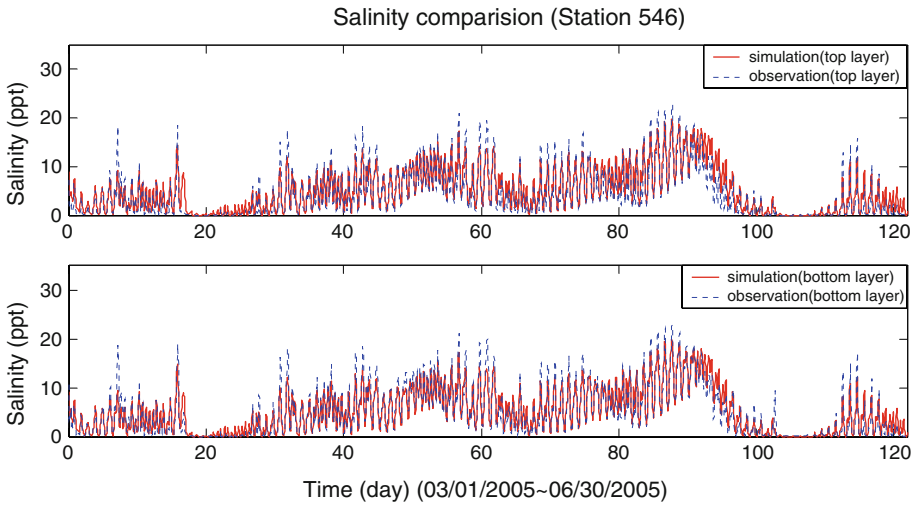


Fig. 4 Comparison of model predicted and observed salinity at Station 546 during model verification period

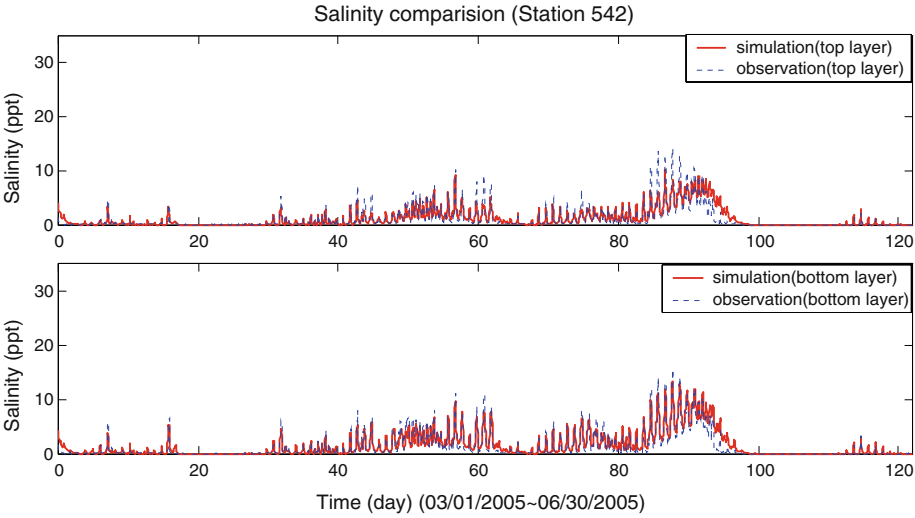


Fig. 5 Comparison of model predicted and observed salinity at Station 542 during model verification period

4 Water ages under various constant inflow scenarios

4.1 Definition of water age in Little Manatee River

Based on Zimmerman's [26] definition, water age at a location in a river is originally defined as the time a "water parcel" has spent since entering the river location through the upstream end boundary. In numerical modeling, the initially released dye will spread because of the dispersion, and will travel toward downstream as a cloud that may spread into several model grids. As the result, some portion of the dye cloud moves faster than others. To approximate the travel time for the dye cloud from the pulse of dye release, the water age in this study

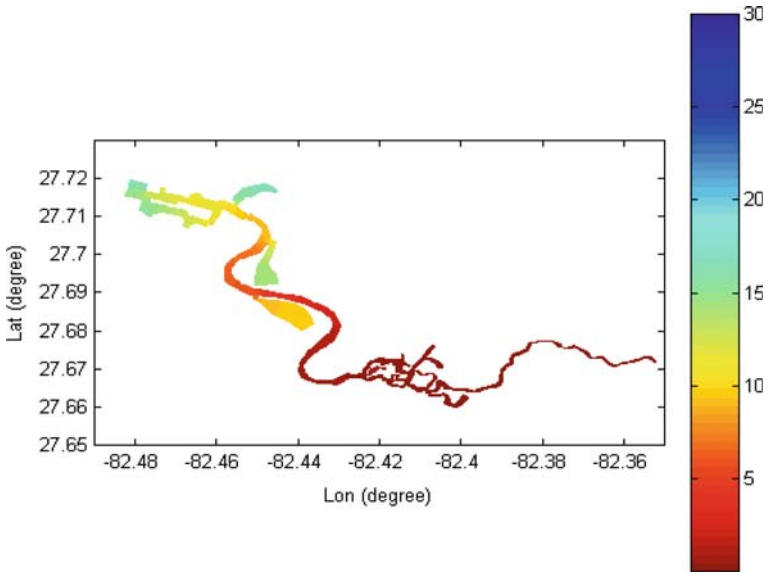


Fig. 6 Salinity field at low tide condition, 2/9/2005

is referred as "median water age" to represent the travel time of the median dye mass. In probability theory and statistics, a "median" is described as the number separating the higher half of a sample, a population, or a probability distribution, from the lower half. Therefore, the median water rage is estimated by tracking the 50% dye mass passing by a river location. Under the idealized condition, when the dye concentration is in normal or Gaussian distribution with the peak concentration in the center, the middle of the dye cloud is the same as the median dye mass (from one end to the center). However, in more complex estuarine mixing and transport, dye concentration may be in non-Gaussian distribution and the peak concentration may not be located in the middle of the dye cloud. In this situation, it is more reasonable to use the "median dye mass" passing by a river location to characterize the center of the dye mass. Using the passage of median dye mass to track the average water age is more reasonable than the use of "peak dye concentration" because the peak concentration may occur near the edge of the dye cloud.

In tidal rivers, the water parcel oscillates upstream-ward and downstream-ward by tidal force with a net movement toward downstream by tidally-averaged residual currents. For examples, time series of salinity variations as shown in Figs. 4–5 indicate the tidal effects at Stations 546 and 542. Figures 4–5 also show the model's capability to predict the mixing and transport processes in the tidal estuary of the LMR system. Because tides cause fluctuations of dye concentrations in the similar ways as salinity variations, water ages are affected by tidal forcing, too. Figure 7 presents the hourly tracking locations of the dye location that shows tide-induced fluctuations. Tidal-induced fluctuation is strongest near the river mouth (the minimum value in the vertical axis), and weakest near the upstream boundary at about 20 km from the river mouth. In order to separate the effects of river flow from tides, tidal effects on water ages were removed by using 24-h moving-average filter, which resulting in the tidally-averaged travel time (or tidally-averaged water age) along the river main channel as shown in Fig. 7.

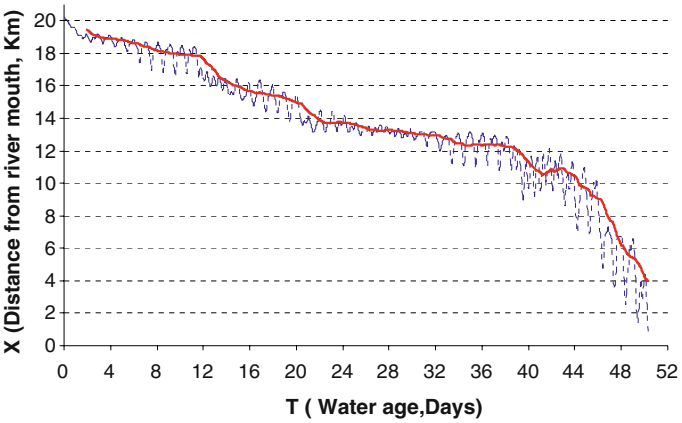


Fig. 7 Tidally-averaged water age along the main channel by using 24 h moving-average filter to remove tidal effects

4.2 Initial and boundary conditions

Using the similar set up as given [17] in modeling a pulse of dye release, dye concentration was initially set to 1 in the most upstream volume elements (only at the main upstream Station 500), and set to zero in all other elements and boundaries (including all ungaged freshwater inputs) in the model grid system. The calibrated hydrodynamic model is capable of calculating the water ages at any locations along the river by determining the travel time it takes for to reach the given location from the upstream boundary. In this study, water age outputs were recorded at locations along the center line of the main channel at 2 km interval from the river mouth to the upstream boundary at USGS flow gage station, in response to the pulse dye release only at the most upstream end boundary at Station 500.

4.3 Seventeen constant inflow scenarios

Seventeen constant gaged river inflow scenarios were selected in hydrodynamic model simulations to examine water ages in the estuarine system. The scenarios were obtained based on the hydrological analysis of long-term flow data. The gaged flow scenarios cover wide range of flow conditions from low flow in drought season to high flow in flood season that are important to environmental and water resources management. In particular, the low and minimum flow scenarios can be used to evaluate the potential environmental impact under drought seasons. Constant ungaged inflows for corresponding gaged flow scenarios, as shown in Table 1, were approximately estimated from the gaged inflows by the percentage as determined from the regression analysis [13]. The total constant inflows were obtained by summing the gaged inflows and all un-gaged inflows from the watersheds. As shown in Table 2, total inflows vary from the minimum 0.26 m^3/s to the maximum 76.6 m^3/s.

4.4 Tidally-averaged water age along the main channel under different inflow conditions

For each constant inflow scenario given in Table 2, hydrodynamic model simulation was conducted, and tidally-averaged water ages were derived by filtering the tidal effects by a 24-h moving-average filter. Tidally-averaged water ages are presented in Table 2. For selected flow scenarios, water ages at different locations along the main channel are presented in Fig. 8.

Table 1 Ratio of ungaged inflow to gaged inflow

	Gaged inflow (m^3/s)	Ungaged inflow at different sub-basins of the watershed (estimated by the percentage of the gaged inflow)									
#	Q_{gaged}	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
%	100%	9%	1%	3%	1.6%	1%	2%	4%	0	5%	11%

Table 2 Tidally-averaged water age $\boldsymbol{T}$ at different locations along the main channel of the river under different constant inflow $\boldsymbol{Q}$ scenarios

Scenario	Upstream gaged inflow (m^3/s)	Total inflow Q(m^3/s)	Water age T (days) at different distance (km) from the river mouth									
			1	3	5	7	9	11	13	15	17	19
1	0.20	0.26	50.0	49.5	49.0	46.8	44.4	39.8	29.1	19.2	13.1	3.1
2	0.30	0.52	39.9	39.6	38.8	36.8	34.5	30.2	20.8	12.5	8.6	2.5
3	0.50	0.78	32.5	32.3	31.5	29.8	27.5	23.5	14.2	9.6	5.5	1.9
4	0.60	0.95	31.3	31.2	30.4	29.2	26.4	22.9	13.6	9.3	5.4	1.8
5	0.80	1.16	28.5	28.5	27.9	26.8	23.5	20.5	12.1	7.2	4.5	1.7
6	0.90	1.38	27.2	27.1	26.6	25.5	22.3	19.5	11.9	6.3	4.1	1.6
7	1.00	1.55	24.4	24.4	23.8	22.8	19.6	17.0	10.1	5.6	3.9	1.5
8	1.20	1.76	22.7	22.5	22.0	21.2	17.9	15.6	9.6	5.2	3.3	1.5
9	1.30	2.02	21.8	21.5	21.3	20.5	17.1	15.0	9.2	4.7	3.1	1.4
10	1.50	2.33	20.9	20.7	20.4	19.8	16.8	14.5	9.0	4.2	3.1	1.3
11	1.80	2.71	19.9	19.6	19.4	19	16.1	13.9	8.3	3.9	2.6	1.2
12	2.40	3.66	15.7	15.5	15.3	14.9	12.4	10.3	4.6	3.4	1.7	1.0
13	3.50	5.38	11.1	11.0	10.8	10.6	8.4	6.6	3.7	2.5	1.5	0.8
14	5.70	8.65	7.3	7.0	6.9	6.6	5.2	3.4	1.6	1.1	0.6	0.3
15	11.50	17.52	3.7	3.3	3.2	2.9	2.1	1.6	1.1	0.7	0.4	0.3
16	20.10	30.53	2.0	1.8	1.6	1.3	1.1	1.0	0.5	0.5	0.3	0.2
17	50.40	76.65	1.2	1.0	0.7	0.6	0.6	0.4	0.3	0.2	0.2	0.1

Note: 1. Total constant inflow scenario = constant gaged inflow at Station 500 + all constant ungaged flows. 2. Ungaged constant inflows were approximately estimated from the gaged inflows by the percentage as determined from the regression analysis [13]

Water ages nonlinearly vary in term of distance along the main channel. This is due mainly to the velocity variations along the channel as the results of the changes in cross section areas. In addition, the travel time not only depends on the advection transport by velocity, but also is affected by non-linear diffusive transport. The maximum water age, 50 days, occurs at the location 1 km from the river mouth under the minimum river inflow, 0.26 m^3/s. Water ages decrease as the locations approach upstream with increasing distance from the river mouth. Water age variations in the downstream channel area portion are generally less than those in the upstream channel area. This is because of the weaker currents as the result of wider channel cross sections in the down stream channel. Because the water age defined in this study referred to the time when inflow water entered the upstream boundary at USGS gaged flow station, the minimum water age (0 day) occurs at the upstream inflow boundary.

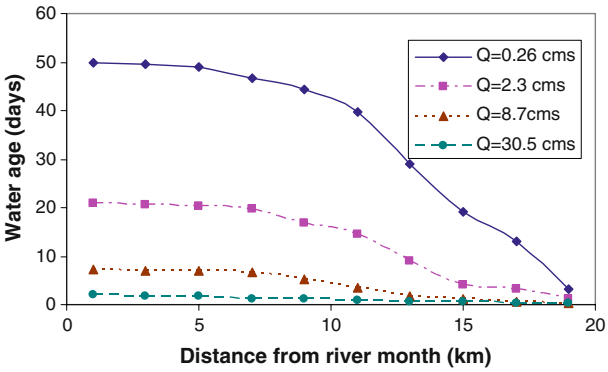


Fig. 8 Tidally-averaged water age at different locations along the main channel of river for a pulse release of tracer at upstream end at USGS station. River inflow unit is cubic meter per second (cms)

At a given location, changes of water ages are different at different flow conditions. Under high flow condition, water ages are very low. The water ages at 1 km from the river mouth are only 2 days under 30.5 m^3/s high flow condition in comparison to the 50 days under 0.26 m^3/slow flow condition. Because the large increase of water ages occurs under low flow conditions, attentions should be paid to low flow conditions in environmental and water resources management to avoid adverse impact to the aquatic ecosystem.

4.5 Correlation between constant river inflow and tidally-averaged water age

In order to investigate the correlations between water ages and inflows, water ages at three locations were plotted against inflows as shown in Fig. 9. The first station, 1 km from the river mouth, can be used to characterize water ages in the lower river area. The 2nd station, 11 km from the river mouth, represents the mid river area. The third station, 17 km upstream from the river mouth, is located in the upper river portion. In all three stations, data obtained from hydrodynamic model simulations indicate that water ages reduce in response to river inflows in a similar pattern. As inflows decrease from 76.6 to 8.6 m^3/s, water ages gradually increase. When inflows further reduce to below 5 m^3/s, water ages substantially increase in terms of flow reductions. Least-square regression analysis was conducted. Empirical best-fitting equations were obtained in forms of power law. The power-law empirical equations fit well with the water ages resulted from the hydrodynamic model simulations, with correlation R-squared values above 0.96 in all three locations. At the upstream location, 17 km from the river mouth, the R^2 value is the highest with a 0.98 value because of very weak tidal effects. At downstream location near the river mouth, where tidal effects are strong, the regression fit ($R^2 = 0.96$) deviate from data when inflows are less than 1 m^2/s.

Empirical regression equations of water ages $\boldsymbol{T}$ (days) verse total river flows $\boldsymbol{Q}$ (m^3/s) are given below:

(a) Location 1 km from the river mouth

$$\boldsymbol{T} = 30.676\boldsymbol{Q}^{-0.7028} \quad (Q > 0, R^2 = 0.962) \tag{1}$$

(b) Location 11 km from the river mouth

$$\boldsymbol{T} = 22.361\boldsymbol{Q}^{-0.8486} \quad (Q > 0, R^2 = 0.9602) \tag{2}$$

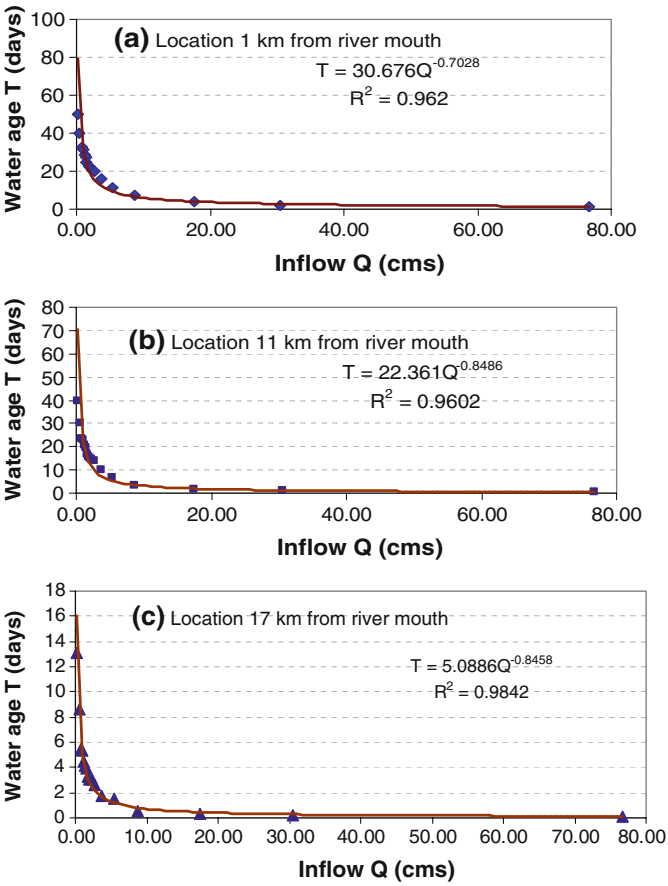


Fig. 9 Tidally-averaged water ages T (days) in responses to river inflows $Q(\mathrm{m}^3/\mathrm{s})$ at different locations along the river main channel. Least-square regressions were done by using power law. Data were obtained from hydrodynamic model simulations

(c) Location 17 km from the river mouth

$$T = 5.0886\,Q^{-0.8458} \quad (\mathrm{Q} > 0, \mathrm{R}^2 = 0.9842) \tag{3}$$

While the hydrodynamic model can be used to simulate water ages for other inflow scenarios at any locations of the river, the empirical equations can provide quick and low-cost preliminary estimations of water ages at selected stations to support environmental and water resources managements. As shown from Fig. 9, water ages substantially increase in terms of flow reduction in low flow conditions, which usually occur or in drought seasons. Therefore, adequate inflows should be maintained in water resources management in the Little Manatee River to control water ages under certain values to preserve the estuarine environmental and ecological system. It should be noticed that the water age scenarios discussed in this study are assumed to be constant inflows.

However, naturally occurring inflow rates are often not near a constant value, which may alter the water age. To approximately estimate water ages by the regression equations, time

series of inflows can be divided into several zones of "quasi-steady flow conditions" so that flows in each zone do not substantially vary. For the time series inflows in Little Manatee River as shown in Fig. 3, inflows do not significantly vary over the most of the period except storm events. Therefore, constant inflow approximations can be used to some flow periods. For example, for the 50-day period from days 90 to 140 (3/31–5/1/2005), constant averaged flow ($3.7\,m^3/s$) can be use to approximate the slightly varied flows (minimum $2.5\,m^3/s$–maximum $4.2\,m^3/s$) for water age simulations. Because water age for $3.5\,m^3/s$ gaged inflow is only 11 days, the use of constant inflow is a reasonable estimation for 50-day period between 3/31 and 5/1/2005. For the storm event between days 77 and 79 (3/18–3/20), gaged flow varied from 23.2 to $27.3\,m^3/s$. Because the maximum water age for $20\,m^3/s$ gaged inflow is only 2 days, it is a reasonable approximation to use the constant averaged flow for the 3-day period 3/18–3/20 to estimate the water age during the storm event.

In case that flow changes significantly, the hydrodynamic model can be readily used to predict water ages under any given time series of unsteady inflows conditions. The existing hydrodynamic model is readily for simulating any unsteady inflow conditions in the same way as the constant inflow conditions without any technical difficulty. However, the limit for simulating time-dependent flow is that it is case by case simulations, and it is difficult to show how water ages response to different unsteady inflows.

5 Conclusions

The tidally-averaged water age has been calculated through the application of a previously calibrated 3D hydrodynamic model to the Little Manatee River estuary. The model employs horizontal curvilinear orthogonal grids to represent the complex tidal river system that consists of bayous and tributaries. Water age in the Little Manatee River is estimated in the hydrodynamic model by tracking the time the dye has spent since its entering a location in the estuary through the boundary at upstream end. Water ages at selected stations along the river main channel at 2-km interval were studied for 17 constant river inflow scenarios through hydrodynamic model simulations. For the 17 inflow scenarios ranging from the minimum $0.26\,m^3/s$ to the maximum $76.60\,m^3/s$ of the total inflows, water ages varied from 1.2 to 50 days. For a given river inflow condition, the maximum water age occurs near the most downstream location because the longest travel distance to the upstream boundary of dye release. The water age decreases as the location approaches to the upstream dye-input boundary. The water age increases in response to the decrease of river inflow. When river inflow decreases to $5\,m^3/s$ or less, water age increases more significantly to the reduction of river inflow. Least-square regression analysis indicates that the relationship between water ages and river inflow at a given river location can be reasonably described by the power law. In comparison to normal and high flow conditions, both regression equations and hydrodynamic model simulations indicate that flow reduction under low flow conditions lead to substantial increase of water age. Because water age is an important indicator for estuarine water quality and ecosystem assessment, results from this study are helpful for environmental and water resources management of the Little Manatee River basin. Inflow scenarios are limited to constants in this study for better understanding of the water age responses to inflows through the regression equation. The regression equations may be used as a quick and cost-effective assessment tool in preliminary water resources planning, especially for those inflows corresponding to short water ages.

6 Discussions

Time series inflows in Little Manatee River generally show no significant variations except short-term periods of storm events. Therefore, constant inflow approximations can be used to estimate water ages most flow periods in the Little Manatee River system. To use the average constant flow for approximating the slightly varied inflows, time series of inflows can be divided into several zones of "quasi-steady flow conditions" so that flows in each zone do not substantially vary. If the water age corresponding to the constant inflow is less than the period of the quasi-steady flow, the constant inflow approximation is reasonable. However, if the time scale of the water age for the constant inflow is longer than the period of quasi-steady flow, hydrodynamic model simulations can be conducted to obtain more accurate estimations of water ages than those from the regression equations. Model simulations can be performed by replacing the constant river inflow with the specific time-dependent time series of inflow scenarios in the model simulation. Water ages at different locations in the river can be estimated for each time-dependent inflow scenario in the same way as those for constant inflow simulations. However, the simulation of water age under time-dependent flow is case specific, which is difficult to show how water ages response to different unsteady inflows.

Acknowledgments Funding support for this study was provided by the Southwest Florida Water Management District. The authors greatly appreciate the valuable comments made by the reviewers for the improvement of the manuscript.

References

1. Andrejev O, Myrberg K, Lundberg PA (2004) Age and renewal time of water masses in a semi-enclosed basin—application to the Gulf of Finland. Tellus 56A:548–588
2. Blumberg AF, Mellor GL (1987) A description of a Three-dimensional Coastal Ocean Circulation Model. In: Heaps NS (ed) Three-dimensional Coastal Ocean Models, Coastal and Estuarine Sciences, Vol 4. AGU, Washington, E.D., pp 1–16
3. Deleersnijder E, Campin J-M, Delhez EJM (2001) The concept of age in marine modelling: I. Theory and preliminary model results. J Mar Syst 28:229–267
4. Deleersnijder E, Mouchet A, Delhez EJM, Beckers J-M (2002) Transient behavior of water ages in the world ocean. Math and Comput Model 36:121–127
5. England MH (1995) The age of water and ventilation time scale in a global ocean model. J Phys Oceanogr 25:2756–2777
6. Galperin B, Kantha LH, Hassid S, Rosati A (1988) A quasiequilibrium turbulent energy model for geophysical flows. J Atmos Sci 45:55–62
7. Gourgue O, Deleersnijder E, White L (2007) Toward a generic method for studying water renewal, with application to the epilimnion of Lake Tanganyika. Estuar, Coast Shelf Sci 74(4):628–640
8. Guyondet T, Koutitonsky VG, Roy S (2005) Effects of water renewal estimates on the oyster aquaculture potential of an inshore area. J Mar Syst 58:35–51
9. Hamrick JM (1996) User's manual for the environmental fluid dynamics computer code special Rep. No. 331. The College of William and Mary, Gloucester Point, VA
10. Hilton ABC, McGillivary DL, Adams EE (1998) Residence time of freshwater in Boston's Inner Harbor. J Waterw, Port, Coast, Ocean Eng, ASCE 124(2):82–89
11. Huang W (2007) Hydrodynamic modeling of flushing time in a small estuary of North Bay, Florida. J Estuar, Coast, Shelf Sci 74(4):722–731
12. Huang W, Spaulding M (2002) Modeling residence time response to freshwater input in Apalachicola Bay, Florida, USA. Int J Hydrol Process 16:3051–3064
13. Huang W, Liu X, Chen X (2008) Numerical modeling of hydrodynamics and salinity transport In Little Manatee River. J Coast Res Spec Issue 52:13–24
14. Intera and Aqua Terra Consultants (2006) Estimating the ungaged inflows in the Little Manatee River basin, Florida. Report prepared for the Southwest Florida Water Management District. Brooksville, Florida

15. Kratzer CR, Biagtan RN (1997) Determination of travel times in the lower San Joaquin River Basin, California, from dye-tracer studies during 1994–1995. U.S. Geological Survey Water-Resources Investigations Report 97-4018
16. Mellor GL, Yamada T (1982) Development of a turbulence closure model for geophysical fluid problems. Rev Geophys Space Phys 20:851–875
17. Miller RL, McPherson BF (1991) Estimating estuarine flushing and residence times in Charlotte Harbor, Florida, via salt balance and a box model. Limnol Oceanogr 36:602–612
18. Monsen NE, Cloern JE, Lucas LV (2002) A comment on the use of flushing time, residence time, and age as transport time scales. Limnol Oceanogr 47:1545–1553
19. Peebels EB, Flannery MS (1992) Fish nursery use of the Little Manatee River estuary (Florida): relationships with freshwater discharge. Southwest Florida Water Management District, Brooksville, Florida
20. Ribbe J, Wolff JO, Staneva J, Gräwe U (2008) Assessing water renewal time scales for marine environments from three-dimensional modelling: a case study for Hervey Bay, Australia. Environ Model Softw 23(10–11):1217–1228
21. Shen J, Haas L (2004) Calculating age and residence time in the tidal York River using three-dimensional model experiments. Estuar, Coast Shelf Sci 61:449–461
22. Shen J, Wang HV (2007) Determining the age of water and long-term transport timescale of the Chesapeake Bay. Estuar, Coast Shelf Sci 74(4):585–598
23. Smagoringsky J (1963) General circulation experiments with the primitive equations. I. The basic experiment. Mon Weather Rev 91:99–164
24. Takeoka H (1984) Fundamental concepts of exchange and transport time scales in a coastal sea. Cont Shelf Res 3:311–326
25. Tetra Tech (2002) User's manual For environmental fluid dynamics code. http://www.epa.gov/athens/wwqtsc/html/efdc.html
26. Zimmerman JTF (1988) Estuarine residence times. In: Kjerfve B (ed) Hydrodynamics of estuaries, V. 1. CRC Press, pp 75–84

Environ Fluid Mech (2010) 10:213–233
DOI 10.1007/s10652-009-9138-3

ORIGINAL ARTICLE

Simulating transport of ^{129}I and idealized tracers in the northern North Atlantic Ocean

Steinar Orre · John N. Smith · Vasily Alfimov · Mats Bentsen

Received: 12 October 2008 / Accepted: 28 May 2009 / Published online: 12 June 2009

Abstract Transport of the radioactive tracer Iodine-129 (^{129}I, $T_{1/2} = 15.7$ Myr) in the northern North Atlantic Ocean has been investigated using a global isopycnic Ocean General Circulation Model (OGCM) and observed data. ^{129}I originates mainly from the nuclear fuel reprocessing plants in Sellafield (UK) and La Hague (France), and is transported northwards along the Norwegian coast, and then into surface and intermediate layers in the Arctic Ocean through the Barents Sea and the Fram Strait, but also partly recirculating south along the eastern coast of Greenland. In the North Atlantic Subpolar Seas, ^{129}I is mainly found in dense overflow waters from the Nordic Seas being exported southwards in the Deep Western Boundary Current, and to a lesser extent in surface and intermediate layers circulating cyclonically within the Subpolar Gyre. Observed concentration of ^{129}I along a surface transect in the eastern Nordic Seas in 2001 is captured by the OGCM, while in the Nansen Basin of the Arctic Ocean the OGCM overestimates the observed values by a factor of two. The vertical profile of ^{129}I in the Labrador Sea, repeatedly observed since 1997 to present, is fairly realistically reproduced by the OGCM. This indicates that the applied model system has potential for predicting the magnitude and depth of overflow waters from the Nordic Seas into the North Atlantic Subpolar Seas. To supplement the information obtained from the ^{129}I distribution, we have conducted a number of idealized tracer experiments with the OGCM, including tracers mimicking pure water masses, and instantaneous pulse releases. New insight into time-scales of tracer transport in this region is obtained by utilizing a few

S. Orre · M. Bentsen
Nansen Environmental and Remote Sensing Center & Bjerknes Centre for Climate Research, Bergen, Norway

S. Orre (✉)
Badger Explorer ASA, Stavanger, Norway
e-mail: steinar.orre@gmail.com

J. N. Smith
Marine Environmental Sciences Division, Bedford Institute of Oceanography, Dartmouth, NS, Canada

V. Alfimov
ETH/PSI Ion Beam Physics Group, Institute of Particle Physics, Zurich, Switzerland

recently developed methods based on the theory of Transit Time Distribution (TTD) and age of tracers. Implications for other types of "anomalies" in the northern North Atlantic Ocean, being anomalous hydrography or chemical tracers, and how they are interpreted, are discussed.

Keywords Iodine-129 · Nordic Seas · Arctic Ocean · North Atlantic Subpolar Seas · Transit Time Distribution · Age

1 Introduction

A variety of anthropogenic compounds entered the world oceans during the latter half of the twentieth century. Examples are radioactive tritium originating from nuclear weapons testing and Chlorofluorocarbons (CFCs) stemming from various industrial applications. These tracer compounds have provided insights into time-scales and mixing processes concerning the ventilation of intermediate and deep waters in the ocean, and thus helped quantifying how efficiently the ocean exchange properties with the atmosphere. Furthermore, anthropogenic tracers have provided an important benchmark for testing the performance of global Ocean General Circulation Models (OGCMs), where the observed surface values are used as boundary conditions (e.g. [15, 16]).

Oceanic discharges of anthropogenic radionuclides from European nuclear fuel reprocessing plants have introduced another set of chemical tracers to the marine environment, for instance ^{137}Cs, ^{90}Sr, ^{99}Tc, and ^{129}I [4]. As with CFCs and tritium, these tracers can be detected at very low concentrations, and are extremely inert in sea water justifying their interpretation as *passive tracers*. In contrast to CFCs and tritium, which show a rather homogeneously distribution across world oceans basins, radionuclides from European nuclear fuel reprocessing plants are specific to European coastal waters. These radionuclides are therefore better suited for a detailed investigation of ocean processes in ocean regions bordering the source regions. Examples are the surface circulation and mixing in the Nordic Seas, the surface and intermediate circulation in the Arctic Ocean, and the dense overflow descending from the Greenland-Scotland Ridge into the deep North Atlantic Subpolar Seas.

Iodine-129 (^{129}I) is a highly soluble and long-lived ($T_{1/2} = 15.7$ Myr) radioactive isotope with a relatively large anthropogenic component. The pre-anthropogenic amount of ^{129}I in the oceans was roughly 50 kg, atmospheric fallout from nuclear weapons testing added 50–150 kg, while direct releases from the nuclear fuel reprocessing plants in Sellafield in the Irish Sea since 1952 and from La Hague on the French coast in the English Channel since 1962 added 5,000 kg [1]. There are some indications that contaminated Russian rivers from Siberia can be a source of ^{129}I, but this contribution is nevertheless of minor importance for the large scale distribution in the North Atlantic and Arctic Oceans [8]. By far the main source of ^{129}I entering the oceans over the last half century, and also the main source for the oceanic inventory, has been direct discharges from Sellafield and La Hague [1].

Observations of ^{129}I have shed light on many of the characteristic features of the ocean circulation along the pathways from the source regions. Assuming that the bulk of tracers from both Sellafield and La Hague are being transported into the North Sea and then entrained in the Norwegian Coastal Current (NCC) [37], mixing between the Norwegian Atlantic Current (NwAC) and the NCC along the Norwegian coast should cause a net transport of tracers from NCC to the NwAC. Gascard et al. [19] estimated that roughly one half of the ^{129}I burden initially carried by the NCC is lost to the NwAC at about 70° N.

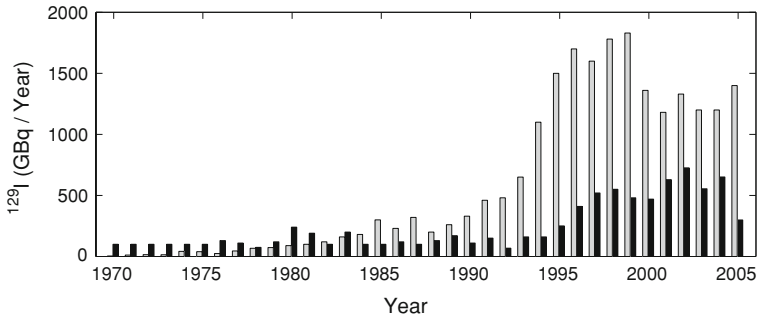


Fig. 1 Annual releases of ^{129}I (10^9 Bq) from the reprocessing plant at Sellafield (*black bars*) and La Hague (*grey bars*)

In the Arctic Ocean, ^{129}I can be seen as Atlantic derived waters, where a pronounced front is found between waters of Pacific and Atlantic origin. This is because the ^{129}I concentrations are more than an order of magnitude higher in Atlantic waters [41, 42]. Going deeper, the Pacific-Atlantic front is observed further into the Canada basin, in line with the observed distribution of intermediate Atlantic Waters (e.g. [39]). Alfimov et al. [3] observed a doubling of ^{129}I concentration between 1996 and 2001 in the Amundsen Basin of the Arctic Ocean down to 1,000 m depth, while in the Nansen and Makarov Basins the concentrations increased in a shallower layer, indicating slower ventilation of Atlantic Waters in these basins.

In the western Nordic Seas, observed ^{129}I and hydrography revealed a complex structure of the different water masses and how they are transformed while being transported southwards with the East Greenland Current (EGC) [2]. Further south in the Irminger and Labrador Seas, ^{129}I concentrations are highest in the deepest layers, providing a direct signature of overflow waters from the Nordic Seas [43].

The timing of the release rate signal of ^{129}I from Sellafield and La Hague is favorable for studying recent ocean processes occurring since 1990, since the discharges increased by six-fold during the 1990s, see Fig. 1 [1]. This "ramp" of increased concentration is right now being observed by ongoing programs in the North Atlantic Subpolar Sea tracing out deep overflow waters from the Nordic Seas.

Simulating the dispersion of ^{129}I from its source regions to the Nordic Seas and further into intermediate layers in the Arctic Ocean and deep layers in the Irminger and Labrador Seas is a very challenging task, requiring a decent representation of numerous ocean processes—some of which are only parameterized in the current generation of OGCMs (e.g. small scale mixing processes). Failure to represent the transport realistically in only some areas could potentially preclude the large scale distribution pattern from being correctly simulated. The vertical dispersion of ^{137}Cs, ^{90}Sr and ^{99}Tc has received little attention, but the surface distributions have been reasonable well reproduced by models that are representative for the current generation of OGCMs used in ocean climate studies (e.g. [18, 29, 36]).

The first objective of this study is to use both observed and simulated distribution of ^{129}I to extract characteristic circulation features of tracer transport in the northern North Atlantic Ocean, and to what extent the applied OGCM is able to reproduce observed values of ^{129}I. The second objective of this study is to apply a few recently developed methods for computing time-scales of tracer transport and the degree of mixing that tracers experience along their pathways. Finally, we discuss how the results obtained in this paper relates to a more general description and interpretation of oceanic tracers, both passive and active.

2 Model description and experimental setup

2.1 Ocean General Circulation Model

The numerical methods and thermodynamics of Miami Isopycnic Coordinate Ocean Model (MICOM) are documented in Bleck and Smith [6] and Bleck et al. [7]. In the version of MICOM used in this study, several important aspects deviate from the original model.

The original MICOM uses potential density with reference pressure at 0 db as vertical coordinate (σ_0-coordinates). This ensures that the very different flow and mixing characteristic in neutral and dia-neutral directions is well represented near the surface since isopycnals and neutral surfaces are similar near the reference pressure. For pressures that differ substantially from the reference pressure, this does not hold. In this study, we choose a reference pressure of 2,000 db (σ_2-coordinates). The non-neutrality of the isopycnals in the world ocean is then reduced in comparison to having the reference pressure at the surface [34].

For advection of tracers (potential temperature, salinity and passive tracers) and layer thickness we use incremental remapping [14] adapted to the grid staggering of MICOM. The algorithm is computationally rather expensive compared to other second order methods with limiters for a single tracer, but the cost of adding additional tracers is modest. In contrast to the original transport methods of MICOM, incremental remapping ensures monotonicity of the tracers.

Traditionally, MICOM expresses the pressure gradient force (PGF) as a gradient of a potential on an isopycnic surface. This is only accurate if the density can be considered to be a function of potential density and pressure alone, which is not the case [9]. Inspired by recent work of Rainer Bleck (pers. comm.), we have based our formulation on Janjić [27] where the PGF is expressed as a gradient of the geopotential on a pressure surface. This allows us to use a more accurate representation of density in the PGF formulation.

For diapycnal mixing we follow the MICOM approach of a background diffusivity dependent on the local stability implemented using the scheme of McDougall and Dewar [33]. To incorporate shear instability and gravity current mixing, we add a Richardson number dependent diffusivity to the background diffusivity.

$$
\begin{aligned}
&\nu_d = \nu_b + \nu_r, \\
&\nu_b = \frac{C}{N}, \quad \nu_r = \nu_0 \max\left\{0, 1 - \left(\frac{Ri_g}{Ri_0}\right)^2\right\}^3, \\
&N^2 = \frac{g}{\rho}\frac{\partial \rho}{\partial z}, \quad Ri_g = \frac{N^2}{(\partial u/\partial z)^2 + (\partial v/\partial z)^2},
\end{aligned}
$$

where N is the Brunt-Väisälä frequency and Ri_g is the local gradient Richardson number. The parameter determining the background diffusivity is set to $C = 1.8 \times 10^{-7}$ m^2 s^{-2}, while the critical Richardson number is $Ri_0 = 1$. The maximum Ri_g dependent diffusivity ν_0 is set to 500×10^{-4} m^2 s^{-1} in the 300 m closest to the ocean floor to parameterize gravity current mixing, and 50×10^{-4} m^2 s^{-1} elsewhere to parameterize shear instability mixing. This greatly improves the water mass characteristics downstream of overflow regions. Lateral turbulent mixing of momentum and tracers is parameterized by Laplacian diffusion, and layer interfaces are smoothed with biharmonic diffusion.

A sea ice module is incorporated, with dynamics using the viscous-plastic rheology [25] as implemented by Harder [24], and the thermodynamics of Drange and Simonsen [13].

The model grid covers the global domain and by using a conformal mapping [5], the poles have been placed to enhance the resolution in the North Atlantic and Arctic region.

The grid size in the Nordic Seas is about 40 km, in the Labrador Sea about 50 km, and then the grid size increases smoothly to about 500 km on the opposite side of the earth. In the vertical, the model has 35 layers of which the uppermost is a mixed layer with temporal and spatial varying density. The potential densities of the isopycnal layers are in the range 1030.12–1037.80 kg m^{-3}.

The model is forced with daily NCEP/NCAR reanalysis fields [28], and the spin-up consisted of two full NCEP/NCAR reanalysis cycles covering the period 1948–2005 giving in total 116 years of spin-up. The following 58 model years are then analyzed.

2.2 Tracer experiment

The following tracer simulations were conducted:

(1) ^{129}I and the age of ^{129}I from Sellafield and La Hague based on historical release rates for the period 1952 to present.
(2) Idealized tracer resembling pure water masses, and their age, from Sellafield and La Hague for the same period.
(3) Idealized pulse releases from Sellafield and La Hague, released once every decade from January 1950 until January 1990 (i.e. five pulse releases from each source).

For experiment (1), release rates with an annual resolution are used for the period 1952–1989, and monthly resolution from January 1990 until January 2006. Figure 1 shows the annual releases from 1970 to 2006 (releases prior to 1970 are negligible). For experiment (2), we prescribe a constant value in the surface grid cell at the source. The pulse releases in experiment (3) are tracer released continuously over January once every decade from 1950 until 1990 (i.e. boxcar pulses, not real delta functions).

3 Ensemble Average of Impulse Boundary Propagators as a surrogate for the Transit Time Distribution

In a diffusive environment like the ocean, there is neither a unique pathway nor a single time-scale for a tracer being transported from a point source. Due to mixing, there is a wide range of possible pathways and transit times, and quantifying oceanic tracers without accounting for the diffusive part of the transport might lead to erroneous conclusions. We will here briefly review the theoretical framework for the Transit Time Distribution (TTD) of a tracer in a generalized advection-diffusion regime, how the TTD relates to the age of tracer, and why the TTD is difficult to obtain for unsteady (time-dependent) flows. A recently developed technique to estimate the TTD in unsteady flows where we take advantage of the pulse releases (experiment 3 listed above) will be discussed.

A passive tracer with concentration C evolves in a fluid flow according to the advection-diffusion equation:

$$\left(\frac{\partial}{\partial t} + \mathcal{L}\right) C = q_s, \tag{1}$$

where $\mathcal{L}$ is the linear transport operator, and q_s is the tracer source. No-flux boundary conditions apply at the sea surface and sea floor. Typically, $\mathcal{L}(C) = \nabla \cdot (\mathbf{v}C - \mathbf{K} \cdot \nabla C)$, where $\mathbf{v}$ is the advection vector, and $\mathbf{K}$ the diffusivity tensor. In a layered model like MICOM, the advection-diffusion equation also involves layer thickness, we will nevertheless use this form for simplicity.

We are often interested in characteristic time-scales for various ocean processes. For instance, the age of a tracer along its pathway can be obtained by coupling an additional advection-diffusion equation as suggested by Deleersnijder et al. [10]:

$$\left(\frac{\partial}{\partial t} + \mathcal{L}\right)\alpha = C, \tag{2}$$

where the product of the age and the concentration, $\alpha(\mathbf{x}, t) = a(\mathbf{x}, t) \cdot C(\mathbf{x}, t)$, is advected and diffused by the model. No-flux boundary conditions apply at the sea surface and sea floor. The age may then be calculated as $a = \alpha / C$. We will refer to this quantity as the *tracer age*, and the age of the idealized tracer resembling water masses (experiment 2 listed above) as the *idealized tracer age*, and prescribe both of them to be zero at the source.

In the present study we are investigating transport of tracers from two sources, and we focus on the combined age of the tracers rather than the age of the tracer from each individual source. If C_S and a_S are concentration and age, respectively, of ^{129}I from Sellafield, and C_L and a_L are the corresponding quantities from La Hague, we define the age of ^{129}I from both sources using the concentration-weighted average: $a_{^{129}\mathrm{I}} = (a_S C_S + a_L C_L) / (C_S + C_L)$. A simple arithmetic average would be flawed by putting equal weight to tracers having different concentrations. This definition will be used later in the text when referring to the age of ^{129}I.

A special case of Eq. 1 is the Green's function of the advection-diffusion equation:

$$\left(\frac{\partial}{\partial t} + \mathcal{L}\right) G = \delta(t - t_0)\delta(\mathbf{x} - \mathbf{x}_0), \tag{3}$$

where a unit amount of tracer is released from the source $\mathbf{x}_0$ at time t_0. No-flux boundary conditions apply at the sea surface and sea floor. Having solved Eq. 3, one can use $G(\mathbf{x}, \mathbf{x_0}, t - \tau)$, where $\tau = t - t_0$ is the transit time, to construct any solution for an arbitrary release function $q_s(t)$ from a convolution integral (e.g. [21, 26]). $G(\mathbf{x}, \mathbf{x_0}, t - \tau)$ contains complete information about the transport operator $\mathcal{L}$, and weights prior releases from the source region over different transit times. Hence, $G(\mathbf{x}, \mathbf{x_0}, t - \tau)$ is the distribution of transit times since a tracer at $\mathbf{x}$ and time t was released from the source region $\mathbf{x_0}$ at $t_0 = t - \tau$, and is termed TTD [26].

One difficulty which arises is how to construct the TTD in an unsteady flow. To illustrate this, let us first consider the case with steady flow. G is then independent of t and t_0, and only a function of the transit time τ. Constructing a Green's function weighting *future* releases from the source region over different transit times would be equivalent to weighting *prior* releases, so $G(\mathbf{x}, \mathbf{x_0}, t - \tau) = G(\mathbf{x}, \mathbf{x_0}, t_0 + \tau)$. We call the latter property the Impulse Boundary Propagator (IBP). The IBP is constructed by releasing a unit pulse at the source and then integrate the model forward in time. In steady flows, IBP equals TTD [22].

In an unsteady flow, the relation between TTD and IBP no longer holds, and the only way to construct the TTD is to use an adjoint tracer equation [26]. For most OGCMs, adjoint versions of the code have not been constructed, hence, it is not possible to obtain the TTD. However, in the recent study by Haine et al. [22], close statistical similarities between the adjoint TTD and the forward IBP in both periodic and aperiodic flows are demonstrated. A number of idealized pulse releases were simulated in an idealized barotropic double gyre circulation model. It was shown that the Ensemble Average of the IBPs (EAoIBP) closely matches the ensemble average of the TTDs obtained from the adjoint version of the same model, even when the circulation was in a regime with strongly chaotic time dependence. Furthermore, while the IBP approach contains random errors, the bias is zero. The EAoIBP may therefore not be relevant to study specific time periods, but could instead be used to explore statistical properties of tracer transport in the area of interest.

We construct the EAoIBP, $\overline{G}(\mathbf{x}, \mathbf{x_0}, \tau)$, from the five individual IBPs listed in experiment (3) in the previous section, by setting $t_0 = 0$:

$$\overline{G}(\mathbf{x}, \mathbf{x_0}, \tau) = \sum_{i=1}^{N} G_n(\mathbf{x}, \mathbf{x_0}, \tau)/N,$$

where G_1 refers to the pulse released in 1950, G_2 the pulse released in 1960, and so forth. Note that τ is 56 years for G_1, but only 16 years for G_5 referring to the pulse released in 1990. We choose to let G_1 contribute to the EAoIBP over the first 16 years, G_2 contribute to the EAoIBP over the first 26 years, and so forth. From the EAoIBP, two other important quantities characterizing tracer transport can be derived. The first moment of the EAoIBP, which we will refer to as the *transit time*, can be calculated as

$$\overline{\Gamma}(\mathbf{x}, \mathbf{x_0}) = \int_0^\infty \tau \overline{G}(\mathbf{x}, \mathbf{x_0}, \tau)\, d\tau. \tag{4}$$

The *width* of the EAoIBP, which is the second moment of the EAoIBP centered around the transit time, can be calculated as

$$\overline{\Delta}^2(\mathbf{x}, \mathbf{x_0}) = \frac{1}{2} \int_0^\infty \left(\tau - \overline{\Gamma}\right)^2 \overline{G}(\mathbf{x}, \mathbf{x_0}, \tau) d\tau. \tag{5}$$

Due to the finite length of the simulation, these integrals are approximated by integration over the simulated time period.

It is reasonable to expect that $\overline{\Delta}$ is strongly dependent on the degree of mixing in the ocean (e.g. [11, 47]). We will use the ratio of the width and the transit time of the EAoIBP, $\overline{\Delta}/\overline{\Gamma}$, as a measure of mixing—or the tendency of a tracer plume to broaden in time. In the limit where $\overline{\Delta}/\overline{\Gamma}$ approaches zero, advection dominates and $\overline{G}$ maintains a narrow *pulse-like* shape. Water masses and tracers are then transported away from the source region while preserving their initial properties. If, however, $\overline{\Delta}/\overline{\Gamma} \sim 1$, diffusive processes are equally important, and $\overline{G}$ typically has a long tail extending well past the transit time $\overline{\Gamma}$. Water masses and tracers are then likely to exchange properties with surrounding water masses. In the following we will drop the overbars, keeping in mind that Δ and Γ only will be used for the ensemble average values.

4 Results

4.1 Large scale dispersion of ^{129}I

Before investigating the large scale horizontal distribution of ^{129}I, we start by examining the integrated vertical distribution obtained from the OGCM. We perform the integral $\int_\Omega C_{\text{iodine}}(x, y, z, t)\, dxdy$, where the domain Ω is the world oceans, and plot this quantity as a function of depth and time (Fig. 2). As a consequence of increasing releases from both Sellafield and La Hague (Fig. 1), we see a steady increase of ^{129}I concentration in the surface for the time span 1970 to the end of 2005. At the same time, tracers are entering deeper layers. Following contours of constant concentration, one sees that the vertical dispersion is relatively slow in the upper 1,800 m, faster between 1,800 and 3,500 m, and then slower again.

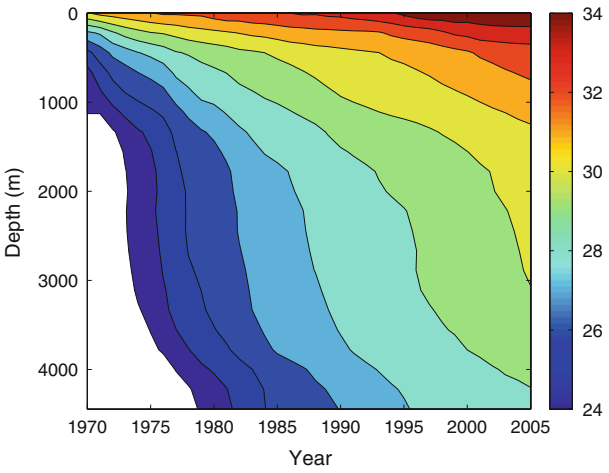


Fig. 2 Simulated concentration of ^{129}I ($\log(^{129}\mathrm{I})$, $\mathrm{Bq\,m^{-1}}$) from Sellafield and La Hague integrated over the world oceans, $\int_\Omega C_{\mathrm{iodine}}(x, y, z, t)\, dxdy$, plotted as a function of depth and time

Table 1 Simulated burden of ^{129}I and pulse tracers at different depths by the end of the simulation, December 2005. Inventory of tracers between the depths H_1 and H_2 is calculated as $\int_{H_2}^{H_1} \int_\Omega C(x, y, z, t_{2005})\, dxdydz$, where Ω is the world ocean, and then divided by the total inventory $\int_{\mathrm{bottom}}^{0} \int_\Omega C(x, y, z, t_{2005})\, dxdydz$

Depth (m)	% of ^{129}I	% of pulse-1990	% of pulse-1970	% of pulse-1950
Surface–200	49	31	18	14
200–500	31	35	25	20
500–1,000	15	25	25	21
1,000–2,000	3	7	19	25
2,000–bottom	2	2	13	20

By the end of the simulation (December 2005), the ^{129}I burden in the upper 200 m of the world oceans is roughly 50%, while 30% is found between 200 and 500 m. The remaining 20% is below 500 m, see Table 1. For comparison, we have calculated the tracer burden from the pulse releases, as described in the previous section. The pulses released from Sellafield and La Hague in 1990 are in December 2005 still largely constrained to the upper 500 m, which contains 66% of the total concentration, and only 9% are found below 1,000 m. The pulses released in 1950 is, as expected, more uniformly distributed in the vertical, with 45% of the tracers located below 1,000 m (see Table 1).

Figure 3 shows the horizontal distribution of ^{129}I at three different depths in 1995 and 2005. At the surface, which is here defined to cover the uppermost 200 m of the water column, concentration of ^{129}I is highest in the shallow North Sea and along the Norwegian Coast in the eastern Nordic Seas. In the western Nordic Seas, one sees tracers having much smaller concentration being transported southwards with the EGC along the east Greenland coast. The surface transport route through the Barents Sea and the Kara Sea, and the entrance to the Arctic Ocean through the St. Anna Through is also clearly depicted, and even more so at intermediate depths between 500 and 1,500 m. At this depth level, relatively high concentration extends from the St. Anna Through and into the Nansen Basin of the Arctic Ocean. Southward transport along the east Greenland coast indicates that the EGC carries tracers both in the surface and at intermediate depths. In deeper layers, here defined as depths below

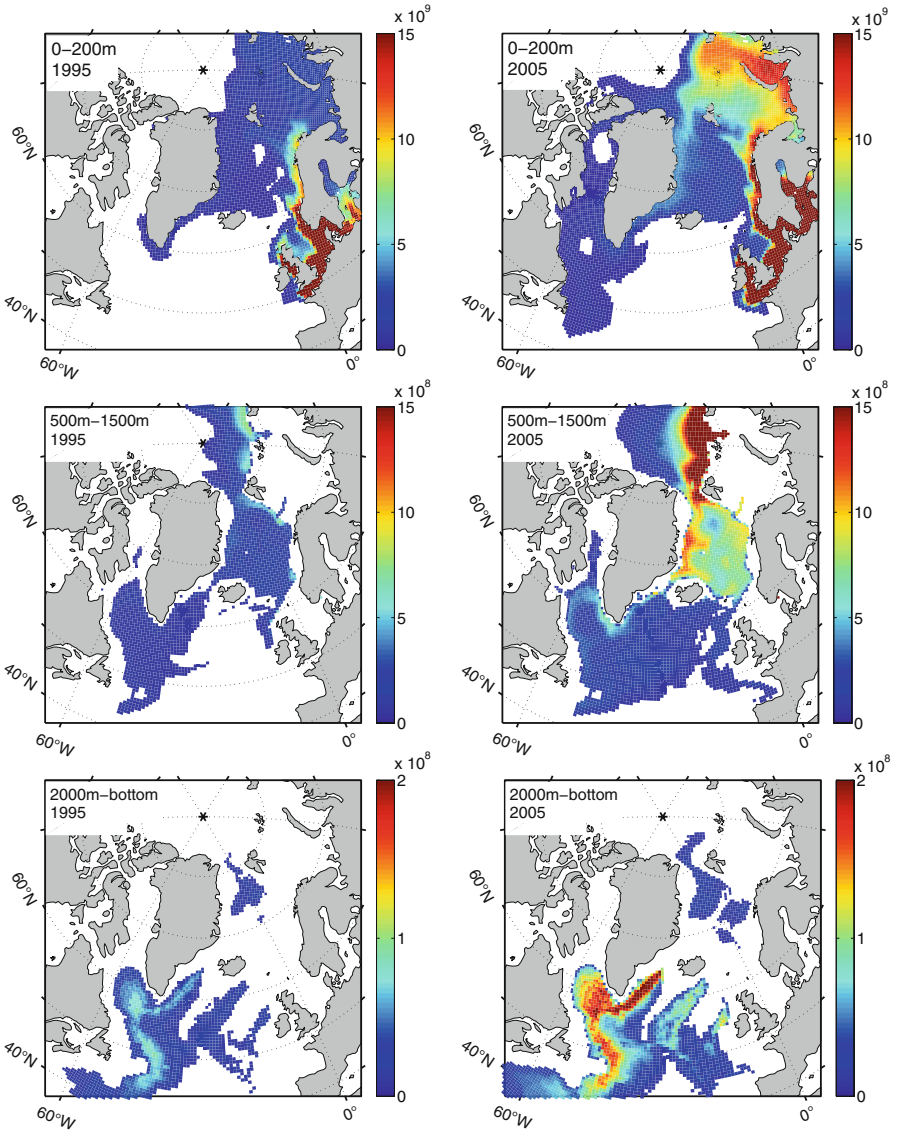


Fig. 3 Simulated concentration of ^{129}I (atoms l^{-1}) from Sellafield and La Hague in 1995 and 2005 averaged over the upper 200 m, for intermediate waters between 500 and 1,500 m, and for deep waters below 2,000 m, see text labels in each subfigure for depth bin and sampled year. Note the different legend scale. ^{129}I concentration less than 10^7 atoms l^{-1} is set to zero

2,000 m, almost all trace of ^{129}I are found in the deep North Atlantic Subpolar Seas. This is a very clear signature of the pathways of overflow waters from the Nordic Seas spilling over the Greenland-Scotland Ridge. For the latter, the largest contribution passes through the Denmark Strait, but also the contribution from the Iceland-Faroe Ridge and the deep Faroe-Shetland Channel are clearly seen.

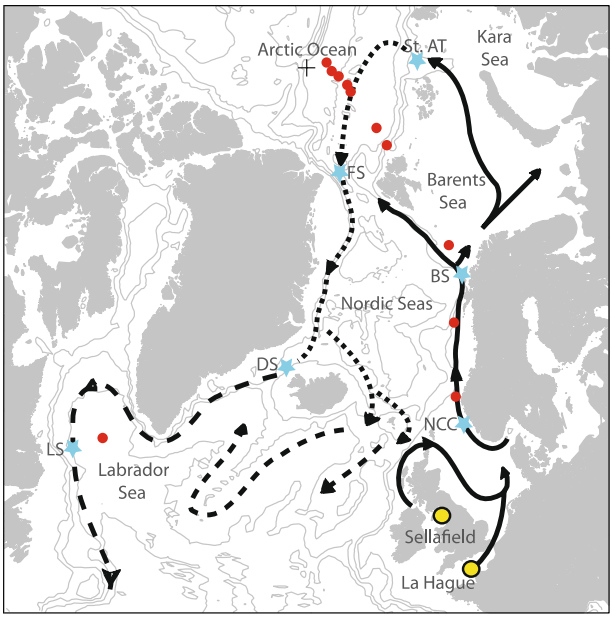


Fig. 4 Schematics of the main pathways of the simulated tracer transport, *solid line* indicates transport mainly in the surface, *dotted line* the transport both in the surface and at intermediate depths, and *dashed line* transport in deeper layers. Locations of observed ^{129}I are shown as *red dots*. The *red dot* in the Labrador Sea is station 17 on the ARW7 (WOCE) section. Stations used for sampling IBP, EAoIBP, and idealized tracer age are shown as *blue stars*: Norwegian Coastal Current (NCC), Barents Sea (BS), St. Anna Through (St. AT), Fram Strait (FS), Denmark Strait (DS), and Labrador Sea (LS)

From the simulated distribution of ^{129}I we can deduce a schematic picture of the *major* transport routes from the source regions on the European shelf to the deep North Atlantic Ocean (Fig. 4). We emphasize that there are, in general, multiple pathways of tracers in the ocean, so this figure is only meant to illustrate the pathways carrying the main burden of tracers from Sellafield and La Hague by the OGCM.

Figure 5 shows observed [3] and simulated surface concentration of ^{129}I from the NCC to the North Pole sampled in July 2001. Along the Norwegian coast and in the Barents Sea the concentration from the OGCM matches the observed values quite closely. The OGCM slightly overestimate the values just north of Spitsbergen in the Nansen Basin, then slightly underestimate the values on the Nansen-Gakkel Ridge, while in the Amundsen Basin close to the North Pole the OGCM shows no tracers at all. The latter is in contrast to the observations showing slightly elevated concentration in the Amundsen Basin compared to the Nansen Basin.

4.2 Vertical profiles of ^{129}I in the Arctic Ocean and the Labrador Sea

In the Nansen Basin, a vertical profile of ^{129}I was collected in July 2001 [3]. Figure 6 shows the observed and simulated concentrations of ^{129}I at this location. Highest concentrations are found at the surface, and at 500 m the concentration are approximately one third of the surface value. The OGCM is overestimating the concentrations roughly by a factor of two in the upper 1,000 m, and seems to be underestimating the concentration at greater depth (although there is only one observation below 1,000 m). Simulated age of ^{129}I shows a

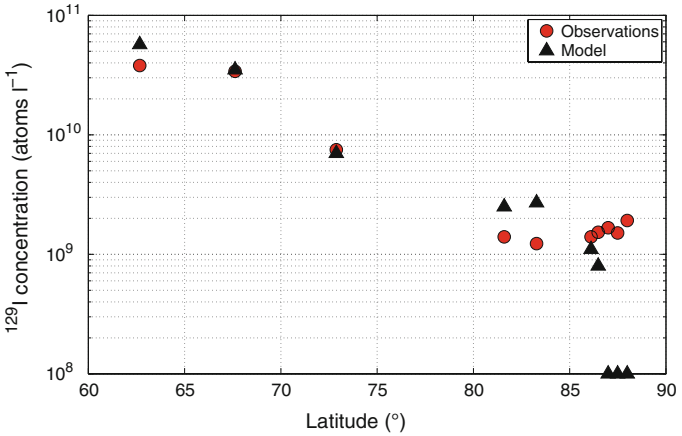


Fig. 5 ^{129}I concentration in the surface from observations and the OGCM along a transect from the Norwegian coast to the North Pole in July 2001. Location of the stations are shown in Fig. 4. The three last values from the model are strictly equal zero—however, to make them appear in the figure, the values are set to 10^8. The observed values are taken from Alfimov et al. [3]

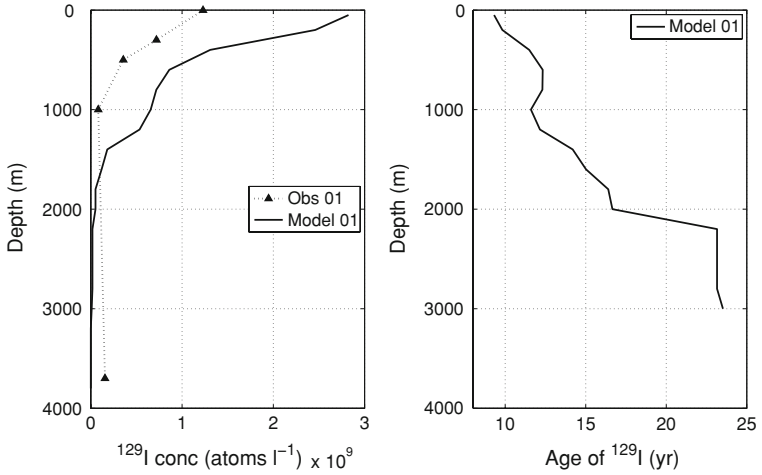


Fig. 6 Concentration and age of ^{129}I at a station in the interior Nansen basin, sampled in 2001. The observed values are taken from Alfimov et al. [3]. Values from the OGCM are interpolated to constant z-levels with 200 m resolution

general increase with depth, from an age of about 10 years at the surface to about 16 years at a depth of 2,000 m.

In the Labrador Sea, vertical profiles show a very different vertical distribution. Figure 7 displays depth profiles of ^{129}I concentration and age at station 17 on the ARW7 (WOCE) section, where the observed values in 1997, 1999, and 2001 are obtained from Smith et al. [43]. The observed values from 2003 and 2005 have not been published previously. Both the observed and simulated values show a slight decrease from the surface to approximately 3,000 m depth, then a pronounced increase down to the deepest layer at about 3,700 m exceeding the surface concentration by more than a factor of two. In the surface and at intermediate

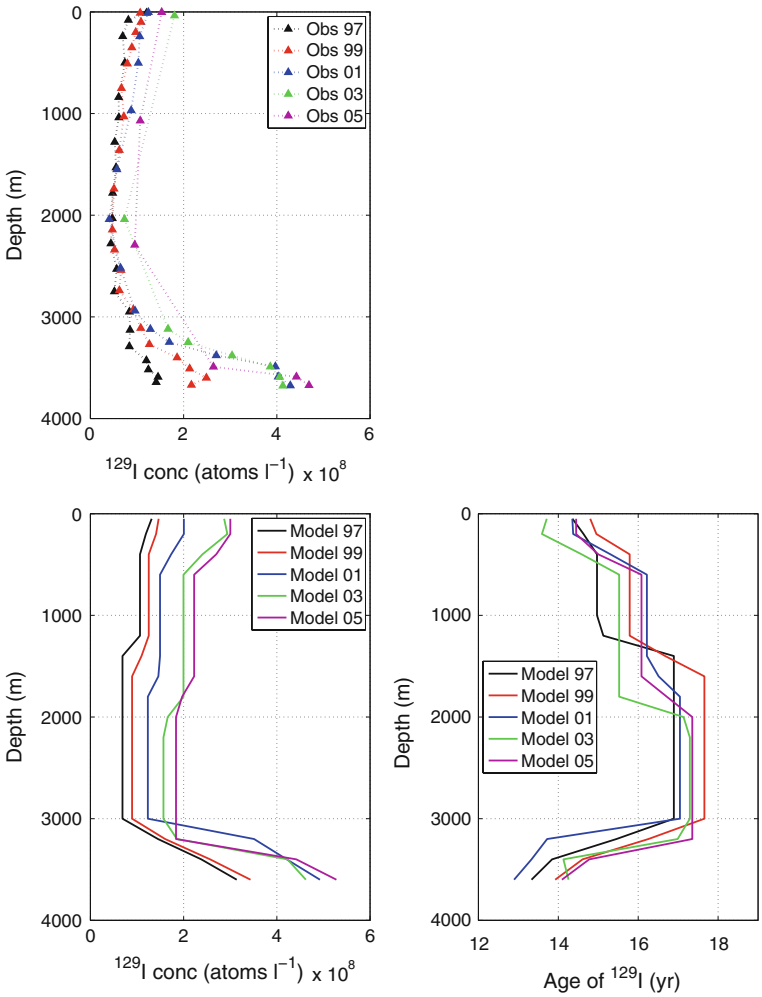


Fig. 7 Concentration and age of ^{129}I at station 17 on the ARW7 (WOCE) section in the Labrador Sea sampled in 1997, 1999, 2001, 2003 and 2005. The observed values from 1997, 1999, and 2001 are taken from Smith et al. [43], the values from 2003 and 2005 are unpublished data from John Smith. Values from the OGCM are interpolated to constant z-levels with 200 m resolution

depths down to 3,000 m the OGCM slightly exaggerates the ^{129}I concentration, although the general vertical structure is maintained. In the observations, a rapid increase is seen in the deep layers from 1997 to 2001, then the ^{129}I concentration levels off, but with the highest concentration found in 2005. The OGCM is capturing both the magnitude and the time evolution of the ^{129}I concentration in the deep layers in a fairly realistic way. The simulated age of ^{129}I in the Labrador Sea increases with depth to about 1,500–2,000 m, but then decreases below 3,000 m. This is consistent with the interpretation that tracers at this depth have been transported with overflow waters from the Nordic Seas, and then exported southwards. It is also worth pointing out that the age of ^{129}I in the deep overflow waters is slightly younger then that at the surface (Fig. 7).

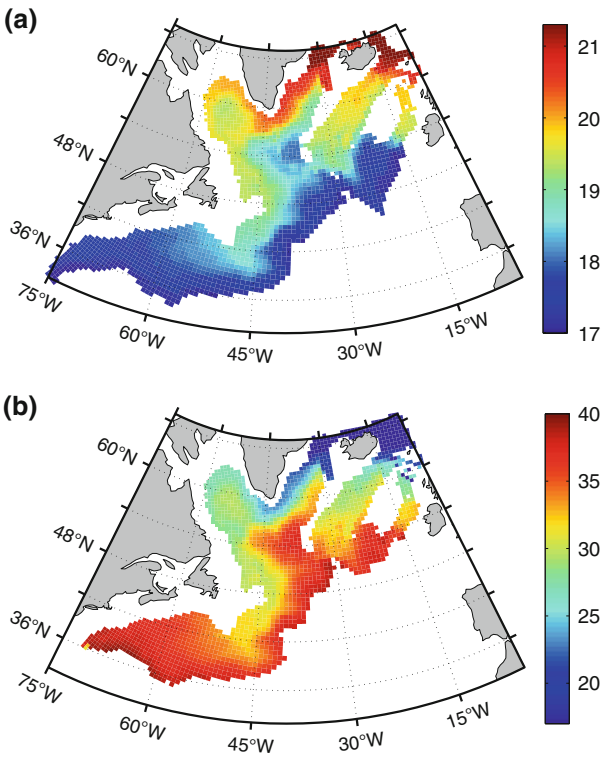


Fig. 8 **a** Simulated concentration of ^{129}I (log(^{129}I), atoms l^{-1}) from Sellafield and La Hague in 2005 on the σ_2-layers 37.07 and 37.12 (roughly equal 27.80 and 27.82 in σ_0-units), and **b** age of ^{129}I (in years) in 2005 on the same isopycnals. The arithmetic mean values from the two layers are shown, and a mask is used to ignore values when the isopycnal thickness is less than 1 m, and when the ^{129}I concentration is less than 10^7 atoms l^{-1}

4.3 ^{129}I in overflow waters from the Nordic Seas in the North Atlantic Subpolar Seas

We have seen that the OGCM quite reasonably reproduce the vertical structure and the temporal evolution of ^{129}I in the deep Labrador Sea, although slightly exaggerating values in the surface and intermediate depths. We identify the isopycnals in the OGCM with highest concentration of ^{129}I, since these layers are representative for overflow waters from the Nordic Seas into the deep North Atlantic Ocean. Figure 8 shows ^{129}I and age of ^{129}I on the σ_2-layers 37.07 and 37.12 in the North Atlantic Subpolar Seas in December 2005. It is readily seen that the Denmark Strait branch is dominant in terms of the southward ^{129}I transport, but also the Iceland-Faroe Ridge and the Faroe-Shetland Channel branch are clearly identified. From the age distribution in the deep Labrador Sea, the youngest ^{129}I is found along the rim of the basin, and older ^{129}I in the interior, confirming the faster transport route along the shelf break [20, 40]. The primary pathways of overflow waters from the Nordic Seas in the Subpolar North Atlantic are therefore visible in Fig. 8 as tongues of both relatively high concentration and young ^{129}I.

4.4 Impulse Boundary Propagators and idealized tracer age

Figure 9 shows the five IBP from La Hague (idealized pulse releases in 1950, 1960, 1970, 1980, and 1990) sampled along the pathway provided in Fig. 4. We will only focus on IBPs from La Hague since tracers from Sellafield are transported roughly along the same pathways as tracers from La Hague, although with a time lag of 1–2 years [36]. The stations used for

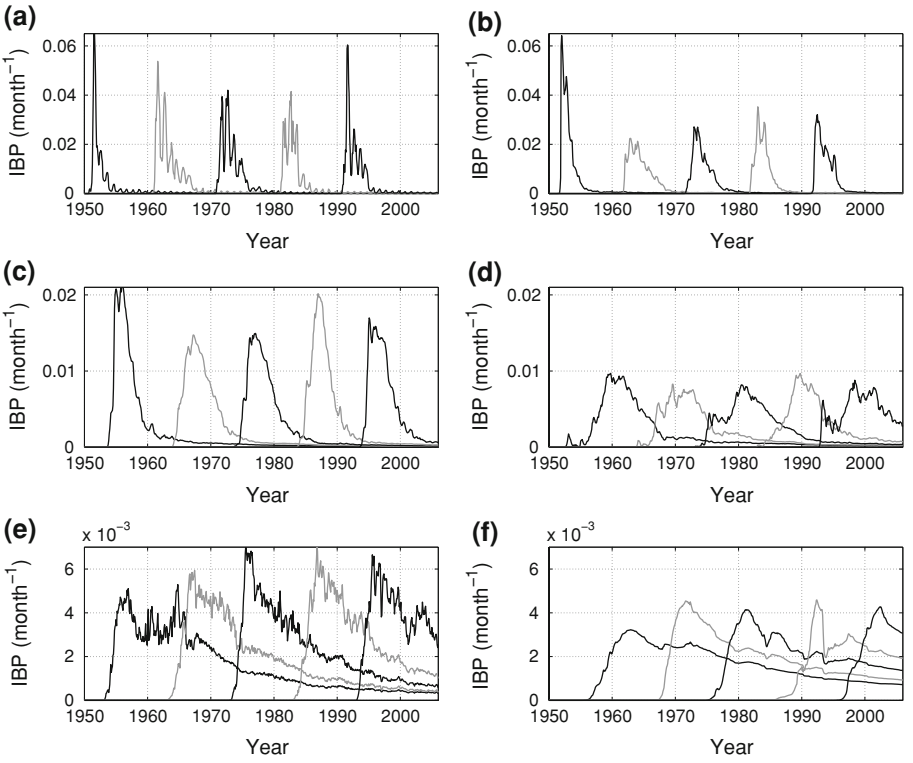


Fig. 9 Simulated Impulse Boundary Propagator (IBP) from La Hague sampled **a** in the Norwegian Coastal Current at 60° N in the depth range 0–100 m, **b** at the entrance to the Barents Sea, 0–100 m, **c** in the St. Anna Through, 300–400 m, **d** in the Fram Strait, 300–400 m, **e** in the Denmark Strait, 400–500 m, and **f** in the deep Labrador Sea, 3,000–4,000 m. Temporal resolution is 1 month in all time-series

sampling the IBPs are chosen such that they reflect the pathway of the core of the La Hague tracer (Fig. 4).

As expected, the magnitude of the IBPs is decreasing along the pathway, and mixing and multiple pathways are causing the shape of the IBPs to broaden. A clear annual signal is seen at the NCC station, to a lesser degree at the Barents Sea station, and it is almost absent in the St. Anna Through. In the Fram Strait and the Denmark Strait, the variations display a more high-frequency signal, while the variations in the Labrador Sea are very slow and contain no systematic signal.

The Ensemble Average of the Impulse Boundary Propagators (EAoIBP) from La Hague is estimated based on the five IBPs (experiment 3). Since the IBPs are released at different times, care has to be taken when calculating the EAoIBP. For instance, the pulse released in 1990 only contributes to the first 16 years of the EAoIBP, the pulse released in 1980 contributes to the first 26 years, and so on. Figure 10 shows the EAoIBP from La Hague along the pathways shown in Fig. 4. Since the magnitude of the EAoIBP is decreasing substantially downstream of the tracer source, we also display EAoIBP normalized by their respective maximum values to highlight the spread in the transit time direction (along the x-axis). Finally, the idealized tracer age from La Hague for the same stations are shown.

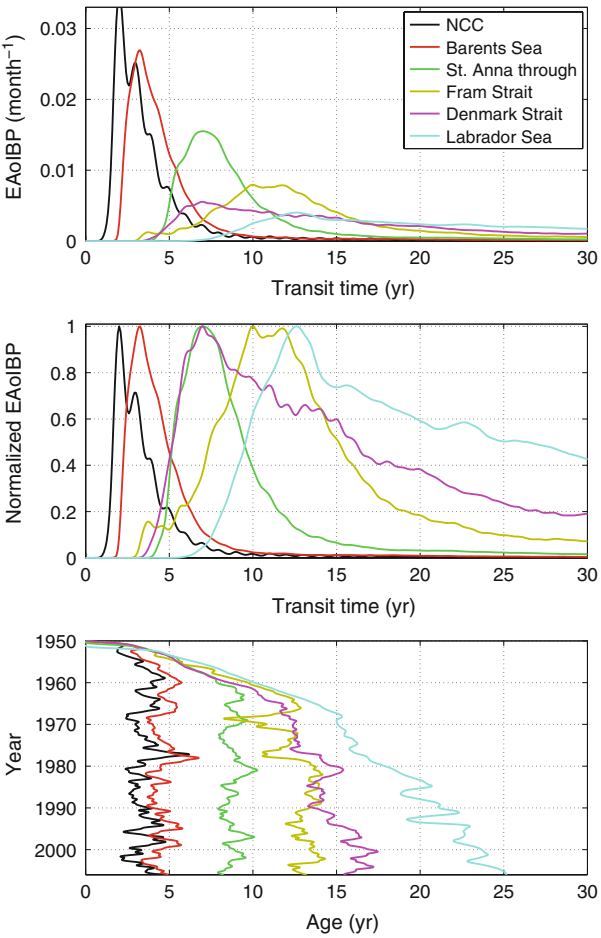


Fig. 10 (*top*) Ensemble Average of the Impulse Boundary Propagator (EAoIBP) from La Hague sampled along the transport pathway, (*middle*) EAoIBP normalized by their respective maximum values, and (*bottom*) time-series of the idealized tracer age of water masses from La Hague

Table 2 compares the mean idealized tracer age (somewhat loosely defined as the mean when the age have reached a "steady-state" value), the transit time of the EAoIBP, Γ, the width of the EAoIBP, Δ, and the degree of mixing expressed in terms of the ratio Δ / Γ.

5 Discussion

Releases of ^{129}I from Sellafield and La Hague, particularly during the 1990s (Fig. 1), causes elevated concentrations far away from the source regions. As a passive tracer, the horizontal distribution of ^{129}I (Fig. 3) can be used to shed light on the ocean circulation and mixing processes.

Integrated vertical distribution of ^{129}I from the OGCM shows the rate of dispersion from the sources to the deep ocean (Fig. 2). Interestingly, while the concentration increases relatively slowly in the upper 1,800 m, the increase below and down to about 3,500 m occurs much faster, e.g. contour lines at this depth level are almost vertical. From the horizontal distribution of ^{129}I below 2,000 m (Fig. 3) it is seen that almost all trace of ^{129}I is in the deep

Table 2 Mean ideal tracer age and the first two moments of the Ensemble Average of the Impulse Boundary Propagator (EAoIBP) from La Hague sampled along the transport pathway. The degree of mixing is calculated as the ratio Δ / Γ

Location	Mean ideal tracer age $\overline{a(\mathbf{x}, t)}$	1st mom. of EAoIBP $\Gamma(\mathbf{x})$	2nd mom. of EAoIBP $\Delta(\mathbf{x})$	Degree of mixing $\Delta(\mathbf{x})/\Gamma(\mathbf{x})$
NCC at 60° N	3.4	2.8	1.8	0.64
Barents Sea	4.6	3.9	1.9	0.49
St. Anna Through	8.7	7.8	3.3	0.42
Fram Strait	12.9	12.1	6.7	0.55
Denmark Strait[a]	14.6	∼14	∼11	∼0.79
Labrador Sea[b]	>20	∼24	∼18	∼0.75

[a] The EAoIBP have been "extrapolated" in time to reach a value close to zero
[b] The ideal tracer age in the deep Labrador Sea have not reached a "steady state", see Fig. 10, these values are therefore uncertain

North Atlantic Subpolar Seas. We therefore ascribe the overflows spilling across the Greenland-Scotland Ridge as the principal mechanism for the relatively fast vertical dispersion of ^{129}I below 1,800 m. The rate at which tracers from Sellafield and La Hague descends to depths are further quantified by the total inventory of tracers at different depth bins (Table 1). In December 2005, ^{129}I is largely constrained to the upper 500 m, while the pulse released from Sellafield and La Hague in 1950 shows a rather homogenous distribution in the vertical.

A surface transect of ^{129}I collected in July 2001 [3] revealed concentrations of roughly 3.5×10^{10} atoms l^{-1} along the Norwegian coast, and about 1.5×10^{9} atoms l^{-1} in the Nansen and Amundsen Basins. This should not only be seen as a general dilution of the tracer patch, but is also a result of the high releases of ^{129}I during the last half of the 1990s. The OGCM reproduces the observed values of ^{129}I rather realistically along the Norwegian coast and in the Barents Sea. We have also demonstrated that a previous version of MICOM reproduced a 6 year time-series of ^{99}Tc at the entrance to the Barents Sea quite satisfactory [36]. Thus, the applied model is able to reproduce key features of the dispersion properties in the eastern Nordic Seas. In the Arctic Ocean, the comparison with observed tracers is not that favorable. While a comprehensive analysis of the OGCMs performance in the Arctic Ocean is beyond the scope of this paper, we point out a few issues.

The mixed layer in the OGCM south-west of Spitsbergen becomes excessively deep in the winter, more than 1,000 m. Atlantic Waters carried by the West Spitsbergen Current are then mixed with colder and fresher waters beneath. In turn, the Atlantic Inflow through the Fram Strait is too weak, too fresh and too cold (not shown). In the Arctic Ocean, the observed cyclonic circulation in the Atlantic Layer beneath the shallow thermocline is shown to be very sensitive to lateral flux of Potential Vorticity (PV). Using a barotropic model, Yang [49] showed that modifying the bathymetry in the Fram Strait could reverse the sense of rotation in the Arctic Ocean. Using a realistic OGCM, Karcher et al. [30] confirmed the importance of lateral flux of PV for the Atlantic Layer circulation, especially in the Eurasian Basin. We believe that improving the Atlantic Inflow through the Fram Strait in MICOM would be instrumental in terms of circulation in the Arctic Ocean as a whole, and consequently to realistically reproduce the observed values of ^{129}I there.

The distinct vertical profile of ^{129}I in the Labrador Sea, with highest concentrations in the deepest layers, is a clear signature of overflow waters from the Nordic Seas (Fig. 7). From 1997 to 2005 there is a general increase throughout the water column, although the most

pronounced increase is found in the overflow layers. As noted in the previous section, the OGCM slightly exaggerates the ^{129}I concentration in the surface and in intermediate layers. We speculate that this is caused by too little tracers in the OGCM entering the Arctic, leading to an overestimation of ^{129}I in surface layers in the western Nordic Seas and the North Atlantic Subpolar Seas.

One might ask why the rapid increase of ^{129}I concentration at depth in the Labrador Sea after 2001 does not continue, as would be expected from the release rate with the highest releases in 1999 and just slightly reduced releases in the following years (Fig. 1), and since the transport from European coastal regions to the Subpolar Seas takes at least a decade. If the large scale circulation transporting tracers through the Nordic Seas and the Arctic Ocean was in steady-state, one would expect tracers to accumulate and show continuously increasing values in the deep Labrador Sea. However, there is in fact a decrease in the simulated southward volume transport of overflow waters in the Denmark Strait, from 2.5 Sv in 1999 to 1.9 Sv in 2003 (not shown). A decrease of 0.6 Sv in this period is consistent with observations [32]. Whether this decrease in volume transport is the only reason why the ^{129}I concentration is not increasing significantly after 2001 is still uncertain. A problem of fundamental character is that there are two entirely different sources of variability in the ^{129}I concentration; the release rate of ^{129}I and circulation changes, and there is no straightforward way to clearly distinguish which one of them is causing ^{129}I anomalies downstream of the source region.

The age of ^{129}I also shows a distinct vertical structure in the Labrador Sea. The general increase of the age down to about 3,000 m indicates that tracers are entering the Labrador Sea in surface waters, and being transported to greater depth with wintertime convection, e.g. formation of Labrador Sea Waters [38]. The decreasing age below 3,000 m clearly illustrates that waters at this depth have a direct source from the overflow regions, rather than being ventilated from above.

The ^{129}I distribution on the overflow isopycnals (Fig. 8) clearly illustrate that the main pathways of dense waters spilling over the Greenland–Scotland Ridge are constrained to the western boundary in the North Atlantic, consistent with the observed Deep Western Boundary Current. In less dense layers, the two eastern overflow branches (the Iceland–Faroe Ridge and the Farao-Shetland Channel) cross the North Atlantic Ridge. Previous versions of MICOM utilizing σ_0-coordinates have all displayed a serious deficiency in that the two eastern overflow branches were trapped on the eastern side of the North Atlantic Ridge [17].

To visualize and quantify the degree of mixing a tracer experience from Sellafield and La Hague in the large scale transport, we have simulated a number of idealized pulse releases (IBP, Fig. 9), and created an ensemble average of these (EAoIBP, Fig. 10). The IBPs displays a clear annual signal at the NCC and the Barents Sea stations. In the eastern Nordic Seas, tracers from Sellafield and La Hague are largely constrained to the surface (Fig. 3), and the tracer variability is therefore mainly influenced by the seasonal mixed layer dynamics. In the Denmark Strait station, sampled between 400 and 500 m depth, the variability is at much higher frequency. The southward flow in the Denmark Strait is rather unique in that wind and buoyancy forcing produce an outflow which is stratified throughout the water column, and that the local synoptic wind stress curl determines the barotropic circulation on intra-annual time-scales (e.g. [31, 44]).

Interestingly, the EAoIBP seen in the Denmark Strait (pink line in upper Fig. 10) peaks at about the same time as the EAoIBP in the St. Anna Through (green line), and approximately three years before the peak is seen in the Fram Strait (yellow line). However, the EAoIBP seen in the Denmark Strait has a very broad distribution with a long tail, and the individual IBPs seen in Fig. 9 shows that some of them even have multiple peaks. As a result, the idealized tracer age in the Denmark Strait is initially younger than the tracer age in the Fram Strait,

and roughly equal to the tracer age in the St. Anna Through. From about 1965, the idealized tracer age in the Denmark Strait is on average older than the tracer age in the Fram Strait. A plausible explanation for this behavior is that tracers transported to the Denmark Strait can follow two distinct pathways; either the *long route* for tracers carried by the NCC to the Barents Sea and then into the Arctic Ocean before leaving through the Fram Strait heading southwards with the EGC, or the *short route* transported with the NwAC and recirculating in the Fram Strait heading southwards with the EGC.

The transit time estimated as the first moment of the EAoIPB is consequently slightly younger than the mean ideal tracer age (Table 2). It has been shown that, for steady flows, the transit time Γ is initially younger than the idealized tracer age a, while at large t when $\partial a(\mathbf{x}, t)/\partial t = 0$, the transit time equals the age [23]. While this result is obtained for highly simplified flows, it indicates that the transit time could be initially younger than the idealized tracer age also in complex flows.

The idealized tracer age from La Hague in the deep Labrador Sea is reaching 25 years by 2006 (lower panel in Fig. 10), while the age of ^{129}I is less than 15 years at the same location at the same time (Figs. 7, 8). This seemingly paradoxical result is caused by the continuously increasing release rate of ^{129}I, which will then put more weight on *recent* than *older* releases from the transient tracer signal. The fact that the age (and other time-scale diagnostics) of tracers is also a function of the release history has been shown by Waugh et al. [47].

The degree of mixing expressed in terms of the ratio of the width and the transit time of the EAoIBP, Δ/Γ, revealed values in the range 0.4–0.8 along the pathways from La Hague (Table 2), showing that mixing in general is a significant contributor to the tracer transport. For comparison, Waugh et al. [48] found that the main features of the observed relationship between CFCs and tritium in the North Atlantic Ocean could be reproduced by fitting a steady-state inverse Gaussian TTD with $\Delta = \Gamma$. It should be noted that quantifying mixing as Δ/Γ cannot distinguish between different types of mixing, for instance vertical mixing or lateral mixing by mesoscale eddies or shear dispersion. Multiple pathways of tracers would also lead to larger Δ/Γ. In a simplified 1-D model with constant velocity u and diffusivity K, Delhez and Deleernijder [11] showed that Δ is directly proportional to K, and that the Peclet number[1] could be expressed as $Pe = (\Gamma/\Delta)^2$. In complex flows like the ocean this relationship may no longer hold.

Based on this study, it seems difficult to interpret tracer transport over large distances (e.g. the northern North Atlantic) as purely advective signals. From the "Great Salinity Anomaly" (GSA) as discussed by Dickson et al. [12], two-thirds of the salt deficit in the Labrador Sea is thought to pass through the Faroe-Shetland Channel by advection of the initial anomaly. However, we have seen that the degree of dilution between a tracer signal (observed and simulated) in the NCC and the Labrador Sea is approximately two orders of magnitude. Since hydrographic anomalies are large values on top of a background signal, it would be very difficult to recognize such a signal far away from the source region without further dynamical changes in the transport. Releasing passive tracers to tag the anomalously low salinity water associated with a GSAs in a coarse resolution OGCM, Wadley and Bigg [46] concluded that GSAs are unlikely to be caused by pure advection of salinity anomalies. Based on hydrographic time series, Sundby and Drinkwater [45] links both high and low salinity anomalies in the northern North Atlantic Ocean to changes in volume fluxes in and out of the Nordic Seas.

[1] The Peclet number is a dimensionless number relating the rate of advection of a fluid flow to its rate of diffusion.

Purposefully release of SF_6 in the interior Greenland Sea has revealed important information on mixing and convection processes there, as well as on the relative composition of the East Greenland Current and overflow waters from the Denmark Strait (see [35], for a recent review). It is suggested to combine information from SF_6 and ^{129}I tracers to reveal even more information on ocean processes both in the interior Nordic Seas and it's overflow regions.

6 Concluding remarks

CFCs, tritium, and other anthropogenic trace compounds have provided important insights into global ventilation of water masses in the ocean. The radioactive tracer ^{129}I has proven to be very useful for more specific studies of processes in the northern North Atlantic; the transformation of Atlantic Waters in the Nordic Seas, circulation of Atlantic derived waters in the Arctic Ocean, and dense overflow waters representing the lower limb of the Atlantic Meridional Overturning Circulation. Simulating the distribution of ^{129}I has a potential for a careful examination of an OGCMs performance in this area.

The specific findings of this study are:

- By the end of the simulation (December 2005), ^{129}I is largely constrained to the upper ocean; 80% of the integrated ^{129}I concentration is found in the upper 500 m. Integrated vertical dispersion is relatively fast below 1,800 m down to about 3,500 m, which we ascribe to the overflows spilling across the Greenland–Scotland Ridge.
- Simulated and observed values of ^{129}I agree rather well along the Norwegian coast and in the Barents Sea. In the Arctic Ocean, the comparison is not that favorable, which is ascribed to too vigorous vertical mixing south-west of Spitsbergen and an unrealistic Atlantic Inflow into the Arctic Ocean through the Fram Strait. In the North Atlantic Subpolar Seas, the OGCM reproduce the observed vertical profile of ^{129}I quite reasonably, with highest concentrations in the deepest layers.
- The isopycnic coordinate ocean model (MICOM) utilizing σ_2-coordinates clearly indicates that the distribution of overflow waters from the Nordic Seas in the North Atlantic Subpolar Seas is more realistically represented compared to previous versions utilizing σ_0-coordinates.
- A surrogate for the Transit Time Distribution based on an ensemble of simulated pulse tracers (EAoIBP) has been investigated. The two first moments of the EAoIBP revealed fundamental information on time-scales and mixing of the tracers in the area.

Acknowledgments This study is a part of the project *Arctic Radioactive Contamination* (ARC), funded by the Research Council of Norway. Support from the G. C. Rieber Foundations is acknowledged. This is contribution A234 from the Bjerknes Centre for Climate Research. We are indebted to Tor Eldevik and Helge Drange for providing internal review, and to two anonymous reviewers who greatly improved the manuscript.

References

1. Aldahan A, Alfimov V, Possnert G (2007) ^{129}I anthropogenic budget: major sources and sinks. Appl Geochem 22:606–618
2. Alfimov V, Aldahan A, Possnert G (2004a) Tracing water masses with ^{129}I in the western Nordic Seas in early spring 2002. Geophys Res Lett 31:L19305.1–L19305.4
3. Alfimov V, Aldahan A, Possnert G, Winsor P (2004b) Anthropogenic Iodine-129 in seawater along a transect from the Norwegian Coastal Current to the North Pole. Mar Pollut Bull 49:1097–1104
4. AMAP (2004) Amap Assessment 2002: radioactivity in the Arctic. Technical report, Oslo, Norway

5. Bentsen M, Evensen G, Drange H, Jenkins A (1999) Coordinate transforming on a sphere using conformal mapping. Mon Weather Rev 127:2733–2740
6. Bleck R, Smith LT (1990) A wind-driven isopycnic coordinate model of the North and Equatorial Atlantic Ocean. 1. Model development and supporting experiments. J Geophys Res 95(C3):3273–3285
7. Bleck R, Rooth C, Hu D, Smith L (1992) Salinity-driven thermocline transients in a wind- and thermohaline-forced isopycnic coordinate model of the North Atlantic. J Phys Oceanogr 22:1486–1515
8. Cooper L, Beasley T, Aagaard K, Kelley J, Larsen I, Grebmeier M (1999) Distribution of nuclear fuel reprocessing tracers in the Arctic Ocean: indication of Russian river influence. J Mar Res 57:715–738
9. de Szoeke RA (2000) Equations of motion using thermodynamic coordinates. J Phys Oceanogr 30:2814–2829
10. Deleersnijder E, Campin J, Delehez E (2001) The concept of age in marine modelling I. Theory and preliminary model results. J Mar Syst 28:229–267
11. Delhez E, Deleersnijder E (2002) The concept of age in marine modelling II. Concentration distribution function in the English Channel and the North Sea. J Mar Syst 31:279–297
12. Dickson R, Meincke J, Malmberg S, Lee A (1988) The "Great Salinity Anomaly" in the Northern North Atlantic 1968-1982. Prog Oceanogr 20:103–151
13. Drange H, Simonsen K (1996) Formulation of air-sea fluxes in the ESOP2 version of MICOM. Technical report no. 125, Nansen Environmental and Remote Sensing Center, Bergen, Norway
14. Dukowicz J, Baumgardner J (2000) Incremental remapping as a transport/advection algorithm. J Comput Phys 160:318–335
15. Dutay J, Bullister J, Doney S, Orr J, Najjar R, Caldeira K, Campin J, Drange H, Follows M, Gao Y, Gruber N, Hecht M, Ishida A, Joos F, Lindsay K, Madec G, Maier-Reimer E, Marshall J, Matear R, Monfray P, Mouchet A, Plattner G, Sarmiento J, Schlitzer R, Slater R, Totterdell I, Weirig M, Yamanaka Y, Yool A (2002) Evaluation of ocean model ventilation with CFC-11: comparison of 13 global ocean models. Ocean Model 4:89–120
16. England M, Maier-Reimer E (2001) Using chemical tracers to assess ocean models. Rev Geophys 39:29–70
17. Gao Y, Drange H, Bentsen M (2003) Effects of diapycnal and isopycnal mixing on the ventilation of CFCs in the North Atlantic in an isopycnic coordinate OGCM. Tellus 55B:837–854
18. Gao Y, Drange H, Bentsen M, Johannessen O (2004) Simulating transport of non-Chernobyl ^{137}Cs and ^{90}Sr in the North Atlantic-Arctic region. J Environ Radioact 71:1–16
19. Gascard JC, Raisbeck G, Sequeira S, Yiou F, Mork KA (2004) The Norwegian Atlantic Current in the Lofoten Basin inferred from hydrological and tracer data (^{129}I) and its interaction with the Norwegian Coastal Current. Geophys Res Lett 31(1):L01308.1–L01308.5
20. Getzlaff K, Böning C, Deng J (2006) Lagrangian perspectives on deep water export from the Subpolar North Atlantic. Geophys Res Lett 33:L21S08
21. Haine T, Hall T (2002) A generalized transport theory: water-mass composition and age. J Phys Oceanogr 32:1932–1946
22. Haine T, Zhang H, Waugh D, Holzer M (2008) On transit-time distributions in unsteady circulation models. Ocean Model 21:35–45
23. Hall T, Haine T (2002) On ocean transport diagnostics: the idealized age tracer and the age spectrum. J Phys Oceanogr 32:1987–1991
24. Harder M (1996) Dynamik, Rauhigkeit und Alter des Meereises in der Arktis. Ph.D. thesis, Alfred Wegener Institutt fur Polar- und Meeresforschung, Bremerhaven, Germany
25. Hibler W (1979) A dynamic thermodynamic sea ice model. J Phys Oceanogr 9:815–846
26. Holzer M, Hall T (2000) Transit-time and tracer-age distribution in geophysical flows. J Atmos Sci 57:3539–3558
27. Janjić ZI (1977) Pressure gradient force and advection scheme used for forecasting with steep and small scale topography. Beitr Phys Atmos 50:186–199
28. Kalnay E et al (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77(3):437–471
29. Karcher M, Gerland S, Harms I, Iosjpe M, Heldal H, Kershaw P, Sickel M (2004) The dispersion of ^{99}Tc in the Nordic Seas and the Arctic Ocean: a comparison of model results and observations. J Environ Radioact 74:185–198
30. Karcher M, Kauker F, Gerdes R, Zhang J (2007) On the dynamics of Atlantic Water circulation in the Arctic Ocean. J Geophys Res 112:C04S02
31. Köhl A, Käse R, Stammer D (2007) Causes of changes in the Denmark Strait overflow. J Phys Oceanogr 37:1678–1696

32. Macrander A, Send U, Valdimarsson H, Jonsson S, Käse R (2005) Interannual changes in the overflow from the Nordic Seas into the Atlantic Ocean through Denmark Strait. Geophys Res Lett 32:L06606.1–L06606.4
33. McDougall TJ, Dewar WK (1998) Vertical mixing and cabbeling in layered models. J Phys Oceanogr 28:1458–1480
34. McDougall TJ, Jackett DR (2005) An assessment of orthobaric density in the global ocean. J Phys Oceanogr 35:2054–2075
35. Messias M, Watson A, Johannessen T, Oliver K, Olsson K, Fogelqvist E, Olafsson J, Bacon S, Balle J, Bergman N, Budeus G, Danielsen M, Gascard J, Jeansson E, Olafsdottir S, Simonsen K, Tanhua T, Scoy V, Ledwell J (2008) The Greenland Sea tracer experiment 1996-2002: horzontal mixing and transport of Greenland Sea Intermediate Water. Prog Oceanogr 78(1):85–105
36. Orre S, Gao Y, Drange H, Nilsen J (2007) A reassessment of the dispersion properties of ^{99}Tc in the North Sea and the Norwegian Sea. J Mar Syst 68:24–38
37. Raisbeck G, Yiou F (1999) ^{129}I in the oceans: origins and applications. Sci Total Environ 237:31–41
38. Rhein M, Fischer J, Smethie W, Smythe-Wright D, Weiss C, Mertens D, Fleischmann U, Putzka A (2002) Labrador Sea Water: pathways, CFC inventory and formation rates. J Phys Oceanogr 32:648–665
39. Rudels B, Jones E, Anderson L, Kattner G (1994) On the intermediate depth waters of the Arctic Ocean. In: Johannessen OM, Muench RD, Overland JE (eds) The Polar Oceans and their role in shaping the global environment. AGU Geophysical Monographs, Washington, pp 33–46
40. Schott F, Brandt P (2007) Circulation and deep water export of the subpolar North Atlantic during the 1990s. In: Schmittner A, Chiang JCH, Hemming SR (eds) Ocean circulation: mechanisms and impacts. AGU Monograph, Washington, pp 91–118
41. Smith J, Ellis K, Kilius L (1998) ^{129}I and ^{137}Cs tracer measurements in the Arctic Ocean. Deep Sea Res 45:959–984
42. Smith J, Ellis K, Boyd T (1999) Circulation features in the central Arctic Ocean revealed by nuclear fuel reprocessing tracers from Scientific Ice Expeditions 1995 and 1996. J Geophys Res 104(C12):29663–29677
43. Smith J, Jones E, Moran S, Smethie W Jr, Kieser W (2005) Iodine 129/CFC 11 transit times for Denmark Strait Overflow Water in the Labrador and Irminger Seas. J Geophys Res 110:C05006.1–C05006.16
44. Spall M, Price J (1998) Mesoscale variability in the Denmark Strait: the PV outflow hypothesis. J Phys Oceanogr 28:1598–1623
45. Sundby S, Drinkwater K (2007) On the mechanisms behind salinity anomaly signals of the northern North Atlantic. Prog Oceanogr 73:190–202
46. Wadley M, Bigg G (2006) Are "Great Salinity Anomaly" advective. J Clim 19:1080–1088
47. Waugh D, Hall T, Haine T (2003) Relationships among tracer ages. J Geophys Res 108(C5):3138
48. Waugh D, Haine T, Hall T (2004) Transport times and anthropogenic carbon in the subpolar North Atlantic Ocean. Deep Sea Res I 51:1475–1491
49. Yang J (2005) The Arctic and Subarctic flux of potential vorticity and the Arctic Ocean circulation. J Phys Oceanogr 35:2387–2407

Environ Fluid Mech (2010) 10:235–255
DOI 10.1007/s10652-009-9135-6

ORIGINAL ARTICLE

Idealized tracer transport models with time-varying transport: applications to ocean boundary currents

F. Terenzi · Timothy M. Hall

Received: 24 November 2008 / Accepted: 17 May 2009 / Published online: 9 June 2009

Abstract One-dimensional advection–diffusion and advection–diffusion–dilution (or "leaky-pipe") models have been widely used to interpret a variety of geophysical phenomena. For example, in the ocean these tools have been used to interpret the penetration and spreading of tracers such as Chlorofluorocarbons(CFCs) along the Deep Western boundary current (DWBC). Usually, the transport coefficients of such models are taken to be constant in time, thus assuming the transport to be in steady state. Here, we relax this assumption and calculate tracer-signal variability in two simple 1D models for the boundary current having low-amplitude time-varying coefficients. Given a background tracer gradient due, for example, to a steady-state source in a boundary region, the resulting tracer field exhibits fluctuations due to the transport acting on the gradients. We compare the transport-induced tracer fluctuations to propagated fluctuations occurring in steady-state models with a periodic source in the boundary region. Using coefficients fitted to DWBC tracer observations, we find that in the North Atlantic propagated tracer fluctuations are larger, while in the sub-tropics transport-induced fluctuations dominate. This contrasts a common view that subtropical and tropical DWBC fluctuations in tracers such as CFCs, temperature and salinity anomalies are propagated signals from the northern formation region. However, the predicted transport-induced fluctuations in these models are still smaller than the observed fluctuations.

Keywords Ocean tracer transport · Deep-western boundary current · Idealized analytic models

F. Terenzi (✉) · T. M. Hall
Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY 10027, USA
e-mail: ft2104@columbia.edu

T. M. Hall
NASA Goddard Institute for Space Studies, New York, NY 10025, USA

1 Introduction

One-dimensional (1D) advection–diffusion and advection–diffusion–dilution (or "leaky-pipe) models are common idealized descriptions of tracer transport used in a wide variety of geophysical domains. In the atmosphere, New and Plumb [13] defined and solved the "Tropical Leaky Pipe" (TLP), an advection–dilution model for stratospheric transport, which explained features of the mean-age distribution present both in observations and in more complex 2D and 3D models by adding a dilution term to the 1D advection–diffusion idealized model by Plumb [15]. Mote et al. [11] obtained improved estimates of vertical diffusion and horizontal mixing by fitting CH_4 and H_2O data to a 1D advection–diffusion–dilution model of the tropical stratosphere. Also, in the context of stratospheric transport, Hall and Waugh [7] computed the TLP residence time and analyzed its relationship to mean-age. In the troposphere, McKenna [9] used similar concepts to explain the dilution of pollution plumes.

In ocean applications, tracer transport along isopycnals [18] and anthropogenic carbon [e.g., 6] have been addressed with the use of 1D advection–diffusion models. Idealized models have also been exploited in order to interpret the propagation of tracer signals that enter the DWBC in the North-Atlantic (NA) formation regions and propagate along the DWBC into the deep tropics [3,5,14,16,21]. In particular, these models were used to interpret anomalies of tracers such as temperature (T) and salinity (S), as well as to explain discrepancies among measured and modeled velocity fields. It was recognized that differences among velocity estimates are due to mixing and recirculation. More recently, Mouchet and Deleersnijder[12] used a 1D advection–diffusion model, in which the deep ocean is represented as a leaky pipe with a decreasing cross section (i.e. a leaky funnel), allowing recirculation of water and tracers toward the surface, in order to study the ventilation rate in the ocean as simulated by a 3D GCM.

Advection–diffusion and advection–dilution analytical models are highly simplified, kinematic descriptions of complex fluid flow and tracer transport. Nonetheless, they offer a convenient interpretation of tracer measurements and an efficient way to explore the sensitivity of tracers to bulk parameters of the flow, which cannot be easily tuned independently in 3D numerical models. In all analyses that we know of, transport coefficients of such models are taken to be constant; that is, the transport is assumed to be in a steady state. This is warranted if the time-averaged behavior is stationary, but it precludes the analysis of certain classes of tracer signals if the transport fluctuates, as we describe below.

In this work we generalize common 1D advection–diffusion and advection–dilution models to include fluctuations in transport coefficients. Our focus is the NA DWBC, and our primary goal is to compare and contrast two mechanisms for fluctuations in tracer concentration: (1) local generation due to fluctuations in transport acting on background tracer gradients; and (2) propagation of tracer fluctuations generated in northern source regions. The effects of mechanism 2 can be accommodated using constant transport coefficients, while mechanism 1 requires time-varying coefficients.

The motivation for comparing these mechanisms is observations of temporal variations in tracers such as CFCs and T and S in the DWBC. CFCs enter the DWBC in NA source regions, and their propagation along the DWBC is a valuable diagnostic of transport. A number of studies have had success constraining advection–dilution and advection–diffusion models with CFC measurements [14,16,17,21]. On the other hand, Steinfeldt and Rhein [17], who analyzed repeated measurements of CFC-11 and CFC-12 in the western tropical Atlantic from 1990 to 2002, observed fluctuations in concentrations that they could not explain using an analysis with constant transport rates. Several studies [2,4,8,10] have interpreted subtropical and tropical DWBC T and S fluctuations as signals propagated from the northern source

regions, where they are generated by variability in air–sea interactions. However, using an advection–dilution model with constant coefficients tuned to CFC and helium–tritium observations, Waugh and Hall [21] were not able to replicate T and S fluctuations. Their analysis indicated that fluctuations propagated from northern source regions should be attenuated beyond detection in the subtropics. Could tracer fluctuations such as these be generated by mechanism 1, transport fluctuations acting on background gradients?

A second goal of our study is to provide solutions to ubiquitous models in the case of time-varying transport that may be applicable to a wide range of tracer analyses. A third goal is to compare and contrast advection–dilution and advection–diffusion models in the case of temporally varying coefficients.

This paper is organized as follows: In Sect. 2 we define both the advection–dilution and advection–diffusion models and compare and contrast their behavior. In Sect. 3, results for both models are discussed for two cases: (1) periodic fluctuations in transport with linearly increasing boundary conditions (BCs) and (2) steady-state transport with periodic BCs. A comparison of CFC observational data and predictions by both models is presented in Sect. 3.3. Finally, a discussion of the main findings of this study follows in Sect. 4. Details on the derivations of the main analytical solutions can be found in the Appendices.

2 Model descriptions

We extend idealized advection–diffusion and advection–dilution 1D models to include small temporal variations in the transport coefficients. Both models can be represented in vector form as:

$$\begin{aligned} \frac{\partial}{\partial t}\mathbf{c} + (1 + \epsilon f\,(t))\,\mathbf{L}\,(\mathbf{c}) &= 0 \\ \mathbf{c}\,(x, 0) &= 0 \\ \mathbf{c}\,(0, t) &= \mathbf{y}\,(\mathbf{t})\,, \end{aligned} \tag{1}$$

where $\mathbf{c} = \mathbf{c}\,(\mathbf{x}, \mathbf{t})$ is the tracer concentration (units tracer mass per unit mass), $\mathbf{L}$ is the steady-state transport operator, ϵ is the dimensionless perturbation parameter ($\epsilon \ll 1$), and $f\,(t)$ is a dimensionless time-dependent function of order one, which summarizes time-variation in transport. Here we consider the case in which a passive tracer penetrates the ocean by air–sea gas exchange at the sea surface. This process, along with transport near the surface, sets a near-surface concentration, that is assumed to be known. This is the boundary condition, BC, at $x = 0$. Initial conditions are applied at all $x > 0$.

2.1 1 D Advection–diffusion model

The BC sets a signal that is transported into the domain by a constant uniform flow along an isopycnal that extends indefinitely in the positive x direction and by diffusive isopycnal mixing. The diffusion coefficient k is positive, and the bulk advection u is non-negative. The transport operator in Eq. (1) is:

$$L = u\frac{\partial}{\partial x} + k\frac{\partial^2}{\partial x^2} \tag{2}$$

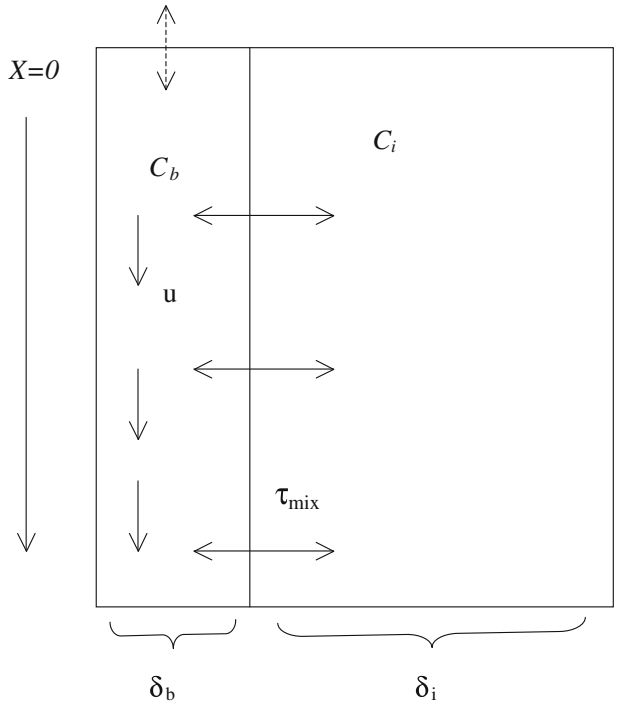


Fig. 1 Schematic of the "leaky-pipe" boundary-current model [21]. The model consists of two coupled one-dimensional semi-infinite domains: a boundary-current core region, "b", with uniform along velocity u and cross-sectional area δ_b, and a larger interior region, "i", with cross-sectional area δ_i and no flow. The cross sectional area of the core to interior region is the ratio $\alpha = \delta_b/\delta_i$. Tracer concentrations mix relaxationally between the two regions at a rate $\sigma = \tau_{\text{mix}}^{-1}$. $c_b(x,t)$ and $c_i(x,t)$ are the tracer concentrations in the boundary current and interior region, respectively

2.2 1 D Boundary-current model ("leaky pipe")

Following Waugh and Hall [21] we define an idealized boundary-current model that consists of two coupled one-dimensional semi-infinite domains: a boundary-current core region with uniform along-stream (positive) velocity u and a stationary surrounding interior ocean region (see illustration in Fig. 1). Tracer concentrations mix relaxationally between the two regions at a rate $\sigma = 1/\tau_{\text{mix}}$, where τ_{mix} is a relaxational time scale. Along-flow diffusion is neglected, as scaling arguments show that is much slower than lateral mixing [14]. As already mentioned, similar models have been used in previous studies for extracting transport rates from tracer observations [3,14], although generally in these models the surrounding interior has been assumed to be an infinite reservoir. We now have to keep track of tracers in two regions: the interior region "i" and the boundary region "b". The vector that represents the tracer concentration $\mathbf{c}$ is:

$$\mathbf{c} = \begin{cases} c_b \\ c_i \end{cases} \tag{3}$$

The transport operator in Eq. (1) can be written as:

$$\mathbf{L} = \begin{pmatrix} u\frac{\partial}{\partial x} + \sigma & -\sigma \\ -\alpha\sigma & \alpha\sigma \end{pmatrix} \tag{4}$$

where α is the cross-sectional area ratio of the core region to the surrounding interior region.

2.3 Physical model comparison

Although their domains of application sometimes overlap, the two models are structurally different and fitted coefficients must be interpreted distinctly. In oceanographic applications the 1D advection–diffusion model has often been used to summarize transport along 2D thermocline isopycnals. In this context the model velocity may only be related to a physical current velocity if the model is applied locally along a streamline of 2D flow. In such an application the along-flow diffusivity would crudely represent the effects of mixing across streamlines. If the model is applied to represent the cross-basin averaged transport away from the outcrop, then the coefficients have no simple interpretation, and the model is merely a convenient two-parameter form that empirically describes tracer transport better than either pure diffusion or pure bulk advection. The advection–dilution model of boundary currents is more physical structurally, in that the model velocity can be identified with the velocity along the fluid plume, if many simplifying assumptions are made. The relaxation time-scale represents crudely the mixing between the boundary and the much larger surrounding regions. The two models can produce similar tracer distributions (the interior region of the advection–dilution model redistributes tracer of core-origin to different regions of the boundary, similar to the effects of along-flow mixing), but the best-fit coefficients will not be the same.

2.4 Perturbation solutions

In order to solve Eq. (1) analytically, we perform a perturbation of order $O(\epsilon)$; that is, we consider solutions of the form $\mathbf{c} = \mathbf{c^{(0)}} + \epsilon\mathbf{c^{(1)}}$. Eq. (1) is thus separated into zero-th and first-order components, both of which can be solved analytically. In the spirit of idealized models we choose a periodic transport perturbation $f = \mathrm{Re}\left(e^{i\omega t}\right)$, capturing one component of the spectrum of variations. (Subsequently, we examine the sensitivity to ω.) For the BC we choose a linear increase, $y(t) = \gamma t$, crudely representing the history of anthropogenic tracers such as CFCs and industrial CO_2 for at least parts of their histories. (Linear is chosen over exponential because it simplifies the solutions greatly.) The zero-th and first-order solutions can be written as

$$\mathbf{c^{(0)}}(x,t) = \gamma\left(t - \mathbf{\Gamma}(x)\right) \tag{5}$$

and

$$\mathbf{c^{(1)}}(x,t) = \mathbf{A^{(1)}}(x)\cos\left(\omega\left(t - \boldsymbol{\tau}^{\mathbf{(1)}}(x)\right)\right), \tag{6}$$

where $\mathbf{\Gamma}$ is the mean transit time [20], $\boldsymbol{\tau}^{\mathbf{(1)}}$ is the phase lag and $\boldsymbol{A}^{\mathbf{(1)}}$ is the amplitude (Appendix A).

2.5 Parameter values

The parameter values used to perform the calculations for both models are reported in Table 1. The values for the boundary-current model are chosen in order to be consistent with the study of Waugh and Hall [21]. Waugh and Hall [21] showed that the parameter combination:

Table 1 Parameter values used in the calculations for the two models

Model	u [cm^2/s]	τ_{mix} [years]	k [m^2/s]	γ [tracer unit/year]	α	T_p [years]
Boundary current	5	1	–	1	0.1	10
		10	–			
		∞	–			
Advection–diffusion	0.5	–	10^{-7}	1	–	10
		–	10^2		–	
		–	10^3		–	

$u = 5\,\mathrm{cm\,s^{-1}}$, $\alpha = 0.1$ and $\tau_{mix} = 1$ year for the boundary-current model gives a good agreement to a range of tracer observations. For the advection–diffusion model, we chose $u = 0.5\,\mathrm{cm\,s^{-1}}$, the value indicated by [21]. This choice results in an agreement between $\Gamma = x/u$, the mean transit time for the advection–diffusion model, and Γ in the boundary region of the advection–dilution model,

$$\Gamma = \tau_{adv}\left(1 + \frac{1}{\alpha}\right). \tag{7}$$

Waugh and Hall [21] also found an upper bound for the diffusion coefficient of $k \sim 10^4\,\mathrm{m^2\,s^{-1}}$. The high value of k and the small value of u are probably due to the fact that the large along-flow diffusion contributes to the tracer propagation, as also noticed in other studies (e.g., [17]). Here we span a wide range of diffusivities (from 10^{-7} to $10^3\,\mathrm{m^2\,s^{-1}}$) to allow for different regimes and document sensitivity. The highest value of k in our range is the "standard" value, equal to the value used by Steinfeldt and Rhein [17], which is more realistic than the low value. The low value is well within the weak mixing limit. (Note that k cannot be set to zero identically because the solutions are singular.) The rate of linear increase of the tracer at the origin is $\gamma = 1$ tracer unit per year. The perturbation magnitude is $\epsilon = 1/3$. The period of transport oscillations, $T_P = 2\pi/\omega$, is 10 years. In what follows, it is understood that the parameters used are those presented in this section, unless otherwise stated. Sensitivity to the choice of some of the parameters is also shown in Sect. 3.

3 Results and discussion

3.1 Periodic fluctuations in transport

The increasing boundary condition establishes background gradients in tracer concentration upon which the transport fluctuations act. Shown in Fig. 2 is the tracer's time series at a fixed location ($x \sim 8{,}000$ km) for the boundary component of the boundary-current model. Not surprisingly, the amplitude of the induced fluctuations depends on the dilution coefficient τ_{mix}. Fig. 3 shows the profiles of the amplitude, $\epsilon A_1\,(x)$, and phase lag, $\tau_1\,(x)$, as a function of non-dimensional distance x, with transport oscillation period $T_p = 10$ years. The dilution coefficients for the boundary-current model are: $\tau_{mix} = \infty$ years (red), $\tau_{mix} = 10$ years (black) and $\tau_{mix} = 1$ year (blue). The diffusion coefficients for the advection–diffusion model are: $k = 10^{-7}\,\mathrm{m^2\,s^{-1}}$ (red), $10^2\,\mathrm{m^2\,s^{-1}}$ (black) and $10^3\,\mathrm{m^2\,s^{-1}}$ (blue). Dimensionless distance x is defined as the ratio of the physical position along the DWBC current to the distance, uT_p, a parcel of water with speed u would cover in time T_p. For $T_p = 10$ years and $u = 5\,\mathrm{cm\,s^{-1}}$, this distance is $\sim 16{,}000$ km. Thus, for the advection–dilution model with these parameter

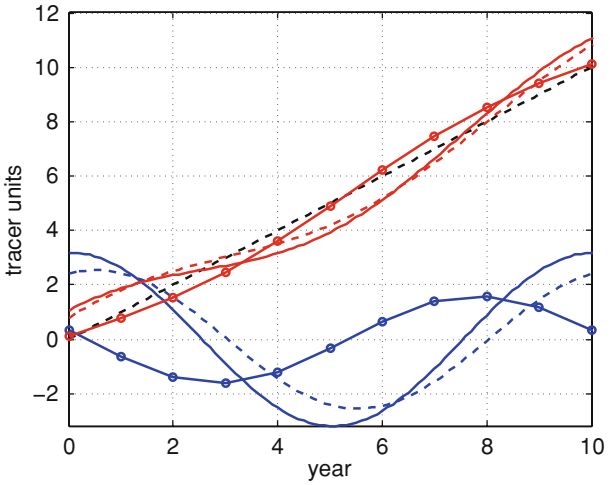


Fig. 2 Tracer concentration time series for the boundary-current model at a fixed location, $x = 8,000$ km: zero-th order solution $c_0(t)$ (*black solid line*); first order solution $\epsilon \cdot c_1(t)$ (*blue lines*) and full solution $c(t) = c_0(t) + \epsilon \cdot c_1(t)$ (*red lines*). The advective velocity is $u = 5\,\mathrm{cm\,s^{-1}}$. Solutions for different values of the dilution coefficient are plotted with different symbols: $\tau_{\mathrm{mix}} = \infty$ (*solid line*), τ_{mix} =10 years (*dash*) and τ_{mix} =1 year (*circles*)

values, only the early part of the first amplitude cycle can be realized physically. (For example, Steinfeldt and Rhein [17] report that for Labrador Sea Water the pathway from the Labrador Sea to 16°N is about 8,500 km, and the pathway to 10°S is 13,000 km. The authors define the distance from the source region as the shortest pathway following the DWBC along the coast line.)

The amplitude and phase profiles in the two models behave similarly. The x structure in both models attenuates with distance from the source and with increasing mixing. We give a physical interpretation of the behavior of both $\epsilon A_1(x)$ and $\tau_1(x)$ for the boundary-current model, as simpler analytical expressions can be found for this model. The curves of Fig. 3a and b illustrate two noteworthy limits, weak and strong mixing via large and small τ_{mix} (small and large σ).

First, consider the weak mixing limit:

$$\sigma \ll \frac{u}{x}\left(\frac{\omega^2 + \alpha^2\sigma^2}{\omega^2}\right) \tag{8}$$

The amplitude becomes:

$$A_b^{(1)} \sim \sqrt{2}\frac{\gamma}{\omega}\left(1 - \cos\left(\frac{\omega\beta_+ x}{u}\right)\right)^{1/2} \tag{9}$$

(see Eq. (27) in Appendix). In this limit, the amplitude oscillates in x; that is, the tracer fluctuation magnitude is modulated by an envelope with periodic structure in x (red curve, Fig. 3a). At short distances from the origin, the boundary current has only had time to experience one phase of the transport perturbation. The further into the domain the water travels, the more time it spends in this phase and the greater the tracer perturbation. Consequently, the amplitude increases with x. However, water at $x > uT_{\mathrm{P}}/2$ has been in the domain long enough to also experience the opposite phase of the transport perturbation, which cancels

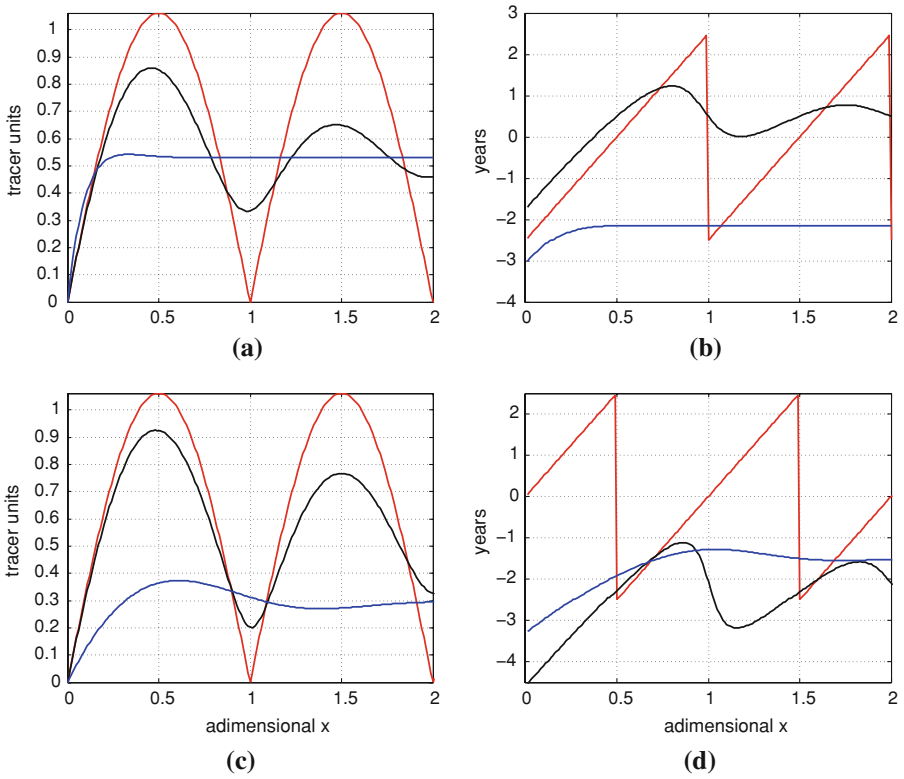


Fig. 3 The x-dependence of (**a**) amplitude $\epsilon A^{(1)}$ for the boundary-current model; (**b**) phase lag $\tau^{(1)}$ for the boundary-current model; (**c**) amplitude $\epsilon A^{(1)}$ for the advection–diffusion model, and (**d**) phase lag $\tau^{(1)}$ for the advection–diffusion model. For the boundary-current model the dilution coefficients are: $\tau_{\rm mix} = \infty$ (red), $\tau_{\rm mix}$ =10 year (*black*) and $\tau_{\rm mix}$ =1 year (*blue*). For the advection–diffusion model, the diffusion coefficients are: $k = 10^{-7}\,{\rm m}^2\,{\rm s}^{-1}$ (*red*), $k = 10^{2}\,{\rm m}^2\,{\rm s}^{-1}$ (*black*) and $k = 10^{3}\,{\rm m}^2\,{\rm s}^{-1}$ (*blue*). Other parameters are $T_{\rm p}$ =10 years, γ =1 tracer unit year^{-1}, and $u = 5{\rm cm\,s}^{-1}$ and $u = 0.5{\rm cm\,s}^{-1}$ for the boundary-current and the advection–diffusion model, respectively. Distance x is non-dimensionalized by the length scale $uT_{\rm p}$

the effect of the first phase, and the amplitude decreases with x. At $x = uT_{\rm P}$, the effects on the tracer of the two phases exactly cancel, and the amplitude is zero. For $x > uT_{\rm P}$ the cycle repeats.

In the strong mixing limit, on the other hand, water at x has a wide range of times since boundary contact, because of the possibility of a wide range of times spent in the interior region "en route" to x. Consequently, there is a wide range of phase of the transport perturbation and the net effect is cancellation and the loss of spatial structure in amplitudes. In this limit the tracer perturbation amplitude is simply determined by the amount of time the transport perturbation acts on the background tracer trend, i.e., $A_b^{(1)} \sim \gamma/\omega$. The mixing rate, σ, required to be in this limit, $\sigma \gg u/x$, depends on x. For small x, there has been little time since boundary contact for mixing to act, so that the mixing rate must be greater to be in this limit. This limit is illustrated by the blue curve in Fig. 3a for non-dimensional $x \sim 0.3$, which corresponds to $x \sim$5,000 km, $\sigma \gg 1 \cdot 10^{-8}\,{\rm s}^{-1}$, or $\tau_{\rm mix} \ll 3$ years.

Similar arguments apply to the phase-lag time, $\tau^{(1)}(x)$. Consider the unmixed case. The tracer perturbation at x results from integrating the transport perturbation over the time

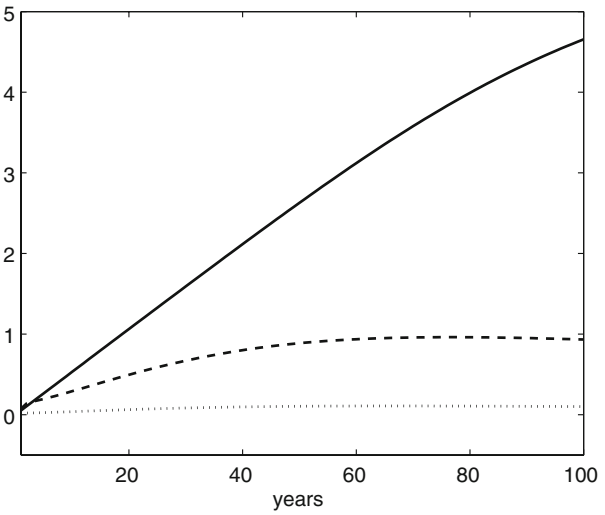


Fig. 4 Amplitude ϵA^1 of transport-induced fluctuations for the boundary-current model versus the period of transport oscillations T_P for $x = [10^2, 10^3, 10^4]$ km (*dotted*, *dashed* and *solid lines*, respectively). $\tau_\mathrm{mix} = 1$ year, while all other parameters are as in Table 1. Dimensions are tracer units

required to advect the unmixed parcel to x from the origin, namely, x/u. As x increases, the integration time increases, and an increasingly large range of perturbation phases have been sampled. The resulting average tracer phase becomes increasingly different from the contemporaneous transport phase; that is, $\tau^{(1)}(x)$ increases with x. However, at $x = uT_\mathrm{P}$, all transport phases have been equally sampled, the phase is indistinguishable from that at $x = 0$, and the cycle repeats. In the strong mixing limit, this argument only applies at small x. Once an x is reached so that $\sigma \gg u/x$, then there is a wide range of transit times, and the net result is a lack of any further x-dependence of $\tau^{(1)}$. In this limit,

$$\tau_b^{(1)} \sim -\frac{1}{\omega} \tan^{-1}\left(\frac{2\sigma\omega(\omega^2 + \alpha^2\sigma^2 + \alpha\sigma^2)}{\sigma^2\omega^2 - (\omega^2 + \alpha^2\sigma^2 + \alpha\sigma^2)^2}\right), \tag{10}$$

is dependent on σ and α, but not x. Similar arguments apply for the advection–diffusion model for both the amplitude and the phase lag (Fig. 3c, d). The difference is that for advection–diffusion the range of transit times in the well-mixed limit is due to the 1D random-walk in the flow direction, rather than the range of times spent in any external reservoir (the interior region).

We show in Fig. 4 the dependence of the tracer-fluctuation amplitude on the forcing period, T_P, for $\tau_\mathrm{mix} = 1$ year at three different x. Three features are clear. (1) At small T_P the amplitude increases linearly with T_P. In this regime the magnitude of a tracer fluctuation is simply determined by the time over which the transport fluctuation acts on the background gradient. The more time in one phase, the greater the tracer perturbation. (2) At large T_P the amplitude is independent of T_P. In this limit, the effect of the transport fluctuation on the tracer gradient has saturated. The variation of the transport is now so slow that, over the entire range of times it takes fluid to reach x from the origin, the transport phase is approximately constant. In this limit, the transport perturbation is equivalent to a slow modulation of the background transport rates, the tracer distribution is in approximate equilibrium with these rates, and zero-order equations suffice to represent the system. (3) The T_P at the transition from one limit to the other increases with x. Larger x corresponds to longer transit times to arrive at x, and, hence, slower transport fluctuations are required to be in the saturated limit. In other words, as x increases at constant

T_P, more phase variation of the transport fluctuation is experienced en route to x from the origin.

Finally, we note that there are other limits for the amplitude and phase not discussed here. The general dependence on the three time-scales, T_P, τ_{mix}, and the advective time $\tau_{adv} = x/u$, results in a rich variety of behavior. Here, we have explored only the behavior that is plausibly relevant to analyses of the DWBC. In other applications other regimes may be more relevant.

3.2 Comparison of tracer-fluctuation mechanisms

In this section we compare two distinct mechanisms to generate temporal fluctuations in tracer concentration: (1) fluctuations in transport acting on background tracer gradients due to a steady source, and (2) fluctuations in tracer source causing fluctuations in concentration in the source region, which are then propagated downstream by steady transport.

Mechanism 1 has been examined above with our two idealized models. To examine mechanism 2 we use the same models, but now with steady transport ($\epsilon = 0$) and a periodic boundary condition on tracer concentration applied at the origin. Solutions can be written $c(x,t) = P(x)\cos(\omega_*(t-\phi(x)))$, where ω_* is the boundary condition frequency. Expressions for $P(x)$ and $\phi(x)$ are found in Appendix C. To compare fluctuating tracer mechanisms, we set ω_* to the frequency of the periodic transport fluctuation of mechanism 1; that is, $\omega_* = \omega = 2\pi/10$ years. This is reasonable if the same mode of climate variability (e.g., the North-Atlantic Oscillation) is the origin for both the transport fluctuation and the source-region air–sea exchange fluctuation. For concreteness, we also set the mechanism 2 source-region amplitude, $P(0) = \gamma T_P$. This equality is not crucial. As we shall see, at some point downstream, mechanism 1 become larger than mechanism 2. The smaller $P(0)$ is, the closer to the origin this crossover point occurs.

Figures 5a and c show the x profiles of the two mechanisms' transport fluctuation amplitudes, P and $\epsilon A^{(1)}$, for the two models. The mixing rates are $\tau_{mix} = 1$ year, $k = 10^3\,\mathrm{m^2\,s^{-1}}$, and other parameters are as in Fig. 2. Waugh and Hall [21] found that these parameters could reproduce observed tracer distributions in the DWBC. In both models P decreases exponentially with x, while $\epsilon A^{(1)}$ increases from zero and asymptotes to constant values, ~0.5 for the boundary current model and ~0.3 for the advection–diffusion model. Fig. 5b and d show the ratios $\epsilon A^{(1)}/P$ as functions of x for the two models. In both models, mechanism 2 (propagated fluctuations) dominates tracer fluctuations at small x and mechanism 1 (in situ transport-generated fluctuations) dominates at large x. For the boundary-current model, the crossover point where $\epsilon A^{(1)} = P$ is $x \approx 4,500$ km, while for the advection–diffusion model it is $x \approx 3,000$ km. To examine further parameter sensitivity of these crossover points we contour in Fig. 6 the ratio $\epsilon A^{(1)}/P$ of the boundary-current model against the boundary current speed, u, and the mixing timescale, τ_{mix}, evaluated at the position $x = 15,000$ km. Slower current speed and more rapid mixing favor transport-induced fluctuations (mechanism 2), while faster current speed and less mixing favor propagated fluctuations (mechanism 1).

The crossover distances 3,000 km (advection–diffusion) and 4,500 km (boundary current) are both less than the distance along the DWBC from the Labrador Sea to the subtropics. We conclude that in the subtropics tracer fluctuations induced by transport fluctuations dominate over fluctuations propagated from North-Atlantic source regions, for a class of tracers with sufficiently large background gradients. The question to which we now turn is whether these fluctuations can explain certain observed subtropical fluctuations.

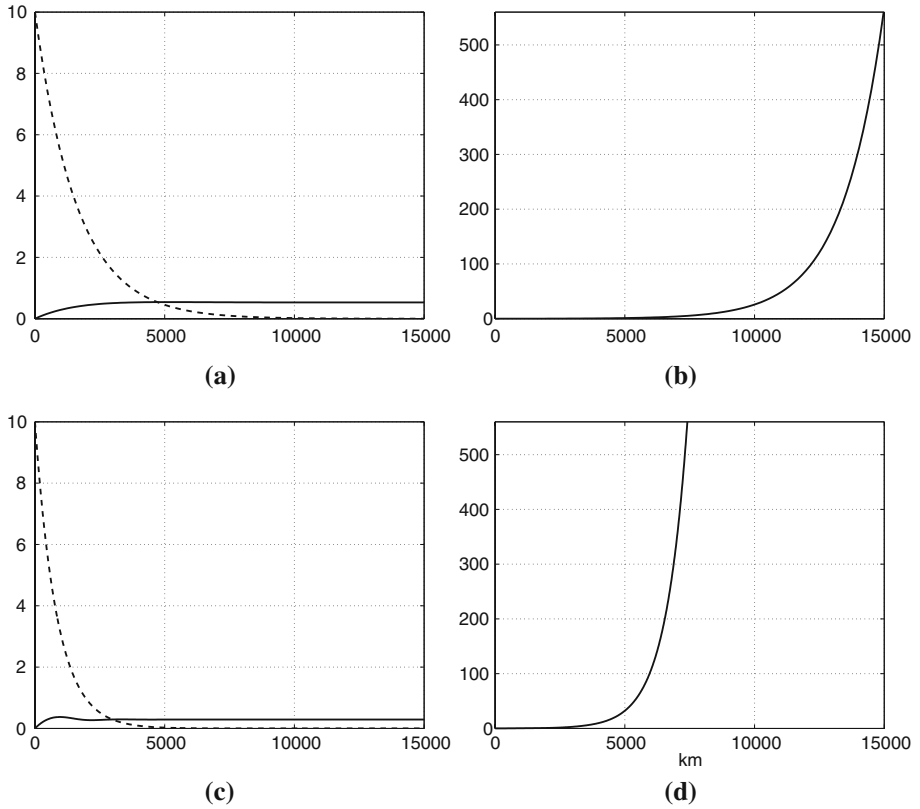


Fig. 5 (**a**) The amplitudes in the boundary current model, ϵA^1, of transport-induced fluctuations (*solid black*) and P of a tracer-signal propagating in response to periodic variation in boundary condition at $x = 0$ (*dashed black*); (**b**) The ratio $\epsilon A^1/P$ for the boundary current model; (**c**) and (**d**) equivalent to (**a**) and (**b**) for the advection–diffusion model. For the boundary-current model $\tau_{\text{mix}} = 1$ year and for the advection–diffusion model $k = 10^3\ \text{m}^2\ \text{s}^1$. Other parameters as in Fig. 3. The propagated tracer's amplitude at $x = 0$ is 10 tracer units, and its period of oscillation is 10 years, identical to the period of transport variation

3.3 Relation to observations

Steinfeldt and Rhein [17] present and analyze a series of CFC-11 and CFC-12 measurements in the western tropical Atlantic within the period 1990–2002. Here, we analyze their CFC-11 at two sites: "44°W", a transect straddling the equator at 44°W, and "10°S", a transect running about 5° east from 35°W at approximately 10°S (see Fig. 2 of [17]). Observed CFC-11 core concentrations for three NADW components and their standard deviations are shown in black in Fig. 7. The three NADW components are: LSW (Labrador Sea Water), ULSW (Upper LSW) and LNADW (Lower NADW). On average, CFC-11 increases in time at all sites and water masses, and Steinfeldt and Rhein [17] are able to capture well this increase by fitting a 1D advection–diffusion model—the same presented here in the unperturbed case—forced by boundary conditions at the remote northern outcrops. In addition to the overall increase, however, there are fluctuations, particularly the post-2000 decline seen in ULSW and LNADW. Steinfeldt and Rhein [17] note that these fluctuations cannot be captured by

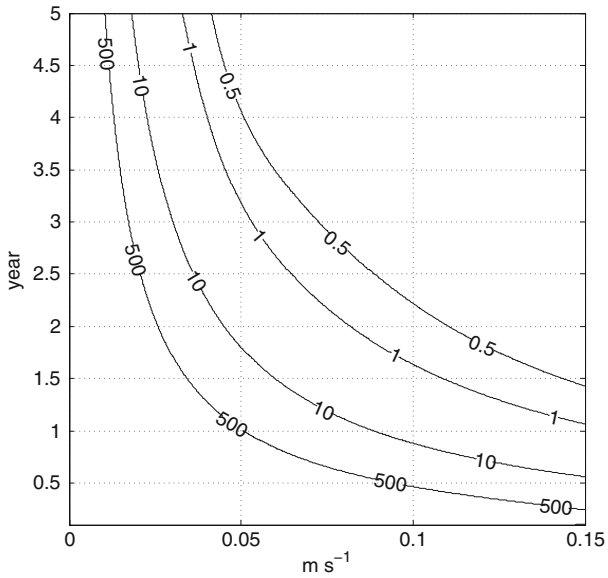


Fig. 6 Ratio between the amplitudes of transport-induced fluctuation, ϵA^1, and periodic amplitude, P, at $x = 15{,}000$ km versus u and $\tau_{\rm mix}$ for the boundary-current model. Other parameters as in Fig. 3

the steady model, and it is these fluctuations that we attempt to explain with our non-steady boundary-current model.

Also shown in Fig. 7 are model time series evaluated at values of x equal to the distance along the boundary current from the northern outcrop to the observational site. These values are 11,000 km to the 44°W site and 13,000 km to the 10°S site for ULSW/LSW and 2,500 km longer to each site for LNADW (Reiner Steinfeldt, personal communications). The blue curve corresponds to the steady-state model, with "best-value" parameters from Waugh and Hall [21]: $\tau_{\rm mix} = 1$ year, $\alpha = 0.1$, and $u = 5\,{\rm cm\,s^{-1}}$. The linear increase, γ, is set to match the linear trend over the time span shown in each panel. The solid red curve is the "best-estimate" fluctuating solution, using the same parameters as steady state and, in addition, a transport oscillation period of $T_{\rm P} = 10$ years. The value of $T_{\rm P}$ is somewhat arbitrary, and in reality a spectrum of transport fluctuations is present. Here, our goal is merely to determine the plausibility of a decadal-scale transport-fluctuation mechanism. Our choice of 10 years allows us to focus on interannual variability but is significantly shorter than the time-scale that has established the background CFC gradient (∼60 years). The observed fluctuations in Fig. 7, although poorly resolved by the infrequent measurements, suggest periods between 6 and 12 years. The perturbation magnitude for the "best-estimate" case is $\epsilon = 1/3$; i.e., as large as possible, but still comfortably consistent with the linearization. The dashed red curve is the "envelope-pushed" solution to maximize the fluctuating tracer amplitude, in which $\tau_{\rm mix} = 3$ years, the largest value still roughly consistent with Waugh and Hall [21], and $\epsilon = 1/2$, pushing the limit of linearity.

There is no model-observational agreement in detail, nor should any be expected, as we have not attempted to model any particular fluctuation in transport (for example, some observed fluctuation in DWBC core velocity). Moreover, even aside from disagreement in detail, our model cannot match the magnitude of post-2000 CFC–11 decline in concentration at 10°S in ULSW and LNADW. The smaller observed fluctuations at 44°W for all water masses and 10°S for LSW have magnitudes that can roughly be matched by the model with "envelope-pushed" parameters. It may be that the observed fluctuations are the result of a

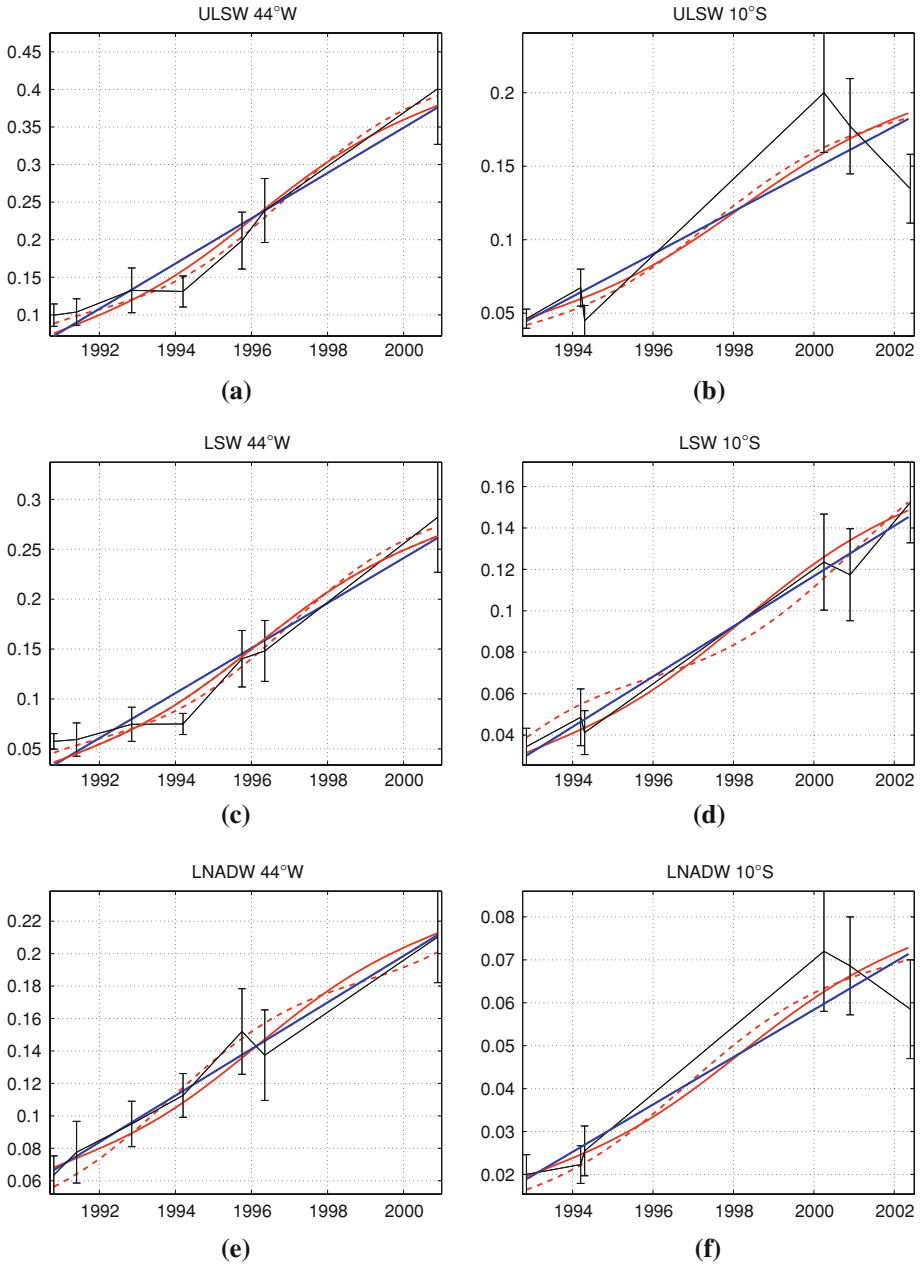


Fig. 7 Comparison between measured and modeled CFC-11 concentrations at stations "44°W" and "10°S". Observed core concentrations for ULSW, LSW and NADW components and their standard deviations for the period 1990–2002 are shown in *black* (data from [17]). *Blue lines* represent the zero-order solution $c_0(x,t)$ of the boundary-current model, assuming linear increase for the boundary condition, and red lines represent the total perturbed solution up to first-order, $c = c_0(x,t) + \epsilon \cdot c_1(x,t)$. The period of transport oscillation is $T_p = 10$ years. *Red solid lines* correspond to "best-estimate" choice of parameters ($\tau_{mix} = 1$ year, $\epsilon = 1/3$), and *red dashed* correspond to $\tau_{mix} = 3$ years and $\epsilon = 1/2$. See text for details on the adopted values of x and u. Units are in picomole kg^{-1}

complete different mechanism, the intermittent measurements sampling eddies that carry water from outside the DWBC into the domain of observation. This water would have very different CFC concentration. Such a process is well beyond the scope of the simple boundary current model, in which the bulk-averaged effect of such eddies is summarized by a single mixing time-scale. The time-variation of our model's coefficients can only be expected to replicate slow modulation of the averaged mixing effect of the eddies. Unfortunately, the frequency of measurement in the Steinfeldt and Rhein [17] data is insufficient to distinguish any low frequency modulation of net-eddy effects from the high-frequency signals of individual eddies.

Several studies have pointed out the presence of a DWBC bifurcation near the equator into an eastward and a southward branches [22]. Although the exact location and extent of the bifurcation are still under debate (e.g., [1]), Steinfeldt and Rhein [17] see some evidence of this circulation feature for LNADW from the equator at 35°W to 10°S. This introduces a third additional possible mechanism to temporal tracer fluctuations, i.e., that these can be driven by temporal variability in the proportion of the DWBC that heads East, rather than that continues South (to 10°S). Although a quantitative study of the bifurcation of the DWBC is of considerable interest, we do not address it here. This analysis should be viewed as a first attempt to quantify tracer-fluctuations mechanisms. In this spirit, each of the 1D models presented in this work could be separately applied to each branch of the DWBC.

Finally, we have assumed a linearly increasing boundary condition to obtain the analytic solutions for the perturbed models. In reality, increasing anthropogenic tracers (e.g., CFCs and anthropogenic CO_2) have surface-water time variations that vary considerably from linearity. Anthropogenic CO_2 has increased approximately exponentially. CFCs had increased exponentially, until leveling in the late 1990s. Nonetheless, linearity offers a first approximation over one or two decades and permits us to solve the models and obtain at least a qualitative description of the effects of transport variations acting on background tracer gradients. In this way, competing explanations for observed tracer variability (variations in formation rates versus variations in transport) can be tested. Here, we are interested in basic testing of mechanisms, sensitivity to parameters and qualitative comparison to observations. For more detailed studies of tracer variations, numerical solutions to the advection–diffusion and/or advection–dilution models could be obtained using more realistic tracer boundary conditions and transport fluctuations. For example, the technique to estimate anthropogenic CO_2 uptake in the Labrador Sea by Terenzi et al. [19], which uses a steady-state advection–diffusion model and is driven by observed atmospheric carbon history, could be generalized to include decadal transport variations. In stratospheric tracer analysis, the effects of transport variations due to natural mechanisms (e.g., quasi-biennial oscillation) and anthropogenic mechanisms (e.g., secular changes in the Brewer–Dobson circulation associated with greenhouse forcing) on the mean-age in the tropical stratosphere could be elucidated with models similar to those presented here.

4 Conclusions

The use of 1D advection–diffusion and advection–dilution (boundary-current) models to interpret tracer observations is ubiquitous in ocean and atmosphere science. As far as we know, these models are always applied using transport coefficients that are constant in time, despite the fact that the real system under study is often clearly variable. Constant coefficients provide a zero-order view of the system, a convenient translation of tracer concentrations to rough transport timescales. However, as we show here, these simple models can be

solved analytically with small-amplitude time-varying coefficients using standard methods. A primary goal of this work has been to present and document such solutions in the hope that they may be of use to researchers who want to go one step beyond the simplest analysis, while still maintaining the convenience and ease of a low-parameter analytical model.

With this goal in mind, we have solved the 1D advection–diffusion and advection–dilution models analytically with time-varying transport coefficients in a semi-infinite domain. A background (zero-order) tracer gradient is established by a linearly increasing concentration boundary condition at the origin. The time variations in transport, assumed uniform in space, act on this background gradient, resulting in fluctuations in tracer concentration. We have analyzed these fluctuations, documenting their amplitude and phase as functions of model parameters. We have also compared these transport-induced tracer fluctuations to fluctuations that can arise from a distinct mechanism: fluctuations in concentration at the boundary, propagated by time-independent transport. We find that, in regions close to the boundary, the propagated fluctuations have larger amplitude, while further downstream the transport-induced fluctuations dominate.

The comparison of propagated- to transport-induced fluctuations is relevant to a key application of the model that we report here: the interpretation of fluctuations in tracer concentrations in the sub-tropical and tropical Atlantic Deep Western Boundary Current (DWBC). Waugh and Hall [21] were able to model well the tracers CFCs, helium–tritium and SF_6 in the North Atlantic using a boundary-current model with constant coefficients. However, they noted that fluctuations in temperature and salinity could not be replicated by the model, which rapidly attenuated any periodic signal specified at the northern outcrop boundary. Similarly, Steinfeldt and Rhein [17] were able to reproduce averaged concentrations and trends in time series of tropical DWBC CFCs using a 1D advection–diffusion model with constant coefficients and a northern-outcrop boundary condition. However, they were not able to reproduce fluctuations in the CFCs.

We have applied the boundary-current model to the tropical DWBC CFC time series of Steinfeldt and Rhein [17]. We find that with the best-estimate parameter values of Waugh and Hall [21], transport-induced fluctuations dominate over propagated fluctuations at these distances from the northern outcrop. The propagated fluctuations are negligible and cannot explain the observed CFC fluctuations. In fact, some of the observed fluctuations are too large even to be explained by the transport fluctuation mechanism. However, other CFC fluctuations can be explained by the mechanism using plausible parameter values.

Acknowledgements We thank Darryn Waugh for useful discussions and comments and Reiner Steinfeldt for providing CFC data that allowed a comparison between observational data and model results. Finally, we are grateful to two anonymous reviewers whose comments lead to a much improved manuscript. This work was supported by NSF Award OCE-06-23366.

Appendix

In what follows we describe in some detail the procedure to find the analytical solutions for Eqns. (5) and (6) for the perturbed advection–diffusion and advection–dilution problems (1). Also, the steady-state case of a periodical boundary condition is treated.

A Solutions for a linear varying BC and fluctuation in transport

In order to solve Eq. (1) analytically, we perform a perturbation of order $O(\epsilon)$; that is, we consider solutions of the form $\mathbf{c} = \mathbf{c^{(0)}} + \epsilon \mathbf{c^{(1)}}$. Eq. (1) is separated into zero-th and first-order components. The zero-th order component is

$$\frac{\partial \mathbf{c^{(0)}}}{\partial t} + \mathbf{L}\left(\mathbf{c^{(0)}}\right) = \mathbf{0}, \tag{11}$$

while the first-order solution satisfies the following PDE:

$$\frac{\partial \mathbf{c^{(1)}}}{\partial t} + \mathbf{L}\left(\mathbf{c^{(1)}}\right) = -f\,(t)\,\mathbf{L}\left(\mathbf{c^{(0)}}\right). \tag{12}$$

The initial condition is $\mathbf{c^{(i)}} = 0$ at $t = 0$ for both $i = 0, 1$. The zero-th order BC at $x = 0$ is $\mathbf{c^{(0)}} = \mathbf{y}\,(\mathbf{t})$ and $\mathbf{c^{(1)}} = 0$. Note that the first-order equation is equivalent to the zero-th order equation, except for the addition of the forcing term $-f\,(t)\,\mathbf{L}\left(\mathbf{c^{(0)}}\right)$, which depends on the zero-th order solution only. Physically, the transport variations act on the zero-th order tracer gradients to force the first-order tracer fluctuations.

Solutions to Eqs. (11) and (12) can be written as

$$\mathbf{c^{(0)}}\,(x,t) = \int_{-\infty}^{t} \mathrm{d}t' \mathbf{y}\left(t'\right) \mathbf{G^{(0)}}\left(x,t;t'\right) \tag{13}$$

and

$$\mathbf{c^{(1)}}\,(x,t) = \int_{-\infty}^{t} \mathrm{d}t' \int_{0}^{\infty} \mathrm{d}x' \mathbf{F}\left(x',t'\right) \mathbf{G^{(1)}}\left(x,t;x',t'\right), \tag{14}$$

where $\mathbf{F}\,(x,t) = -f\,(t)\,\mathbf{L}\left(\mathbf{c^{(0)}}\right) = f\,(t)\,\frac{\partial}{\partial t}\mathbf{c^{(0)}}$ and $\mathbf{G^{(0)}}$ and $\mathbf{G^{(1)}}$ are the Green's functions for the respective PDEs (see Appendix B).

Linear varying boundary condition and periodic fluctuation in transport. We consider a linearly increasing BC, $\mathbf{c^{(0)}}\,(0,t) = \mathbf{y}\,(t) = \gamma t$ and a periodic fluctuation in transport $f\,(t) = \mathrm{Re}\left(e^{i(\omega t)}\right)$. We assume that the BC has been present for sufficiently long that we can neglect transients in the zero-th order solution for all x of interest, and the long-time solution is given by Eq. (5). For the advection–diffusion model the mean transit time is simply the advective time, given by

$$\Gamma = \tau_{\mathrm{adv}} = \frac{x}{u}. \tag{15}$$

For the boundary-current model

$$\begin{cases} \Gamma_b = \tau_{\mathrm{adv}}\left(1 + \frac{1}{\alpha}\right) \\ \Gamma_i = \Gamma_b + \frac{\tau_{\mathrm{mix}}}{\alpha} \end{cases}, \tag{16}$$

(see [21]). In this case, the forcing term in Eq. (14) becomes:

$$\mathbf{F} = f\,(t)\,\frac{\partial \mathbf{c^{(0)}}}{\partial t} = \gamma \mathrm{Re}\left(e^{i(\omega t)}\right). \tag{17}$$

We now turn to the solutions $\mathbf{c}^{(1)}$ in (6) for the 1D advection–diffusion and boundary-current models.

A.1 1 D Advection–diffusion model

The solution $\mathbf{c}^{(1)}$ is

$$c^{(1)}(x,t) = \mathrm{Re}\left\{ \int_0^\infty \mathrm{d}x' \int_{-\infty}^t \mathrm{d}t' \gamma e^{i\omega(t-t')} G^{(1)}\left(x,t;x',t'\right) \right\} = \gamma \mathrm{Re}\left\{ e^{i\omega t} |I| e^{-i\phi} \right\}. \tag{18}$$

with $G^{(1)}$ given by Eq. (38) and

$$I = \int_0^\infty \mathrm{d}x' \int_{-\infty}^t \mathrm{d}t' e^{-i\omega t'} G^{(1)}\left(x,t;x',t'\right) = I_\mathrm{R} - i I_\mathrm{C} = |I| e^{-i\phi}, \tag{19}$$

where $|I| = \sqrt{I_\mathrm{R}^2 + I_\mathrm{C}^2}$, and $\tan\phi = \frac{I_\mathrm{C}}{I_\mathrm{R}}$. In order to solve the above integral, the "trick" is to recast and regroup the exponential arguments in Eq. (38) for both $x < x'$ and $x > x'$, so that the exponent in time has the form $-\left(\frac{u^2}{4k} + i\omega\right) t'$. We then recognize that the resulting integrand is a Laplace transform with complex Laplace variable $s = \left(\frac{u^2}{4k} + i\omega\right)$ and find the transform from standard tables. Finally (18) can expressed as

$$c^{(1)}(x,t) = A^{(1)}(x) \cos\left(\omega\left(t - \tau^{(1)}(x)\right)\right), \tag{20}$$

where

$$A^{(1)}(x) = \gamma |I| \tag{21}$$

and

$$\tau^{(1)}(x) = \frac{\phi}{\omega}. \tag{22}$$

In the above expressions $I_\mathrm{R} = \left[E_-(x) - E_+(x)\right] - \left[E_-(0) - E_+(0)\right]$, $I_\mathrm{C} = \left[O_-(x) - O_+(x)\right] - \left[O_-(0) - O_+(0)\right]$, with

$$\left\{ \begin{array}{c} E_{-,+}(x) \\ O_{-,+}(x) \end{array} \right\} = \frac{1}{2k\sqrt{|\frac{s}{k}|}|Z_{-,+}|} e^{a_{-}x} \left\{ \begin{array}{c} \cos\left(bx + \frac{\theta}{2} + \alpha_{-,+}\right) \\ \sin\left(bx + \frac{\theta}{2} + \alpha_{-,+}\right) \end{array} \right\}. \tag{23}$$

Here, $a_{-,+} = \frac{u}{2k} \mp \sqrt{|\frac{s}{k}|}\cos\frac{\theta}{2}$, $b = \sqrt{|\frac{s}{k}|}\sin\frac{\theta}{2}$, and $\alpha_{-,+} = \arctan\left(\mp\frac{b}{a_{-,+}}\right)$, with $Z_{-,+} = \frac{u}{2k} \mp \sqrt{\frac{s}{k}}$, where: $\sqrt{\frac{s}{k}} = \sqrt{|\frac{s}{k}|} e^{i\frac{\theta}{2}}$, $|\frac{s}{k}| = \sqrt{\left(\frac{u}{2k}\right)^4 + \left(\frac{\omega}{k}\right)^2}$, and $\theta = \arctan\left(\frac{4k\omega}{u^2}\right)$.

A.2 Boundary-current model

The expressions for the components of $\mathbf{c}^{(1)}$ become:

$$c_b^{(1)} = \gamma \mathrm{Re}\left(\int_0^\infty \mathrm{d}x' \int_{-\infty}^t \mathrm{d}t' e^{i\omega t'} (G_{bb}^{(1)} + G_{bi}^{(1)}) \right) \tag{24}$$

$$c_i^{(1)} = \gamma \mathrm{Re}\left(\int_0^\infty \mathrm{d}x' \int_{-\infty}^t \mathrm{d}t' e^{i\omega t'} (G_{bi}^{(1)} + G_{ii}^{(1)}) \right). \tag{25}$$

The time integrals can be evaluated by changing the time variable to $t'' = t - t'$, grouping terms so that the exponent has the form $-(\alpha\sigma + i\omega)t''$, recognizing the resulting integrand as

a Laplace transform with complex Laplace variable $s = \alpha\sigma + i\omega$ and finding the transform in standard tables. (For transforming the term in I_1 we exploit $\mathrm{I}_1(y) = \frac{\mathrm{d}}{\mathrm{d}y}\mathrm{I}_0(y)$ and the rule for Laplace transforms of derivatives.) The remaining x integrand is comprised solely of simple exponentials. The result for the boundary-current region is

$$c_b^{(1)}(x,t) = A_b^{(1)}(x)\cos\left(\omega(t-\tau_b^{(1)}(x))\right), \tag{26}$$

where the amplitude

$$A_b^{(1)} = \frac{\gamma}{\omega}\left(1 + e^{-2\sigma\beta_- x/u} - 2e^{-\sigma\beta_- x/u}\cos\left(\omega\beta_+ x/u\right)\right)^{1/2}, \tag{27}$$

the phase-lag time

$$\tau_b^{(1)} = -\frac{1}{\omega}\arctan\left(\frac{C_2(1-e^{-\sigma\beta_- x/u}\cos(\omega\beta_+ x/u)) + C_1 e^{-\sigma\beta_- x/u}\sin(\omega\beta_+ x/u)}{C_1(1-e^{-\sigma\beta_- x/u}\cos(\omega\beta_+ x/u)) - C_2 e^{-\sigma\beta_- x/u}\sin(\omega\beta_+ x/u)}\right), \tag{28}$$

with

$$C_1 = \sigma^2\omega^2 - (\omega^2 + \alpha^2\sigma^2 + \alpha\sigma^2)^2 \tag{29}$$

$$C_2 = 2\sigma\omega(\omega^2 + \alpha^2\sigma^2 + \alpha\sigma^2), \tag{30}$$

and

$$\beta_\pm = 1 \pm \frac{\alpha^2\sigma^2}{\omega^2 + \alpha^2\sigma^2}. \tag{31}$$

The interior region solution $c_i^{(1)}$ has the following simple phase and amplitude relation to $c_b^{(1)}$:

$$A_i^{(1)} = \frac{\alpha\sigma}{\sqrt{\omega^2 + \alpha^2\sigma^2}}\, A_b^{(1)} \tag{32}$$

and

$$\tau_i^{(1)} = \tau_b^{(1)} - \frac{1}{\omega}\arctan\left(\frac{\omega}{\alpha\sigma}\right). \tag{33}$$

B Green's functions

The Green's functions that appear in Eqs. (13) and (14) are the solutions to

$$\frac{\partial \mathbf{G^{(k)}}}{\partial t} + \mathbf{L}\left(\mathbf{G^{(k)}}\right) = \mathbf{S^{(k)}}, \tag{34}$$

with index $k = 0, 1$. For the zero-th order solution, $\mathbf{S^{(0)}} = 0$, $\mathbf{G^{(0)}}(x,0) = 0$, and $\mathbf{G^{(0)}}(0,t) = \boldsymbol{\delta}(t)$; For the first-order solution, $\mathbf{S^{(1)}} = \mathbf{I}\cdot\delta\left(t-t'\right)\delta\left(x-x'\right)$, where $\mathbf{I}$ is the identity matrix; $\mathbf{G^{(1)}}$ satisfies both homogeneous boundary and initial conditions.

For the advection–diffusion model the solution has only one component, $\mathbf{G^{(k)}} = \mathbf{G^{(k)}}\left(\mathbf{x},\mathbf{t}|\mathbf{x}',\mathbf{t}'\right)$. For the boundary-current model we have

$$\mathbf{G^{(0)}} = \begin{pmatrix} G_b^{(0)}\left(x,t|x',t'\right) \\ G_i^{(0)}\left(x,t|x',t'\right) \end{pmatrix} \tag{35}$$

and

$$\mathbf{G^{(1)}} = \begin{pmatrix} G_{bb}^{(1)}\left(x,t|x',t'\right) & G_{bi}^{(1)}\left(x,t|x',t'\right) \\ G_{ib}^{(1)}\left(x,t|x',t'\right) & G_{ii}^{(1)}\left(x,t|x',t'\right). \end{pmatrix}. \tag{36}$$

For instance, the matrix element $G_{bi}^{(1)}\left(x,t|x',t'\right)$ is the response at time t and position x in the boundary-current core region to a point source at a previous time t' and position x' in the surrounding interior region.

B.1 1 D Advection–diffusion Model

B.1.1 Zero-th order solution

$$G^{(0)}\left(x,t|t'\right) = \frac{x}{\sqrt{4\pi k\left(t-t'\right)^3}} e^{-\frac{\left(x-u\left(t-t'\right)\right)^2}{4kt}} \tag{37}$$

B.1.2 First-order solution

$$G^{(1)}\left(x,t|x',t'\right) = \frac{1}{\sqrt{4k\pi\left(t-t'\right)}} e^{-\frac{u\left(x'-x\right)}{2k}} e^{-\frac{u^2\left(t-t'\right)}{4k}} \begin{cases} e^{-\frac{\left(x'-x\right)^2}{4k\left(t-t'\right)}} - e^{-\frac{\left(x'+x\right)^2}{4k\left(t-t'\right)}} & x < x' \\ e^{-\frac{\left(x-x'\right)^2}{4k\left(t-t'\right)}} - e^{-\frac{\left(x+x'\right)^2}{4k\left(t-t'\right)}} & x > x' \end{cases} \tag{38}$$

B.2 Boundary-current model

B.2.1 Zero-th order solution

Waugh and Hall [21] found that:

$$G_b^{(0)}\left(x,t\right) = \widehat{G_1}\delta\left(t-\tau_{\text{adv}}\right) + \widehat{G_2}\Theta\left(t-\tau_{\text{adv}}\right) \tag{39}$$

and

$$G_i^{(0)}\left(x,t\right) = \widehat{G_3}\Theta\left(t-\tau_{\text{adv}}\right) \tag{40}$$

where

$$\widehat{G_1} = \exp^{-\frac{1}{Pe}}, \qquad \widehat{G_2} = \frac{\alpha}{\zeta\tau_{\text{mix}}}\exp -\frac{1+\zeta^2}{Pe} I_1\left(\frac{2\zeta}{Pe}\right),$$
$$\widehat{G_3} = \frac{\alpha}{\tau_{\text{mix}}}\exp -\frac{1+\zeta^2}{Pe} I_0\left(\frac{2\zeta}{Pe}\right)$$

$\tau_{\text{adv}} = \frac{x}{u}$, $Pe = \frac{\tau_{\text{mix}}}{\tau_{\text{adv}}}$, $\zeta^2 = \alpha\left(\hat{t}-1\right)$, $\hat{t} = \frac{t}{\tau_{\text{adv}}}$, I_0 and I_1 are modified Bessel functions of zero-th and first-order.

B.2.2 First-order solution

$$G_{bb}^{(1)} = \frac{1}{u}\sqrt{\frac{\alpha\sigma^2(x-x')}{uT}} e^{-\frac{\sigma}{u}(x-x')}\, e^{-\alpha\sigma T}\mathrm{I}_1\left(2\sqrt{\frac{1}{u}\alpha\sigma^2(x-x')T}\right)\Theta(T)$$
$$+ \frac{1}{u}e^{-\frac{\sigma}{u}(x-x')}\delta(T), \tag{41}$$

$$G_{ib}^{(1)} = \frac{\alpha\sigma}{u} e^{-\frac{\sigma}{u}(x-x')} e^{-\alpha\sigma T} \mathrm{I}_0\left(2\sqrt{\frac{1}{u}\alpha\sigma^2(x-x')T}\right)\Theta(T), \tag{42}$$

$$G_{ii}^{(1)} = \sqrt{\frac{\alpha\sigma^2 T}{u(x-x')}} e^{-\sigma(x-x')T/u} e^{-\alpha\sigma T} \mathrm{I}_1\left(2\sqrt{\frac{1}{u}\alpha\sigma^2(x-x')T}\right)\Theta(T), \tag{43}$$

and

$$G_{bi}^{(1)} = \frac{1}{\alpha} G_{bi}^{(1)}, \tag{44}$$

where $T = t - t' - \frac{1}{u}(x - x')$ and I_n is the modified Bessel function of order n. Because the zero-th order transport is in steady-state, the solutions depend only on the difference $t - t'$.

C Solutions for a periodic tracer BC

In this section we give the solutions to the steady-state versions of the two models where periodic BCs apply. The solutions can be written as $c\,(x, t) = P\,(x) \cos\left(\omega\,(t - \phi\,(x))\right)$. For the advection–diffusion model

$$P\,(x) = P_0\sqrt{\mathcal{P}_R^2 + P_C^2}, \tag{45}$$

$$\phi\,(x) = \frac{1}{\omega} \cdot \frac{P_{\mathrm{C}}}{P_{\mathrm{R}}} \tag{46}$$

$P_{\mathrm{R}} = e^{Z_{\mathrm{R}}x} \cos\left(Z_{\mathrm{C}}x\right)$ and $P_{\mathrm{C}} = e^{Z_{\mathrm{R}}x} \sin\left(Z_{\mathrm{C}}x\right)$, where $Z_{\mathrm{R}} \equiv a_-$ and $Z_{\mathrm{C}} \equiv b$, with a_- and b given in (A.1). For the boundary-current model the solution in the core region is given by

$$P(x) = P_0 \exp\left(-\frac{x\sigma\omega^2}{u(\omega^2 + \alpha^2\sigma^2)}\right) \tag{47}$$

$$\phi\,(x) = \frac{x}{u} \cdot \frac{\omega^2 + \alpha\,(\alpha + 1))\sigma^2}{\omega^2 + \alpha^2\sigma^2} \tag{48}$$

(see [21]).

References

1. Andrie C, Ternon JF, Messias MJ, Memery L, Bourles B (1998) Chlorofluoromethane distributions in the deep equatorial atlantic during january-march 1993 - water masses and transports near the equator. Deep Sea Res I 45:903–930(28)
2. Curry RG, McCartney MS, Joyce TM (1998) Oceanic transport of subpolar climate signals to mid-depth subtropical waters. Nature 391:575–577
3. Doney SS, Jenkins WJ (1994) Ventilation of the deep western boundary current and abyssal western North Atlantic: estimates from tritium and ^{3}He distributions. J Phys Oceanogr 24:638–659
4. Freudenthal S, Andrié C (2002) The arrival of a "new" Labrador Sea Water signal in the tropical Atlantic in 1996. Geophys Res Lett 29(15): 1741. doi:10.1029/2002GL015062
5. Haine TW, Watson AJ, Liddicoat MI, Dickson RR (1998) The flow of Antarctic bottom water to the southwest Indian Ocean estimated using CFCs. J Geophys Res 103(C12):27637–27654
6. Hall TM, Haine TWN, Waugh DW (2002) Inferring the concentration of anthropogenic carbon in the ocean from tracers. Glob. Biogeochem. Cycles 16(4):1131. doi:10.1029/2001GB001835
7. Hall TM, Waugh DW (2000) Stratospheric residence time and its relationship to mean age. J Geophys Res 105:6773–6782

8. Johnson GC, Bullister JL, Gruber N (2005) Labrador Sea Water property variations in the northeastern Atlantic Ocean. Geophys Res Lett 32:L07602
9. McKenna DS (1997) Analytic solutions of reaction diffusion equations and implications for the concept of an air parcel. J Geophys Res 102(D):13719–13725
10. Molinari RL, Fine RA, Wilson WD, Curry RG, Abell J, McCartney MS (1998) The arrival of recently formed Labrador Sea water in the deep western boundary current at 26.5°N. Geophys Res Lett 25:2249–2252
11. Mote PW, Dunkerton TJ, McIntyre ME, Ray EA, Haynes PH, J MRIII (1998) Vertical velocity, vertical diffusion, and dilution by midlatitude lower stratosphere. J Geophys Res 103(D8):8651–8666
12. Mouchet A, Deleersnijder E (2008) The leaky funnel model, a metaphor of the ventilation of the World Ocean as simulated in an OGCM. Tellus A 60:761–774
13. Neu J, Plumb AR (1999) Age of air in a "leaky pipe" model of stratospheric transport. J Geophys Res 104:19243–19255
14. Pickart R, Hogg N, Smethie WM (1989) Determining the strength of the deep western boundary current using the chlorofluoromethane ratio. J Phys Oceanogr 19:940–951
15. Plumb AR (1996) A tropical pipe model of stratospheric transport. J Geophys Res 101:3957–3972
16. Rhein M (1994) The deep western boundary current—tracers and velocities. Deep Sea Res 41:263–281
17. Steinfeldt R, Rhein M (2004) Spreading velocities and dilution of North Atlantic deep water in the tropical Atlantic based on CFC time series. J Geophys Res 109:C03046.
18. Sunnerup RE (2001) On the relations among CFC derived water mass ages. Geophys Res Lett 28(9):1739–1742
19. Terenzi F, Hall TM, Khatiwala S, Rodehacke CB, LeBel DA (2007) Uptake of natural and anthropogenic carbon by the Labrador Sea. Geophys Res Lett 34:L06608
20. Waugh DW, Hall TM (2002) Age of stratospheric air; theory, observations, and models. Rev Geophys 40(4):1010
21. Waugh DW, Hall TM (2005) Propagation of tracer signals in boundary currents J Phys Oceanogr 35: 1538–1552. doi:10.1175/JPO2779.1
22. Weiss RF, Bullister JL, Gammon RH, Warner MJ (1985) Atmospheric chlorofluoromethanes in the deep equatorial Atlantic. Nature 314:608–610

Environ Fluid Mech (2010) 10:257–273
DOI 10.1007/s10652-009-9159-y

ORIGINAL ARTICLE

Helium isotopic constraints on simulated ocean circulations: implications for abyssal theories

Jean-Claude Dutay · Julien Emile-Geay · Daniele Iudicone · Philippe Jean-Baptiste · Gurvan Madec · Claire Carouge

Received: 13 November 2008 / Accepted: 18 November 2009 / Published online: 18 December 2009

Abstract There is ongoing controversy as to the dynamical significance of geothermal heat flow in shaping the abyssal circulation. In this paper, we gauge the impact of geothermal heating and vertical mixing parameterizations in the general circulation model OPA. The experiments are evaluated by comparing simulated mantle ^{3}He with observations collected during the GEOSECS and WOCE programs. This tracer is particularly adapted to the validation of our numerical simulations because its injection into the ocean interior is tightly linked to geothermal processes. In agreement with previous studies, the model circulation is found very sensitive to the parameterization of the vertical mixing. The meridional overturning circulation (MOC) is globally intensified when moving from a constant mixing to a version with enhanced mixing near the ocean bottom, with the most drastic variation observed for AABW (+50%). Adding the geothermal heat flux mainly affects AABW circulation in the model, enhancing it all the more as the meridional circulation is slow (low vertical mixing), but proportionally less so when it is more vigorous (enhanced vertical mixing). This can be understood from the requirement of the abyssal ocean to maintain heat balance. The evaluation with mantle ^{3}He simulations reveals that the version with low vertical mixing, with its sluggish circulation, produces unrealistically high a ^{3}He isotopic composition. However, with a vertical mixing that is enhanced at depth, the ^{3}He distribution falls within an acceptable range of values in the deep ocean. Finally, adding the geothermal heating to this enhanced mixing case provides a substantial improvement of the simulation of AABW in all basins but

J.-C. Dutay (✉) · P. Jean-Baptiste · C. Carouge
LSCE, IPSL/CEA, UVSQ, CNRS, Gif sur Yvette, France
e-mail: jean-claude.dutay@lsce.ipsl.fr

J. Emile-Geay
Department of Earth Sciences, University of Southern California, Los Angeles, CA, USA
e-mail: julieneg@usc.edu

D. Iudicone · G. Madec
IPSL/LOCEAN, Paris, France

D. Iudicone
Stazione Zoologica "Anton Dohrn", Naples, Italy

the Indian Ocean. ^{3}He isotopic composition is then in good agreement with the observations. Taken jointly with observational estimates of the MOC intensity, these independent isotopic constraints suggest that both geothermal heating and enhanced diapycnal mixing at depth are key ingredients in the realistic simulation of abyssal circulation.

Keywords Ocean · Tracer · Helium-3 · Geothermal flux · Model

1 Introduction

The ocean circulation plays an important role in climate variability, e.g. [40,53], with the ocean having a great capacity to store and transport heat and CO_2, as well as exchanging these with the atmosphere. Thus it contributes significantly to the poleward transfer of excess solar energy received in the low latitudes to higher latitudes. In addition, it governs atmospheric CO_2 concentrations, since it contains about 50 times more CO_2 than the atmosphere, with the majority in the deep ocean [6]. Hence, understanding the dynamics of the deep-ocean is essential to the study of global climate change, via the carbon cycle.

Ocean general circulation models (OGCMs) provide the only controllable environment with which to answer such questions. It is thus crucial to understand the processes and forcings that control the strength and patterns of the simulated deep-ocean circulation. Many previous studies have shown the importance of the parameterization of small-scale mixing processes on ocean circulation [9,10,16,35,58]. Indeed, the models' deep-ocean circulation are known to be particularly sensitive to the vertical mixing parameterization [7,54]. Since the deep ocean is essentially aphotic, diapycnal mixing is chiefly responsible for the transfer of water mass characteristics from top to bottom away from convective regions (i.e. over much of the global ocean), thereby mapping surface density gradients onto the deep ocean. These gradients, in turn, support the abyssal circulation via geostrophic balance.

There is, however, another process supplying heat to the abyss. Although it has long been neglected by oceanographers, e.g. [26], geothermal heating has recently been recognized as a significant actor in abyssal dynamics and thermodynamics [1,2,13,45,52].

While its magnitude is small compared to surface heat fluxes, geothermal heating acts on a much broader area—the entire seafloor. Overall, it contributes a comparable amount of buoyancy to bottom waters as air/sea fluxes—albeit of opposite sign over the Southern Ocean. Furthermore, it acts below the thermocline, in regions of the ocean where small thermal gradients mean that the downward diffusive heat flux is also small. So small, in fact, as to be commensurate to geothermal heat flow over much of the seafloor's surface [13]. As a result, the same authors found a realistic representation of geothermal heating was able to generate a deep circulation of the order of several Sverdrups[1] for Antarctic Bottom Water (AABW), and suggested that it constitutes a driving force of the AABW cell. This statement must be tempered by the observation than the deep circulation response to geothermal forcing is much reduced when diapycnal mixing is very strong—as appears to be the case in the vicinity of rough topography. Nonetheless, the convergence of idealized [1,45,52] and more realistic [13] studies of its impact justifies the claim that geothermal heat flow significantly shapes the abyssal circulation. Can these model-based theories be substantiated by observational constraints?

In this paper we use ^{3}He tracer data to evaluate the suite of modeling experiments. The model circulation is used to simulate natural ^{3}He distributions and compare them to obser-

[1] 1 Sverdrup (Sv) $= 10^6\ \mathrm{m^3\ s^{-1}}$.

vations. This inert tracer, whose oceanic behavior is tightly linked to geothermal heat flow [25], is particularly apt to probing the effect of geothermal heat flow on the model's deep circulation. Indeed, its source consists of mantle ^{3}He that is injected into the oceans by the hydrothermal circulation at deep sea spreading ridges, where most of the geothermal heat flux is also transferred to the ocean [8,31,44]. This injection of hot hydrothermal fluids enriched in ^{3}He produces large scale ^{3}He plumes (with ^{3}He excesses up to 30–40% relative to the solubility equilibrium with the atmosphere), that are dispersed by the prevailing currents and can be used for tracing deep ocean currents and for evaluating ocean general circulation model [12,17]. The comparison from six OGCM during the Ocean Carbon Model Intercomparison Project (OCMIP2) has shown that the modeled ^{3}He distribution is very sensitive to the simulated circulation and that the parameterization proposed by Farley et al. [17] for its source function is robust [12]. During OCMIP-2, six different global OGCMs were tested with mantle ^{3}He simulations [12]. This inter-comparison has revealed that the resulting circulation was a direct consequence of the models' characteristics. More precisely, the OGCMs coupled to a sea-ice model were found to produce a substantial amount of bottom water in the Southern Ocean, which tended to overestimate AABW ventilation and to produce too low a ^{3}He value in this region. Conversely, models that were not coupled with a sea-ice model were systematically underestimating AABW ventilation and simulating excessive ^{3}He values. Moreover, the parameterization of sub-grid scale mixing (which is recognized to generate large effect on the simulated circulation and tracer redistribution, e.g. [14]) also accounted for significant discrepancies. The two OCMIP2 models coupled to a sea-ice model were not using the eddy-induced velocity parameterization of Gent et al. [20]. Sensitivity tests performed during that project clearly demonstrated that this parameterization tended to reduce ventilation of the deep ocean, and to generate higher ^{3}He excesses. The wide range of simulated distributions during OCMIP2 has been a motivation to explore in more details the impact of the physics on the modeled deep ocean circulation and to further evaluate it with mantle ^{3}He simulations. Taking into account the experience gained during the OCMIP2 exercise, we have initiated our new modeling efforts with a more recent configuration of the OPA model with characteristics that provide a more realistic circulation. This analysis focuses on the impact of the geothermal heating and the parameterization of the vertical mixing on the circulation simulated by the model.

2 Description of the model and the simulations

The ice-ocean coupled model is OPA, developed at IPSL/LOCEAN [33], in its global configuration ORCA2-LIM. The horizontal mesh is based on a 2° by 1.5° Mercator grid. It has been modified poleward of 20°N in order to include two numerical inland poles and meridional resolution is refined up to 0.5° at the equator [32]. There are 31 vertical levels, with the highest resolution (10 m) in the upper 150 m. The upper boundary uses a free surface formulation [42]. The model is coupled to the dynamical-thermodynamical sea ice model LIM [21]. Lateral mixing is performed along isopycnal surfaces, and the eddy-induced velocity parameterization of Gent and McWilliams (1990) is added. The vertical mixing scheme uses a turbulent closure [4], and a diffusive bottom boundary layer parameterization [3] is included.

Two versions of the model background vertical diffusivity are tested in our simulations. A first version (CONTROL) uses a constant values of $0.12 \times 10^{-4}\,\mathrm{m^2\,s^{-1}}$ throughout the entire ocean volume, and a second version (High K_z) where it increases linearly from the surface to the bottom in order to mimic the effects of decreased stratification and increased

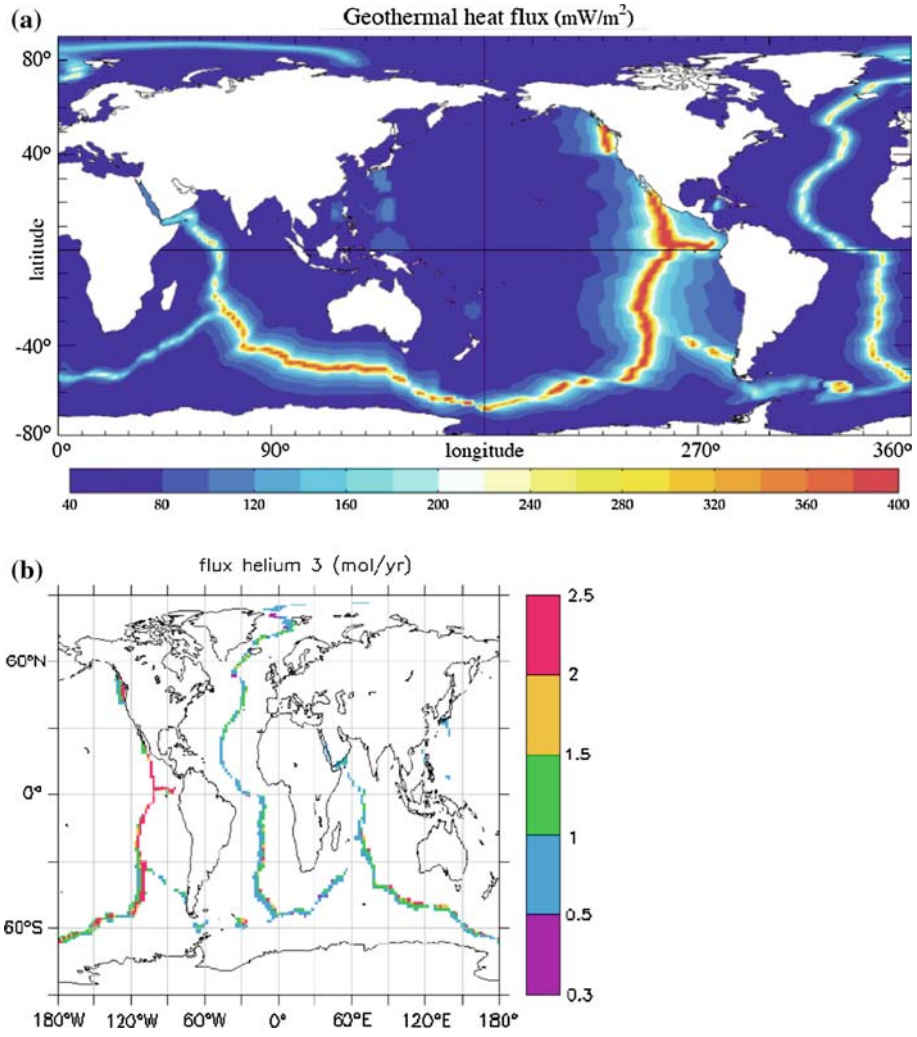


Fig. 1 **a** Geothermal heat flux (units: mW m^{-2}), **b** mantle helium-3 flux (units: mol year^{-1})

small-scale turbulence near the bottom (values ranges from 0.12×10^{-4} m^2 s^{-1} in the first 1000 m to 1.2×10^{-4} m^2 s^{-1} at depth).

At the surface, the model is forced by fluxes of heat and freshwater prescribed with bulk formulae using monthly climatologies. A penetrative shortwave solar radiation formulation is used. A restoring to climatological surface salinities was also added, that is only significant in the North Atlantic and Greenland Sea. Climatological ERS1/2 scatterometer monthly mean wind stresses were used for the tropics while the NCEP/NCAR climatology was used poleward of 50°N and 50°S.

At the bottom, we include in some experiments the geothermal heating term (Q_{geo}) developed by Emile-Geay and Madec [13], who parameterized the heat source as a function of the age of the bedrock. The global mean value is 86.4 mW m^{-2}, close to observational estimates [48]. Its spatial distribution is shown in Fig. 1a, which closely mirrors seafloor age [34]. The

Table 1 Mean δ^3He value (units: %)

	Global	Atlantic	Pacific	Indian	Southern Ocean (south of 45°S)
Exp 1: Control	8.5	5.6	15.6	11.7	9.8
Exp 2: Q_{geo}	7.8 (8)	4.7 (16)	14.5 (7)	10.8 (8)	8.6 (12)
Exp 3: High K_z	6.2 (27)	3.5 (38)	11.9 (23)	8.8 (25)	6.4 (35)
Exp 4: Q_{geo} + High K_z	5.7 (32)	3.0 (46)	11.0 (29)	8.0 (32)	5.5 (44)

Values in brackets correspond to the percentage of variation with the control experiment

highest values are observed along the ridge system where the lithosphere formed at high temperature, then was displaced outward by lithospheric plate divergence. Along ridge axes, geothermal heat flux is directly related to the spreading rates: maxima are located in the eastern Pacific Ocean, where the spreading rate is highest; lowest values are observed in the Atlantic, with intermediate values in the Indian and Southern Oceans.

We have performed four simulations in order to investigate the sensitivity of the ocean circulation to both vertical mixing parameterization and geothermal heating (Table 1). Our first experiment (CONTROL) uses a constant vertical mixing coefficient and the geothermal forcing is not included. In the second experiment (Q_{geo}), the geothermal heating is added. The third experiment (High K_z) does not include geothermal heating, but has a vertical mixing parameterization with values increasing with depth (see Sect. 2). The fourth experiment includes both geothermal heating and the vertical mixing coefficient that varies with depth ($\exp 4 = \text{High}\, K_z - Q_{geo}$). This last experiment has been extensively evaluated and analyzed in detail in Iudicone et al. [22,23].

3 Global meridional stream functions

We first analyze the global meridional stream functions of the different experiments. In the control experiment, the circulation associated with the North Atlantic deep water overturning (NADW) (between 1000 and 2500 m depth) reaches 15 Sv, and below 3000 m the circulation related to the AABW is on the order of 5 Sv (Fig. 2). These diagnostics are not easily constrained by the observations, as diagnosed stream functions or transports usually carry large uncertainties (3–5 Sv, [50]). However, the mean strength of the NADW in the Atlantic Ocean overturning is comparable to observational estimates, 18 ± 5 Sv [50] and 16 ± 3 Sv [18]. In contrast, the AABW circulation seems weak: using geochemical tracers budgets Orsi et al. [37] estimated the ventilation of AABW on shelves of Antarctica to be 8–10 Sv, and global AABW overturning is estimated from hydrographic data to be 22 Sv [50]. Adding the influence of geothermal heating (exp 2), the main adjustment is an increase of 50% in the strength of the AABW cell (Fig. 2). Conversely, a quasi-insignificant change is observed for the NADW cell. These variations are necessary to compensate for the additional heat supplied to the abyssal ocean [45]. It affects predominantly AABW because the geothermal heating is injected at the ocean floor and only weakly affects the shallower NADW. In addition, the magnitude of geothermal heating is smallest in the Atlantic Ocean. In the third experiment, when the vertical mixing increases with depth, the thermohaline circulation is enhanced. In comparison to the control run (exp 1), the strength of the AABW cell is doubled (to 10 Sv), and the NADW cell increases to 17.5 Sv (+17%). This large increase in the thermohaline circulation is due to the more efficient vertical transfer of the characteristics of the water

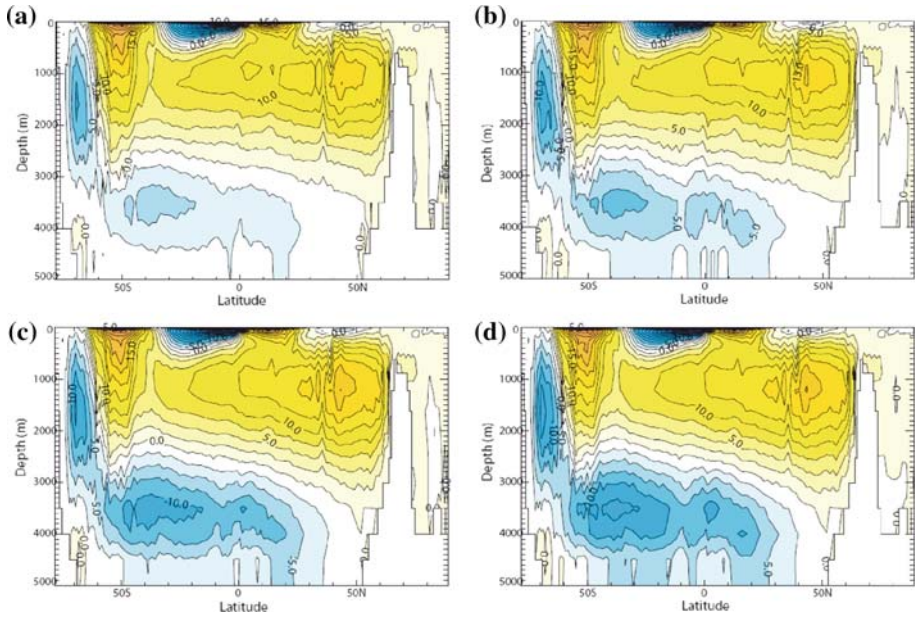


Fig. 2 Global meridional stream functions (units: Sv; 1 Sv $= 10^6$ m^3 s^{-1}). **a** Exp 1, **b** exp 2, **c** exp 3, **d** exp 4

masses produced in the surface layer, which magnifies density gradients in the deep ocean and produces a more intense circulation via geostrophy.

Including the geothermal heating in the version with high vertical mixing leads to almost no change in NADW, but an additional increase of 1.8 Sv (+17.5%) for the AABW cell. The variation obtained by including the geothermal forcing is smaller than for the control run (exp 2 versus exp 1) because the high diapycnal mixing produces an abyssal circulation that is already more vigorous and hence more efficient at transporting heat. This causes only a modest increase in circulation to compensate for the geothermal heat injected at depth and keep the deep ocean in thermal equilibrium [13].

In the model, about 5 Sv of extremely dense Antarctic shelf water converges into the AABW class [22] and is subsequently transformed by the model bottom boundary layer parameterization, which mimics the dense gravity currents descending the slopes of Antarctica to produce AABW. This process is not well simulated because of the excessive lateral diffusion which dilutes the water mass before it reaches the ocean bottom, as occurs in the majority of coarse resolution OGCMs. The result is an AABW strength at 30°S of about 10 Sv, weaker than most observational estimates, e.g., [46,50]. The export of bottom waters toward the Pacific Basin is similar to the estimate in Talley et al. [50]. The CDW import is 50% weaker than observed and so too is AABW export, though observational estimates can vary by as much as 50 Sv [22,23]. The overturning of deep and bottom water in the Indian basin (about 3 Sv) is largely underestimated compared to estimates from inverse models, which range from 7 [19] to 24 Sv [43] (see [50] for a discussion). The model's deep circulation in the Atlantic basin is the closest to observations: NADW/CDW at 30°S is 17.1 Sv in the model as compared to 17.8 Sv in Talley et al. [50], AABW is 2.8 Sv in the model relative to 3.8 Sv in Talley et al. [50] and 6 Sv in Sloyan and Rintoul [46].

Recent studies on the dynamics of the AABW transport have proposed that AABW in the Atlantic Ocean is sensitive to the surface salinity along the Antarctic coast [39] and mixing by lateral diffusion [27]. In this context the model intercomparison study by Weber et al. [56] on the MOC during glacial periods gives interesting indications on the dynamics of the bottom cell. They found that the bottom overturning scales well with both the deep meridional density contrast and—as proposed by Kamenkovich and Goodman [27]—the vertical extent of the deep cell.

From the analysis of our global simulations a complex picture of the sensitivity of AABW to deep mixing emerged, whose main features are summarized here (a more detailed discussion will be presented in a separate manuscript (Iudicone et al., in preparation)).

The most relevant results are related to the Southern Ocean overturning of NADW, which is extremely high and weakly dependent on the deep mixing intensity (in both cases about 80% of NADW is transformed into other water masses). The difference lies in the partitioning of NADW into bottom (AABW) and subsurface waters (SAMW, AAIW). A lower (higher) deep mixing generates a larger (lower) upwelling and a lower (higher) downwelling, which in turn corresponds to a shallower (deeper) separation between the upper and bottom overturning cells.

Furthermore, in the case of low deep diffusivity, most of the weakening of the AABW cell is associated with a reduced deep upwelling in the Pacific basin while the inflow into the Indian Ocean is less affected. In the Atlantic Ocean, a reduced northward penetration of AABW is observed together with a deepening of NADW, in good agreement with the theory of Komenkovich and Goodman [27]. Interestingly, the flow of AABW through the Drake Passage is also strongly affected, with a transport that largely exceeds observational estimates in the case of low diffusivity. Therefore an intense vertical mixing is necessary to reproduce the vertical water mass distribution in the Southern Ocean, as suggested also by observations, e.g. [36].

In principle, the effect of deep mixing on AABW is not limited to a direct influence on the deep and bottom stratification. First, the Southern Ocean upwelling maps the deep anomalies onto the sub-surface water masses. Therefore, deep mixing can also influence the properties of AABW by altering the water mass properties at the surface and thus it can in principle feedback onto AABW formation. This can be particularly true for salinity, which dominates the density variations at low temperatures. However, the role of this latter process could not be fully evaluated in our simulations. Indeed, in our ocean-forced simulations, even in presence of a sea-ice model, the subsurface emergence of deep salinity anomalies in the Southern Ocean is somewhat damped by partially restorating to climatological values.

The other indirect impact of deep diffusivity on AABW formation is related to the overflow mechanism. After surface AABW formation, these very cold waters cascade along the continental slopes of Antarctica. In this process, shear and thermobaric processes increase the volume of dense water by promoting mixing with the incoming NADW, whose thermohaline properties depend on the same diffusivity.

In summary, deep diffusivity impacts the AABW distribution in the ocean by altering both the water mass formation and consumption rates (see discussion in [23]). For the reasons discussed above, only fully coupled ocean-atmosphere-sea-ice models allow for a thorough exploration of the sensitivity of AABW to diapycnal mixing. In particular, the impact of regional mixing variability will be left for future study.

These results illustrate the sensitivity of the model to the parameterization of vertical mixing and geothermal forcing. We will now evaluate the impact of these changes with natural ^{3}He simulations.

4 Natural ^{3}He simulations

4.1 Parameterization of mantle helium injection

The four simulations used the same parameterization for mantle helium injection. The protocol is very similar to that proposed by Farley et al. [17], and further validated using simulations from six OGCM during OCMIP-2 [12]. We simulate both ^{3}He and ^{4}He distributions, and mantle ^{3}He is injected along ocean ridges with an isotopic ratio (δ^3He[2]) of 700%. It is injected at a global rate of 1000 mol ^{3}He/year at 300 m above the ridge axis. δ^3He is next calculated from the outputs of the simulation (^{3}He and ^{4}He concentrations) using its classical definition. The only difference with the OCMIP-2 protocol (http://www.ipsl.jussieu.fr/OCMIP/) is that the mantle helium flux along the ridge axis is proportional to the geothermal flux, instead to be proportional to the ridge spreading rate. This adaptation, formulated for better consistency with the dynamical forcing of our model, induces no significant changes for the spatial distribution of the mantle helium source in the ocean. The depth of injection varies between 2000 and 3000 m. Maximum fluxes occur in the Pacific Ocean at the East Pacific Rise (EPR), intermediate fluxes are found in the Indian and Southern Oceans, while minimum fluxes are associated with the Mid-Atlantic Ridge (MAR) (Fig. 1b). The only sink for mantle helium is loss to the atmosphere, which is prescribed using the gas exchange formulation of Wanninkhof [55] (http://www.ipsl.jussieu.fr/OCMIP/). Simulations are integrated off-line until they reached a quasi-steady state when the globally integrated ^{3}He drift were less than 4 mol year^{-1}.

4.2 Results

The horizontal distributions of the depth-averaged tracer distributions provide a global description of the performance of the simulations (Fig. 3). The inter-basin tracer distribution is qualitatively similar among the simulations. The highest δ^3He values are found in the Pacific Ocean where the rate of injection is the most elevated (Fig. 1b), and no newly ventilated deep water is formed. The zonal contrast is important facet of this basin, with high ^{3}He values simulated close to the East Pacific Rise where the mantle helium flux reaches a maximum. The lowest δ^3He values are simulated in the Atlantic Ocean where the rate of injection is the lowest, and recently ventilated NADW waters with a low δ^3He value are overflowing. In the Indian and southern Ocean, where the rate of tracer injection is intermediate, δ^3He values are in between those simulated in the Pacific and Atlantic basins. Still lower values are observed in the Southern ocean where upwelling in the Antarctic Circumpolar Current (ACC) tends to remove helium by bringing it to the surface where it degasses to the atmosphere. However, large quantitative differences are observed in δ^3He among the simulations, illustrating the large sensitivity of this tracer to the ocean circulation. Modeled δ^3He values are straightforwardly interpreted as a function of the strength of meridional overturning circulation. The control experiment with the weakest overturning circulation (Fig. 2), features a longer residence time of the tracer in the ocean (Table 2) and the highest simulated δ^3He values. The increase in meridional overturning simulated in the other experiments acts to reduce the residence time of the tracer in the ocean (Table 2) and generates globally lower δ^3He values (Table 1). The global mean ^{3}He isotopic composition is reduced by 8, 27 and 32%, respectively, from exp 2 to exp 4 (Table 2). These percentages are also more elevated in the

[2] $\delta^3\mathrm{He} = 100((\mathrm{R_{ocean}}/\mathrm{R_{atmosphere}}) - 1)$ and $\mathrm{R} = {}^3\mathrm{He}/{}^4\mathrm{He}$.

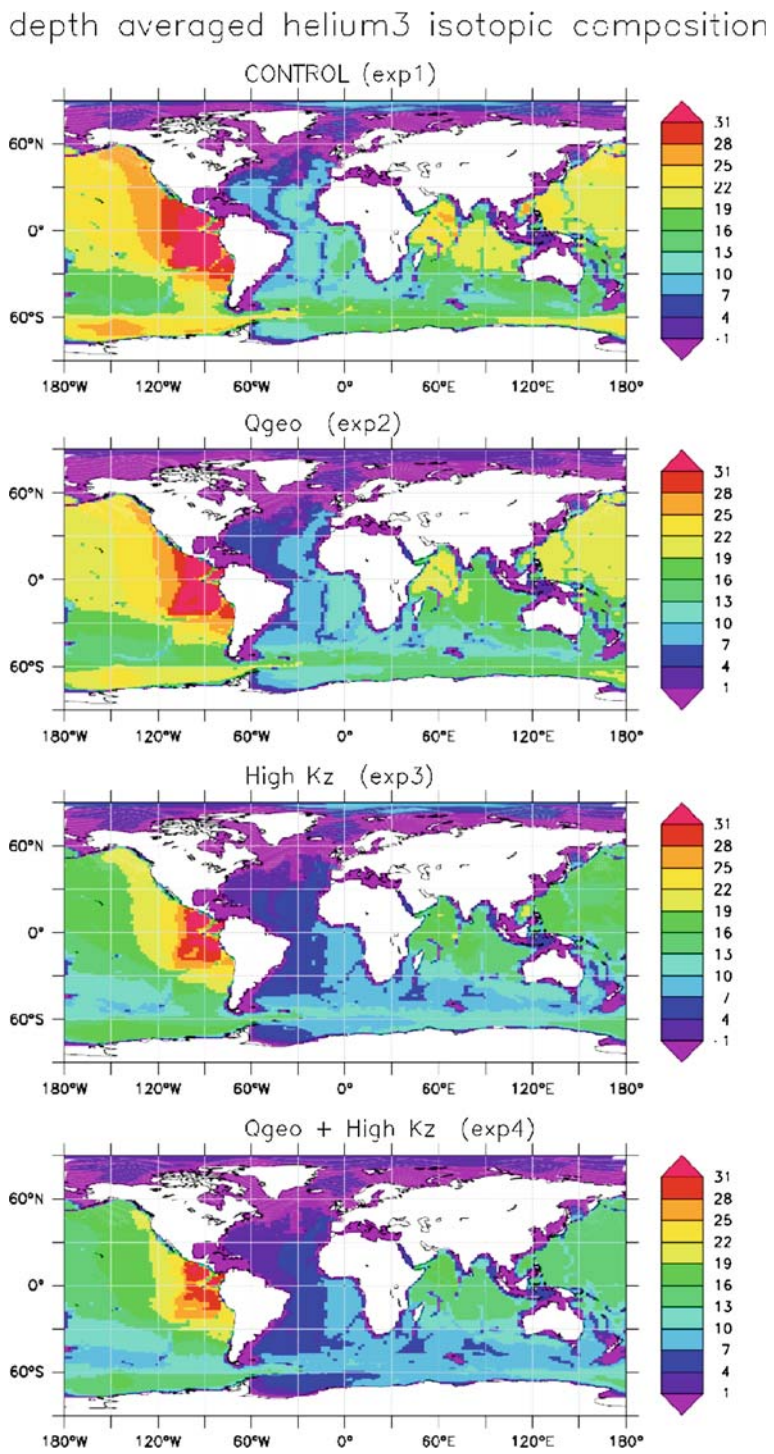


Fig. 3 Depth averaged helium-3 isotopic composition (units: %)

Table 2 Mantle ^{3}He residence time (units: year) defined as the ratio of the global mantle ^{3}He inventory to the mantle ^{3}He source

	^{3}He residence time
Exp 1: Control	855
Exp 2: Q_{geo}	789
Exp 3: High K_z	651
Exp 4: Q_{geo}+High K_z	604

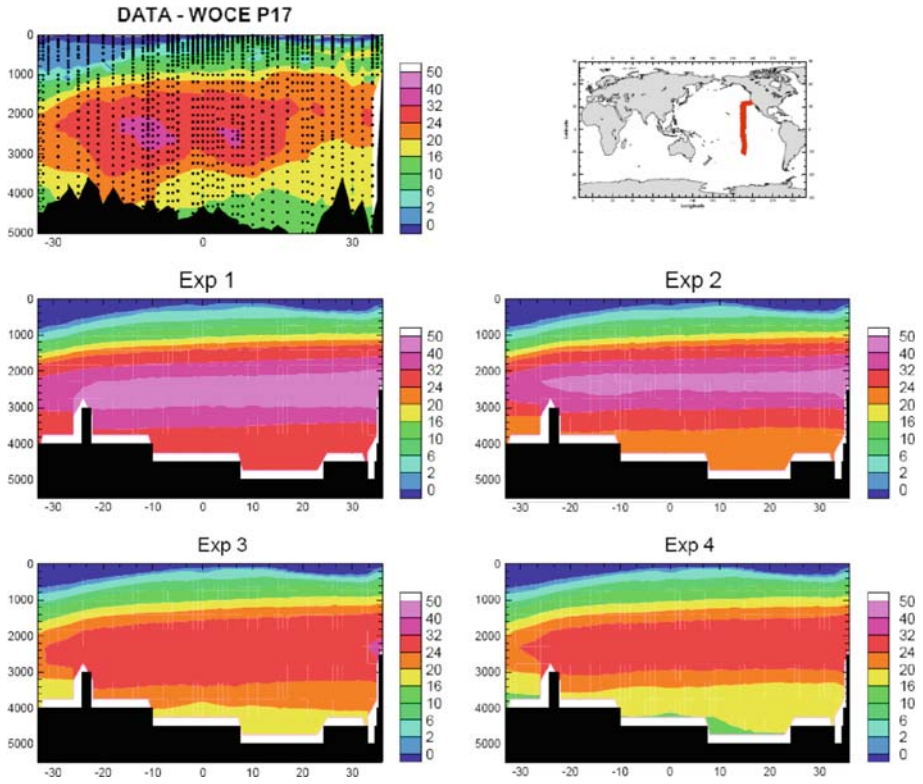


Fig. 4 Observed and simulated helium-3 isotopic composition along the WOCE P17 section in the Pacific Ocean (units: %)

Atlantic and Southern Ocean, where deep waters are formed and ^{3}He isotopic composition is lower.

An opportunity to evaluate the simulated natural ^{3}He distributions is provided by the data collected during the GEOSECS and WOCE programs (http://www.ldeo.columbia.edu/res/fac/etg/text/woce_data.html) [29,30,44,47,57]. Model outputs are compared with observations along the WOCE P17 and WOCE P04 in the Pacific Ocean (Figs. 4, 5), the GEOSECS section in the Atlantic Ocean (Fig. 6) and the WOCE I07 sections in the Indian Ocean (Fig. 7). In the Pacific Ocean, the control experiment (exp 1) has ^{3}He isotopic compositions which are largely overestimated compared to the two WOCE sections (Figs. 4, 5). The weak circulation generated in this experiment produces δ^3He values almost twice as large as those observed. When geothermal heating is added in the second experiment, δ^3He values are globally reduced. The variations are most noticeable below 3000 m in the AABW cell, where modeled values

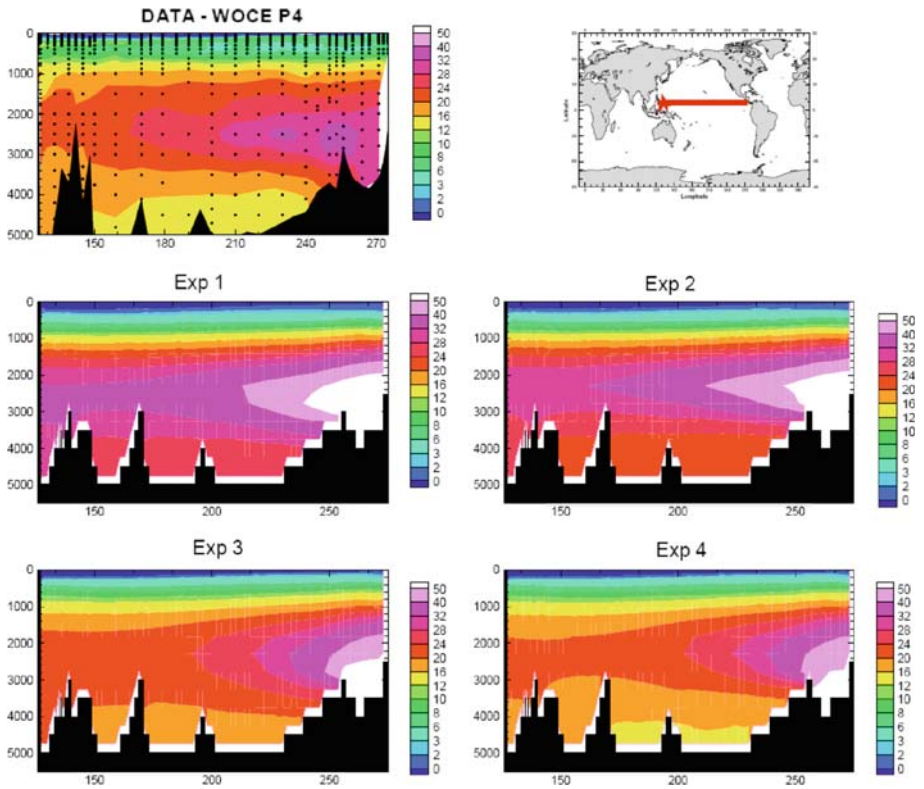


Fig. 5 Observed and simulated helium-3 isotopic composition along the WOCE P4 section in the Pacific Ocean (units: ‰)

are reduced by 20%, but still remain too high relative to the observations. Changes in δ^3He isotopic composition are still more pronounced when vertical mixing is increased (exp 3). The reduction in δ^3He values is again globally accentuated, with a substantial decrease now also occurring in the deep and intermediate water (1000, 3000 m) near the tracer injection depth. These changes lead to a range of modeled δ^3He values now comparable to observations in the Pacific Ocean. When geothermal forcing is added to the version with high vertical mixing (exp 4), very pronounced changes again occur in the simulated bottom waters. The simulated ^{3}He isotopic compositions then compare much more favorably to observations along each section, especially for AABW. There however remain discrepancies in the simulations that are generally attributable to the coarse resolution of the model. The westward propagation of the enriched ^{3}He plume is too weak (Fig. 4) and the two δ^3He maxima associated with plume jets on both sides of the equator are not simulated (Fig. 5).

In general, a comparison to observations reveals similar performances in other basins. The progressive decrease of simulated δ^3He values following the evolution of the configuration of the model observed in the Pacific Ocean, is found again in the Indian Ocean (Fig. 7). However, if the modeled δ^3He values are globally reduced from exp 1 to exp 4, they remain, in general, largely overestimated compared to the observations, especially at mid-depth where the tracer is injected. The only water mass showing a realistic improvement in ^{3}He isotopic composition is AABW, which is quite realistic in exp 4.

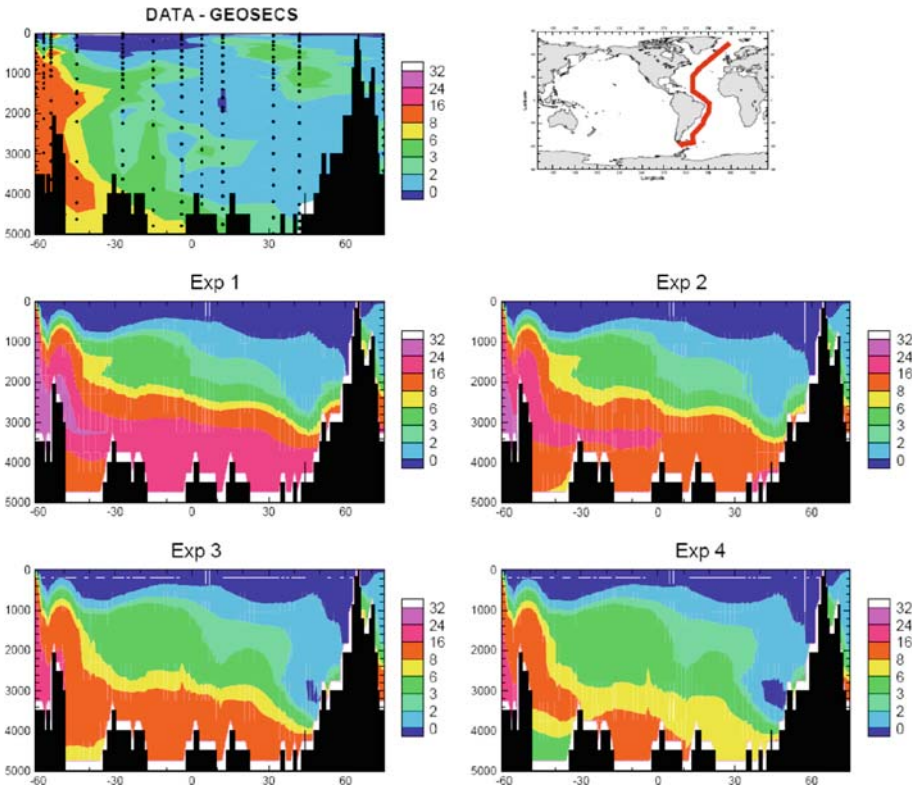


Fig. 6 Observed and simulated helium-3 isotopic composition along the GEOSECS section in the Atlantic Ocean (units: %)

In the Atlantic Ocean where model sensitivity is most pronounced (Table 1), comparison with the GEOSECS observations (Fig. 7) reveals that the main improvement among the different simulations occurs for bottom waters flowing from the Southern Ocean. Large variations are observed in the AABW helium isotopic composition, with the control experiment displaying a higher bias here than in other basins. A noticeable improvement is observed in the Antarctic Circumpolar Current (ACC) for the experiments with higher vertical mixing (exp 3 and exp 4), that is in turn propagated through the AABW isotopic signature and provides more realistic values compared to bottom water observations. This is especially true of the last experiment where geothermal forcing is included (exp 4). Conversely, all simulations successfully produce a realistic helium isotopic composition at depth in the North Atlantic Ocean close to the formation region of NADW. However, NADW gets inconsistently enriched in ^{3}He during its southward journey through the Deep Western Boundary Current, in a generally similar fashion for all simulations.

5 Discussion

It is clear that the main reason for the discrepancies between the control simulation and observations is an excessively weak abyssal circulation, especially for bottom water. Though

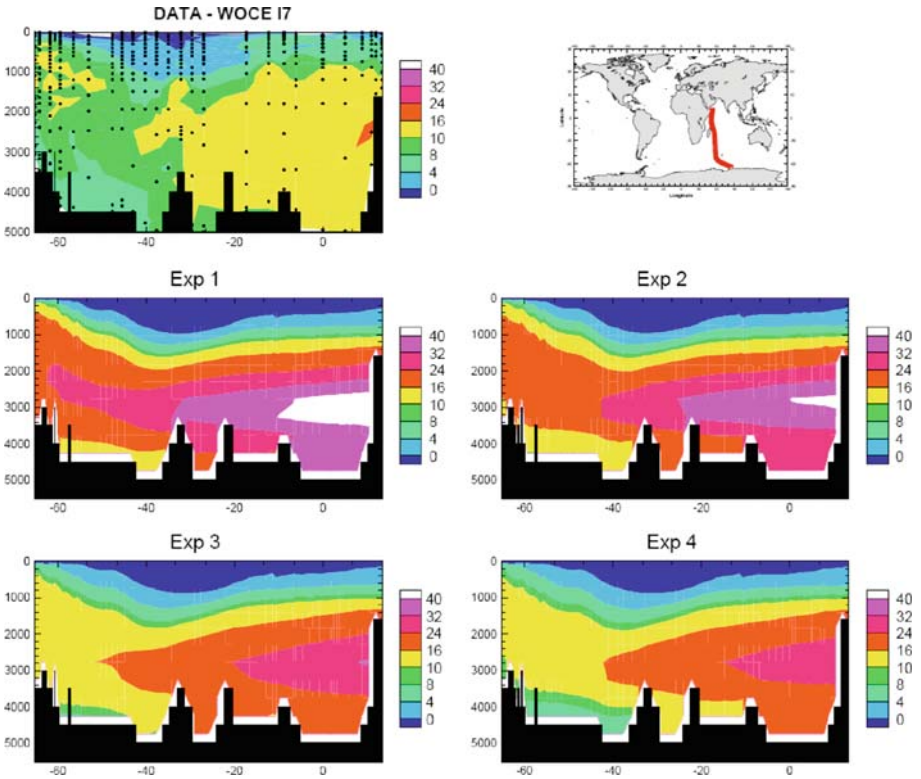


Fig. 7 Observed and simulated helium-3 isotopic composition along the WOCE I7 section in the Indian Ocean (units: %)

modeled NADW overturning in exp 1 (15 Sv) ranks close to the lower limit of a variety of observationally-based estimations (13–15 Sv in [18], 12–23 Sv in [45], 18 Sv in [50]), the intensity of the simulated AABW cell (2.5 Sv) is far below observational estimates (5–8 Sv production rate [37], 22 Sv [50]). This sluggish circulation generates a long ^{3}He residence time and unrealistic isotopic mantle ^{3}He values that largely exceed the observations. The slow circulation is due in part to the weak vertical mixing prescribed throughout the water column. The value of the mixing coefficient (0.1 cm^2 s^{-1}) is adequate for the thermocline—where stratification is strong by definition—but inadequate for bottom water where greater mixing is observed, especially in oceanic regions of rough topography [24].

The modeled thermohaline circulation can be strengthened by increasing vertical mixing as suggested by Bryan [7], but also by including geothermal flux as a forcing, as first suggested by Adcroft et al. [1]. Here we have evaluated in more detail the impact of these two processes on the ocean circulation with mantle ^{3}He simulation.

Adding the geothermal heat flux (exp 2) strengthens the AABW cell, but has little impact on the circulation above 3000 m depth (Fig. 2). When compared to the control experiment, this simulation leads to unrealistically high δ^3He values above 3000 m depth. While the intensity of the AABW cell is increased (5 Sv), it is still too moderate to provide a realistic range of helium isotopic composition near the bottom. We find that changing the vertical mixing parameterization to a version where mixing is enhanced in the deep ocean

globally improves the model's performance (exp 3). The overturning circulation and the simulated $^{\delta 3}$He values are both adjusted towards more realistic values. The improvement in AABW is more significant than when including the geothermal heating alone, even if near-bottom $^{\delta 3}$He values remain slightly too high. Increasing vertical mixing leads to noticeable improvement in the simulation at mid-depth in the ocean where the tracer is injected. Deep ocean ventilation is then improved to the point of simulating a realistic range of $^{\delta 3}$He values where the enriched plumes of mantle helium are observed.

When the geothermal heating is added to the version with high vertical mixing (exp 4) the relative change is less intense than when the mixing was low (exp 1/exp 2) because the deep ocean circulation is already vigorous enough to readily remove the additional heating injected at the bottom [1,45]. Equivalently, strong diapycnal mixing generates important lateral temperature gradients at depth by mapping surface density patterns onto the bottom. The simple scaling law of Emile-Geay and Madec [13] (their Eq. 11) explains why only a small increase in circulation is then necessary to keep the deep ocean in thermal equilibrium for a given geothermal input.

This small circulation change notwithstanding, a noticeable additional improvement is observed on the distribution of isotopic mantle helium in AABW, where $^{\delta 3}$He values become comparable to the observations in all basins. With the exception of the Indian ocean, the simulated helium isotopic composition now appear globally quite reasonable, even if some discrepancies (classically attributed to coarse resolution models) remain:

- The overflow of the densest NADW formed in the Nordic Seas is poorly simulated [5,11,15,16] such that NADW ventilation does not penetrate below 3000 m depth. Consequently, the Deep Western Boundary Current is too sluggish to maintain a realistic ^{3}He isotopic composition of newly formed NADW while it is transported to the south.
- The model failed to correctly reproduce the eastward propagation of the enriched ^{3}He plume in the deep ocean Pacific observed along the WOCE PO4 Section (Fig. 5), maintaining too high values close to the East Pacific Rise source.

The Indian Ocean is the only basin where the model failed to produce a realistic range of ^{3}He altogether, with all experiments largely exceeding the observations. The source function could be to blame, but the results obtained during the multimodel OCMIP2 project tend to indicate that this formulation of the source function is not inaccurate [12]. On the other hand, the deep overturning circulations of all our simulations are very weak in this basin (<2 sv). This is far below all observational estimates: 27 ± 10 Sv [51], 12 ± 3 Sv [41], 11 ± 3 Sv [19], 23 ± 3 Sv [46]. This large discrepancy in the overturning circulation in the Indian Ocean is common to a large set of general circulation models [38]. In our simulations, it is unambiguously the main reason for the overestimation of the helium isotopic composition in this basin.

6 Conclusion

Having taken into account the experience gained during the OCMIP2 project, we have performed simulations with a new configuration of the OPA model in order to improve its deep ocean ventilation. Natural ^{3}He, with its large sensitivity to ocean circulation, proves a useful constraint for the evaluation of the validity of the simulated deep circulation, as large uncertainties remain in the estimations of ocean overturning derived from hydrographic data [50].

It was found that the control version of the model, with low vertical mixing, generated an unrealistically sluggish circulation that led to an overestimated ^{3}He isotopic composition. Including the parameterization with enhanced vertical mixing in the deep ocean provided a major improvement to the model's circulation. It is thus recommended to pursue this effort and to test some more sophisticated parameterizations, such those developed for representing the effect of tides on the topography [24].

Second, the inclusion of a geothermal forcing provided an additional—though modest—enhancement of the AABW circulation. However, its impact on simulated ^{3}He was substantial and brought the simulated tracer distributions in a close agreement with observations for most basins. This supports the notion that ^{3}He is a key oceanographic tracer, providing constraints independent of, and complementary to, hydrographic transport estimates. This additional information corroborates the suggestion that geothermal heating has a significant dynamical impact on the simulated circulation of our model [13] and, we expect, other ocean GCMs besides those where it has already been found [1,45,52]. Since it was found in Emile-Geay and Madec [13] that the inclusion of a realistic geothermal heat flow both improved the fit to both circulation and temperature field, ^{3}He is now the third observational constraint to be positively affected by this process in OPA. Taken together, these constraints strongly suggest that geothermal heating is indeed essential to abyssal dynamics.

The next step would be to carry out a similar analysis with different ocean GCMs, possibly including a broader suite of tracers to diversify the constraints and help discriminate between the most relevant physical processes and parameterizations. Using higher resolution models would also offer the opportunity to better simulate the detailed structure of the δ^3He distribution observed in the deep ocean and even allow to simulate other dynamical process that can affect the δ^3He distribution, such as entrainment and vertical pumping associated with the geothermal heat flux [28,49]. This will be left for future investigation.

References

1. Adcroft A, Scott J, Marotzke J (2001) Impact of geothermal heating on the global ocean circulation. Geophys Res Lett 28:1735–1738
2. Adkins JF, Ingersoll AP, Pasquero C (2005) Rapid climate change and conditional instability of the glacial deep ocean from the thermobaric effect and geothermal heating. Q Sci Rev 24:581–594
3. Beckmann A, Döscher R (1997) A method for improved representation of dense water spreading over topography in geopotential-coordinate models. J Phys Oceanogr 27:581–591
4. Blanke B, Delecluse P (1993) Variability of the tropical ocean simulated by a general circulation model with mixed layer-physics. J Phys Oceanogr 23:1363–1388
5. Böning CW, Holland WR, Bryan FO, Danabasoglu G, McWilliams JC (1995) An overlooked problem in model simulations of the thermohaline circulation and heat transport in the Atlantic Ocean. J Clim 8:515–523
6. Broecker WS, Peng TH (1982) Tracers in the sea, Lamont Doherty Geological Observatory. Palisades, New York, 690 pp
7. Bryan F (1987) Parameter sensitivity of primitive equation ocean general circulation models. J Phys Oceanogr 17:970
8. Craig H, Clarke WB, Beg MA (1975) Excess ^{3}He in deep water on the east Pacific rise. Earth Planet Sci Lett 26:125–132
9. Danabasoglu G, MCWilliams JC, Gent PR (1994) The role of mesoscale tracer transports in global ocean crirculation. Science 264:1123–1126
10. Duffy PB, Caldeira K, Selvaggi J, Hoffert MI (1997) Effects of subgrid-scale mixing parametrisations on simulated distributions of natural ^{14}C, temperature and salinity in a three-dimensional ocean general circulation model. J Phys Oceanogr 27:498–523
11. Dutay J-C, Bullister JL, Doney SC, Orr JC, Najjar R, Caldeira K, Campin J-M, Drange H, Follows M, Gao Y, Gruber N, Hecht MW, Ishida A, Joos F, Lindsay K, Madec G, Maier-Reimer E, Marshall JC,

Matear RJ, Monfray P, Mouchet A, Plattner G-K, Sarmiento J, Schlitzer R, Slater R, Totterdell IJ, Weirig M-F, Yamanaka Y, Yool A (2002) Evaluation of ocean model ventilation with cfc-11: comparison of 13 global ocean models. Ocean Model 4:89–120
12. Dutay J-C, Jean-Baptiste P, Campin JM, Ishida A, Maier-Reimer E, Matear RJ, Mouchet A, Totterdell IJ, Yamanaka Y, Rodgers K, Madec G, Orr JC (2004) Evaluation of OCMIP-2 ocean model's deep circulation with mantle ^{3}He. J Mar Syst 48:15–36
13. Emile-Geay J, Madec G (2009) Geothermal heating, diapycnal mixing, and the abyssal circulation. Ocean Sci 5:203–218
14. England MH, Hirst AC (1997) Chlorofluorocarbon uptake in a world ocean model 2. Sensitivity to surface thermohaline forcing and subsurface mixing parameterizations. J Geophys Res 102:15709–15731
15. England M, Holloway G (1998) Simulations of cfc content and water mass age in the deep north Atlantic. J Geophys Res 103:15885–15901
16. England MH, Rahmstorf S (1999) Sensitivity of ventilation rates and radiocarbon uptake to subgrid-scale mixing in ocean models. J Phys Oceanogr 29:2802–2827
17. Farley KA, Maier-Reimer E, Schlosser P, Broecker WS (1995) Constraints on mantle ^{3}He fluxes and deep-sea circulation from an oceanic general circulation model. J Geophys Res 100:3829–3839
18. Ganachaud A (2003) Large-scale mass transports, water mass formation, and diffusivities from world ocean circulation experiment (WOCE) hydrographic data. J Geophys Res 108. doi:10.1029/2002JC001565
19. Ganachaud A, Wunsch C, Marotzke J, Toole J (2000) Meridional overturning and large-scale circulation of the Indian Ocean. J Geophys Res 105:26117–26134
20. Gent PR, Willebrand J, McDougall TJ, McWilliams JC (1995) Parameterizing eddy-induced tracer transports in ocean circulation models. J Phys Oceanogr 25:463–474
21. Goosse H, Campin JM, Fichefet T, Deleersnijder E (1997) Impact of sea-ice formation on the properties of Antarctic bottom water. Ann Glaciol 25:276–281
22. Iudicone D, Madec G, Blanke B, Speich S (2008) The role of the southern ocean surface forcings and mixing in the global conveyor belt. J Phys Oceanogr 38:1357–1376
23. Iudicone D, Speich S, Madec G, Blanke B (2008) The global conveyor belt from a southern ocean perspective. J Phys Oceanogr 38:1401–1425
24. Jayne SR, Laurent LCS (2001) Parameterizing tidal dissipation over rough topography. Geophys Res Lett 28:811–814
25. Jean-Baptiste P (1992) Helium-3 distribution in the deep world ocean: its relation to hydrothermal ^{3}He fluxes and to terrestrial heat budget. In: Consultants meeting on isotopes of noble gases as tracers in environmental studies, Vienna
26. Joyce T, Warren B, Talley L (1986) The geothermal heating of the abyssal subarctic Pacific Ocean. Deep Sea Res 33(8):1003–1015
27. Kamenkovich IV, Goodman PJ (2000) The dependence of AABW transport in the Atlantic on vertical diffusivity. Geophys Res Lett 27:3739–3742
28. Lupton JE (1995) Hydrothermal plumes: near and far field. In: Humphris S, Zierenberg R, Mullineaux L, Thomson R (eds) Seafloor hydrothermal systems: physical, chemical, biological, and geological interactions. Geophysical Monograph 91. American Geophysical Union, Washington, DC, pp 317–346
29. Lupton JE (1996) A far-field hydrothermal plume from Loihi Seamount. Science 272:976–979
30. Lupton JE (1998) Hydrothermal helium plumes in the Pacific Ocean. J Geophys Res 103:15853–15868
31. Lupton JE, Craig H (1981) A major helium-3 source at 15°S on the east pacific rise. Science 214:13–18
32. Madec G, Imbard M (1996) A global ocean mesh to overcome the north pole singularity. Clim Dyn 12:381–388
33. Madec G, Delecluse P, Imbard M, Lévy C (1998) Opa8.1 ocean general circulation model reference manual. Notes du pôle de Modélisation de l'IPSL 11
34. Müller RD, Roest WR, Royer J-Y, Gahagan LM, Sclater JG (1997) Digital isochrons of the ocean floor. J Geophys Res 102(32):3211–3214
35. Munk W, Wunsch C (1998) Abyssal recipes II: energetics of tidal and wind mixing. Deep Sea Res 45:1977–2010
36. Naveira Garabato AC, Stevens DP, Watson AJ, Roether W (2007) Short-circuiting of the oceanic overturning circulation in the Antarctic Circumpolar Current. Nature 447(7141):194–197. doi:10.1038/nature05832
37. Orsi AH, Johnson GC, Bullister JL (1999) Circulation, mixing, and production of antarctic bottom water. Prog Oceanogr 43:55–109
38. Palmer MD, Naveira Garabato AC, Stark JD, Hirschi JJ-M, Marotzke J (2007) The influence of diapycnal mixing on quasi-steady overturning states in the Indian Ocean. J Phys Oceanogr 37:2290–2304

39. Paul A, Schäfer-Neth C (2003) Modeling the water masses of the Atlantic Ocean at the Last Glacial Maximum. Paleoceanography 18(3):1058. doi:10.1029/2002PA000783
40. Rahmstorf S (2002) Ocean circulation and climate during the past 120,000 years. Nature 419:207–214
41. Robbins PE, Toole JM (1997) The dissolved silica budget as a constraint on the meridional overturning circulation of the Indian Ocean. Deep Sea Res I 44:879–906
42. Roullet G, Madec G (2000) Salt conservation, free surface and varying volume: a new formulation for ocean GCMs. J Geophys Res 105:23927–23942
43. Schmitz WJ Jr (1995) On the interbasin-scale thermohaline circulation. Rev Geophys 33:151–173
44. Schlösser P, Bullister JL, Fine R, Jenkins WJ, Key R, Lupton J, Roether W, Smethie W (2001) Transformation and age of water masses. In: Siedler G, Church J, Gould J (eds) Ocean circulation and climate. Academic Press, San Diego
45. Scott J, Marotzke J, Adcroft A (2001) Geothermal heating and its influence on the meridional overturning circulation. J Geophys Res 106:1–14
46. Sloyan BM, Rintoul SR (2001) The southern ocean limb of global deep overturning circulation. J Phys Oceanogr 31:143–173
47. Srinivasan A, Top Z, Schlosser P, Hohmann R, Iskandarani M, Olson DB, Lupton JE, Jenkins WJ (2004) Mantle ^{3}He distribution and deep circulation in the Indian Ocean. J Geophys Res 109:C06012. doi:10.1029/2003JC002028
48. Stein CA, Stein S (1992) A model for the global variation in oceanic depth and heat flow with lithospheric age. Nature 359:123–129
49. Stommel H (1982) Is the south Pacific helium-3 plume dynamically active?. Earth Planet Sci Lett 61: 63–67
50. Talley LD, Reid JL, Robbins PE (2003) Data-based meridional overturning streamfunctions for the global ocean. J Clim 16:3213–3226
51. Toole JM, Warren BA (1993) A hydrographic section across the subtropical south Indian Ocean. Deep Sea Res I 40:1973–2019
52. Urakawa LS, Hasumi H (2009) A remote effect of geothermal heat on the global thermohaline circulation. J Geophys Res 114:C07016. doi:10.1029/2008JC005192
53. Visbeck M (2002) Climate—the ocean's role in Atlantic climate variability. Science 297(5590): 2223–2224
54. Visbeck M (2007) Concept oceanography: power of pull. Nature 447:383. doi:10.1038/447383a
55. Wanninkhof R (1992) Relationship between wind speed and gaz exchange over the ocean. J Geophys Res 97:7373–7382
56. Weber SL, Drijfhout SS, Abe-Ouchi A, Crucifix M, Eby M et al (2007) The modern and glacial overturning circulation in the Atlantic ocean in PMIP coupled model simulations. Clim Past 3:51–64
57. Wijffels SE, Toole JM, Bryden HL, Fine RA, Jenkins WJ, Bullister JL (1996) The water masses and circulation at 10N in the Pacific. Deep Sea Res 43:501–544
58. Wunsch C, Ferrari R (2004) Vertical mixing, energy, and the general circulation of the oceans. Annu Rev Fluid Mech 36:281–314

Environ Fluid Mech (2010) 10:275–295
DOI 10.1007/s10652-009-9154-3

ORIGINAL ARTICLE

Boundary impulse response functions in a century-long eddying global ocean simulation

Mathew Maltrud · Frank Bryan · Synte Peacock

Received: 22 May 2009 / Accepted: 17 September 2009 / Published online: 9 October 2009

Abstract Results are presented from a century-long 1/10° global ocean simulation that included a suite of age-related passive tracers. In particular, an ensemble of five global Boundary Impulse Response functions (BIRs, which are statistically related to the more fundamental Transit Time Distributions, TTDs) was included to quantify the character of the TTD when mesoscale eddies are explicitly simulated rather than parameterized. We also seek to characterize the level of variability in water mass ventilation timescales arising from eddy motions. The statistics of the BIR timeseries are described, and it is shown that the greatest variability occurs at early times, followed by a remarkable conformity between ensemble members at longer timescales. The statistics of the first moment of the BIRs are presented, and the upper-ocean spatial distribution of the standard deviation of the first moment of the BIRs discussed. It is shown that variations in the BIR first moment with respect to the ensemble average are typically only a few percent, and that the variability slightly decreases with increasing ensemble size, implying that only a few ensemble members may be necessary for a reasonable estimate of the TTD. The completeness of the estimated TTD, i.e., the degree to which the century long BIRs capture the range of global ocean ventilation timescales is discussed, and the potential for extrapolation of the BIR to longer times is briefly explored. Several regional BIRs were also simulated in order to quantify the relative abundance of fluid parcels that originate in specific geographical locations.

Keywords Modeling · Ocean circulation · Tracers · Timescales

M. Maltrud (✉)
Los Alamos National Laboratory, Los Alamos, NM 87545, USA
e-mail: maltrud@lanl.gov

F. Bryan · S. Peacock
National Center for Atmospheric Research, Boulder, CO 80307, USA

1 Introduction

Analysis of temperature and salinity distributions to characterize water mass properties, their formation rates, and pathways and timescales of transport in the ocean interior, has been at the foundation of observational oceanography for over a century. The availability of global scale observations of large numbers of natural and anthropogenic tracers during the last several decades has provided a more quantitative basis for characterizing the timescales associated with ocean transport and mixing. One particular timescale, the time since a water parcel was last in contact with the atmosphere (often referred to as "water-mass age") is of special interest in climate research. Tracer based estimates of water-mass age have been used to directly infer timescales of ocean ventilation, which gives bounds on the rate at which the ocean can sequester materials such as carbon dioxide.

In practice, however, various tracers yield different age estimates due to their distinct atmospheric time histories, their particular chemical properties, and oceanic sources and sinks. These differences can be understood through the conceptual framework of Transit Time Distributions (TTDs) (also called "age spectra" and "transit time probability density functions" [10,26]). The TTD is a probability distribution of age, which explicitly accounts for the fact that a water mass is comprised of a mixture of components with different ages, each with different advective and diffusive histories. Holzer and Hall [10] clarify the mathematical connection between this probability distribution function and the Greens function for the advection–diffusion equation governing tracer transport.

Idealized age tracers, and more recently TTDs, have been carried in simulations of the ocean ([4,11,12,14,25] and [24] (hereafter referred to as PM06)). Calculation of TTDs from ocean models with resolution typical of that used in climate research (horizontal grid spacing of order 1° latitude and longitude) is relatively straightforward, and achieved by exploiting the connection to the Greens function for the transport equation. The TTD in coarse resolution models can be obtained by a single forward integration of the model transport equation for a conservative tracer with an impulse boundary condition. This is possible because models in this resolution regime have essentially steady flow (excepting the seasonal cycle) and the transport by turbulent eddies is parameterized through some form of diffusive process. As we shall see, the simulation of a TTD-like tracer becomes much more complex when the model resolution becomes high enough to permit a fully turbulent, non-steady flow.

Both observational and modeling studies suggest that the ocean TTD typically has a long tail, an indication that mixing processes are important in determining the mean ventilation rate. Thus, the model estimates of the TTD from coarse-resolution models are critically dependent on the parameterization of turbulent mixing. However, it is not clear from coarse-resolution model simulations alone how to select an optimal mixing parameterization. Since the majority of the horizontal mixing in the ocean is due to turbulence from mesoscale eddies, directly assessing the effects of these motions requires a model that can resolve them. Here we describe the first ever simulation designed to estimate the ocean TTD using a global general circulation model in which mesoscale eddies are simulated directly rather than parameterized.

Computing the TTD from an unsteady turbulent circulation is complicated by the fact that the transit times themselves become time-dependent. The result of a single forward integration of the transport equation with an impulse boundary condition, formally referred to as the Boundary Impulse Response (BIR), can no longer be interpreted as a true probability distribution function of transit times, as is the case with steady flow. However, Haine et al. [8] (hereafter referred to as HZWH08) have shown that even for unsteady flow, the BIR and TTD share the same statistical moments under certain conditions. We exploit this property

to estimate the TTD as the average of a (modest sized) ensemble of BIRs, which are directly simulated in a global eddying simulation.

The high computational cost of running a global eddy resolving simulation precludes an integration of the several thousand years required to capture the full range of ocean ventilation time scales (PM06). We will present results from a century-scale simulation, which provides a useful basis for studying ocean ventilation processes in the more rapidly ventilated parts of the ocean. These are just the regions that are most relevant with respect to anthropogenic impacts on the ocean and the uptake of transient tracers such as CO_2 and Chlorofluorocarbons.

This paper begins with a brief overview of the TTD theory; the model configuration and experiment design is described, and the simulated BIRs are presented. There is a discussion of both the ensemble statistics of the BIR timeseries, the statistics of the first moment, and how these relate to the TTD. The issue of BIR convergence is also explored. Finally, preliminary results from a set of BIRs which are forced by boundary impulses in specific geographical regions are presented.

2 Boundary propagators and TTDs

The theory of TTDs has previously been developed and applied to oceanic flows [9,7], so only the most basic development will be presented here. Starting with the equation for the evolution of the concentration of a conservative passive tracer, C,

$$\left(\frac{\partial}{\partial t} + L(\mathbf{x}, t)\right) C(\mathbf{x}, t) = 0 \tag{1}$$

where $L(\mathbf{x}, t)$ represents advection and diffusion, an equation can be derived for the quantity $G'(\mathbf{x}, t)$,

$$\left(\frac{\partial}{\partial t} + L(\mathbf{x}, t)\right) G'(\mathbf{x}, t; \mathbf{x}_0, t') = 0 \tag{2}$$

which satisfies the convolution integral,

$$C(\mathbf{x}, t) = \int_0^t dt' \int_\Omega d^2\mathbf{x}_0 G'(\mathbf{x}, t; \mathbf{x}_0, t') C(\mathbf{x}_0, t') \tag{3}$$

where $\mathbf{x}_0$ is a two dimensional vector defined over an ocean surface region Ω. Thus, the boundary propagator, G', "propagates" the surface value of C from Ω into the interior of the ocean. If the convolution integral (Eq. 3) is rewritten as a function of $\tau = t - t'$ (the "transit time"), the kernel has a clear interpretation as a weighting based on the last time a water mass component was at the ocean surface.

Note that G' is seven dimensional, being a function of all locations in the ocean ($\mathbf{x}$), the "field time" (t), all locations at the surface ($\mathbf{x}_0$), and the "source time" (t'), so is typically intractable to calculate and analyze in simulated oceanic flows. A forward model simulation would need to carry a different tracer for each source time and surface location, which would need to be an extremely large number (order 1,000 even for a small number of surface locations) in order to accurately approximate the TTD. For the case of steady flow, the boundary propagator is invariant under time translation: $G'(\mathbf{x}, t; \mathbf{x}_0, t - \tau) = G'(\mathbf{x}, 0; \mathbf{x}_0, -\tau)$, and the TTD depends only on transit time, not the field time and source time individually, so a single forward integration from source time $t' = 0$ is sufficient to compute the TTD. For the case of

unsteady flow, the time translation invariance no longer applies. However, HZWH08 provide a proof that for statistically stationary flow, the distribution of $G'(\mathbf{x}, t_1; \mathbf{x}_0, t_1 - \tau)$ (the TTD) over $t_1$is identical to the distribution of $G'(\mathbf{x}, t_2 + \tau; \mathbf{x}_0, t_2)$ (the BIR) over t_2. If we are primarily interested in the TTD as a statistical characterization of the turbulent flow rather than as a representation of a specific realization of the flow, as we are in this study, then we can estimate the mean TTD as an average of an ensemble of BIRs integrated forward from a set of source times t_2. This relationship between the TTD and BIR defines the computational approach used in our study: we simulate an ensemble of BIRs by forward integration in an eddying ocean model subject to a repeating annual cycle forcing, and estimate the TTD as their ensemble mean.

3 Model description

3.1 Model configuration

The fully global simulations described here were performed using the Parallel Ocean Program (POP, [3]) developed at Los Alamos National Laboratory. The configuration is based upon that described by Maltrud and McClean [18], hereafter referred to as MM05, but with several notable refinements. The horizontal grid retained the nominal longitudinal resolution of 0.1°, but the grid design has been changed from a dipole, where the northern computational pole was smoothly displaced into Canada, to a tripole layout [23] with poles in Canada and Russia, resulting in a more isotropic distribution of resolution in high northern latitudes. The vertical resolution is the same as in MM05 with two additional deep levels that extend the maximum depth of the model from 5,500 to 6,000 m. The discretization of bottom topography has been changed from full-cell to partial bottom cells [1], creating a more accurate and smoother representation of topographic slopes.

Biharmonic viscosity and diffusivity of tracers are used with a cubic dependence on local grid size as described in Maltrud et al. [19]. In this simulation we have reduced the level of explicit dissipation by a factor of three relative to MM05, using equatorial values of $\nu_0 = -90\ \mathrm{m}^4/\mathrm{s}$ for momentum and $\kappa_0 = -30\ \mathrm{m}^4/\mathrm{s}$ for tracers. As in MM05, vertical mixing coefficients for momentum and tracers are calculated by an implementation of the K-Profile Parameterization [16] that includes the use of large diffusion coefficients ($0.1\ \mathrm{m}^2\ \mathrm{s}^{-1}$) to resolve gravitational instabilities. Background values for vertical tracer diffusion range from $10^{-5}\ \mathrm{m}^2/\mathrm{s}$ near the surface to $10^{-4}\ \mathrm{m}^2/\mathrm{s}$ at depth, with viscosity values an order of magnitude higher. Density is calculated from the equation of state for seawater derived by McDougall et al. [22]. Initial temperature and salinity fields were interpolated from the annual mean WOCE Global Hydrographic Climatology [6].

When simulating passive tracer concentrations, it is especially important that numerical dispersion errors arising from the discretization of advection don't result in spurious negative values. The advection scheme used in these experiments is similar to the third order direct space time with flux limiting scheme of Adcroft et al. [2]. The primary differences are that we replace the underlying third order scheme with the second order Lax–Wendroff scheme, to reduce numerical diffusion, and modify the flux limiters to take spatial variations in velocity and grid spacing into account. However, we found this scheme to be undesirably diffusive when applied to the active tracers, so standard centered advection (with no flux limiting) was used for temperature and salinity, as well as for momentum.

Since one of the fundamental aspects of this simulation was to quantify the eddy component of tracer transport, and the experimental design is predicated on a statistically

stationary flow field, we chose to restrict the timescales introduced by the surface forcing. The atmospheric state was based on the repeat annual cycle (normal-year) Coordinated Ocean Reference Experiment (CORE) forcing dataset [17], with the 6-hourly forcing averaged to monthly. Wind stress was calculated offline using bulk formulae and the Hurrell Sea Surface Temperature (SST) climatology [13,15]; evaporation and sensible heat flux were calculated online using the same bulk formulae and the model predicted SST. Precipitation was also taken from the CORE forcing dataset. Monthly river runoff from 46 major rivers [5] was added to the fresh water flux at the locations of the actual outflow, with the remaining ungauged runoff distributed evenly along the coasts of all of the continents. Ice cover was prescribed based on the $-1.8\,\mathrm{C}$ isoline of the SST climatology, with both temperature and salinity restored on a timescale of 30 days under diagnosed/climatological ice. There is also weak surface restoring of salinity in the open ocean with a timescale of 1 year to prevent global salinity drift.

3.2 BIR implementation and simulation design

In theory, like the TTD, the BIR is the response to Delta function source. In practice, the implementation of the BIR was achieved using a 1 year long top hat function in order to average over the seasonal cycle (the same method was employed by Khatiwala et al. [14] and PM06 in their TTD simulations). This is accomplished by fixing the top layer value of the tracer to (1/year) for 1 year, then resetting to zero at all later times. As a result, the surface becomes a sink for the BIR tracers after the first year.

Because TTDs and BIRs share the same statistics, an ensemble of BIR realizations is necessary to estimate the TTD. Instead of running the model several times with a single BIR tracer, each member of the ensemble was run within the same ocean simulation, but the start times for the first four surface impulses were separated by 2 years. The fifth ensemble member was added after an interval of 30 years. The decorrelation time for oceanic eddies is much smaller than 2 years, so each BIR ensemble member can be considered to be transported by an independent realization of the eddy flow field.

Figure 1 shows a schematic of the simulation design. The model was spun up from rest with no passive tracers for 15 years, which is long enough for the upper ocean circulation to reach a statistically steady state. At this point, tracers were added at the appropriate times and the run continued for another 105 years. An ensemble of five realizations of "global" BIRs (i.e., where the source region Ω is the entire global ocean surface) were simulated. The majority of the presented results will focus on these global BIRs. A single realization of six "regional" BIRs (where Ω is not the entire ocean surface, but a subset thereof) were included to provide additional information on the relative contributions of different surface regions

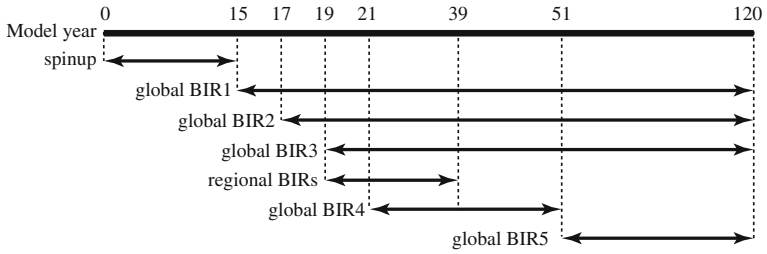


Fig. 1 Schematic of the simulation design showing the starting time and duration of each BIR

in ocean ventilation. Just as with the global BIRs, the surface layer in each defined region is reset to a value of (1/year) every timestep for a year. Outside of each region, the surface value is reset to zero for all times.

As can be seen in Fig. 1, each of the five global BIRs span a different amount of time, but there are two clear choices of record length when performing ensemble statistics. For long time scales, we will consider the first 100 years of BIR1, BIR2, and BIR3, where time is defined relative to January 1 of the year in which the particular BIR was introduced into the simulation. For shorter time scales, the ensemble consists of the first 30 years of all five BIRs.

4 Results

The results shown in this paper will focus on the characteristics of the simulated BIRs. Detailed analysis of the model circulation can be found in Maltrud et al. [20], where a wide array of metrics are applied to demonstrate a high degree of model fidelity when compared to data, theory, and previous high-resolution ocean simulations. As discussed by Maltrud et al. [20], the simulated geographical distribution and amplitude of eddy-activity shows a very close correspondence with observations, and there are many features of the simulated circulation that show a significant improvement over those seen in this model's predecessor (MM05, [21]).

This section begins with a general overview of the global BIR distribution; we then consider the statistics of the timeseries of the ensemble BIR members, and the statistics of the first moment of the BIRs. Finally, we briefly show some results from a set of regional BIR simulations.

4.1 BIR evolution

One way of looking at the BIRs is as a dye tracer with an impulse surface boundary condition. Figure 2 shows snapshots of two BIR realizations in the North and South Atlantic (left and right panels, respectively). In the North Atlantic snapshot, a tongue of high tracer values (red colors) can clearly be seen being advected around Flemish Cap, superimposed on a very turbulent background. The differences in tracer values between the two BIR realizations shown are clearly discernible by eye. In both realizations, very strong tracer gradients are evident across very small spatial scales. Much of this high-frequency variability is smoothed out in the lowermost panel (Fig. 2e), which shows the five-member ensemble average of the fourth year annual mean BIRs. The differences between the two realizations of the tracer in the South Atlantic are not quite so pronounced as in the North Atlantic; however, individual eddies can still clearly be seen as Agulhas rings and turbulence associated with Antarctic Circumpolar Current (ACC) fronts. Again, the ensemble-mean (Fig. 2f) shows a very smoothed version of the snapshots, and all semblance of individual eddies have been erased by the averaging.

The redistribution of the BIRs as the simulation progresses can be seen in the column integrals in Fig. 3. After only 10 years, the Atlantic Deep Western Boundary Current (DWBC) has transported the surface signal from the high-latitude North Atlantic into the Southern Hemisphere. By contrast, in the coarse resolution simulation of PM06 (their figure 1), the net mass transport of the DWBC is fairly realistic, but the model was not able to resolve the narrow width and high speeds of the core of the boundary current. As a result, several decades were required for the DWBC tracer signal to reach the Southern hemisphere in the coarse-resolution simulation.

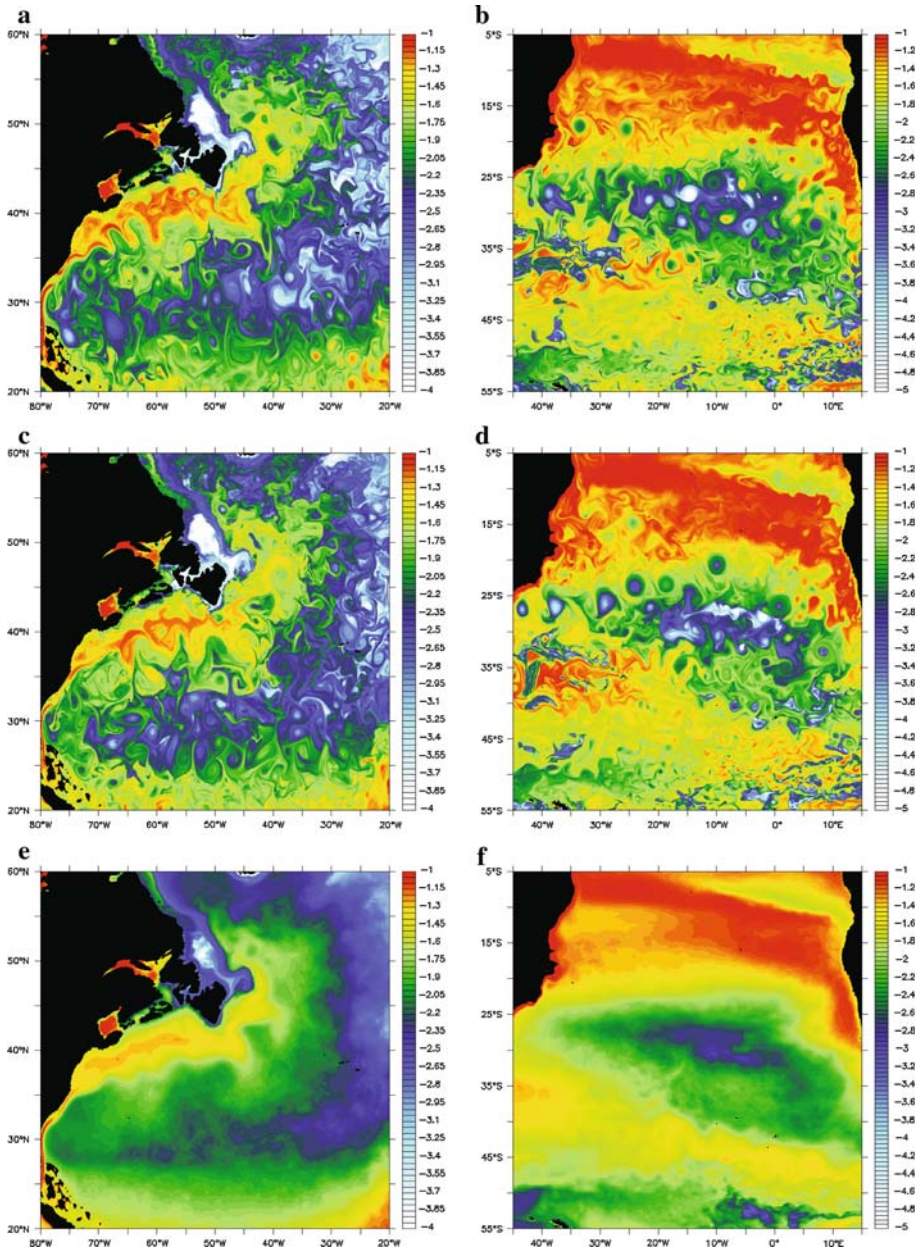


Fig. 2 Snapshots of log(BIR1) (**a**, **b**) and log(BIR2) (**c**, **d**) at 112 m depth on July 1 of year 4 in the Gulf Stream (*left*) and Southern Atlantic (*right*) regions. **e**, **f** Show the five member ensemble average of the year 4 annual mean BIRs for the same regions

After 50 years of model integration (Fig. 3b), some of the tracer transported southward by the Atlantic DWBC has entered the ACC as Circumpolar Deep Water (dark blue colors in the South Atlantic), and some Antarctic Bottom Water (AABW) has made its way into the North Pacific basin (dark blue colors in the Western Pacific). Notable signals of dye are

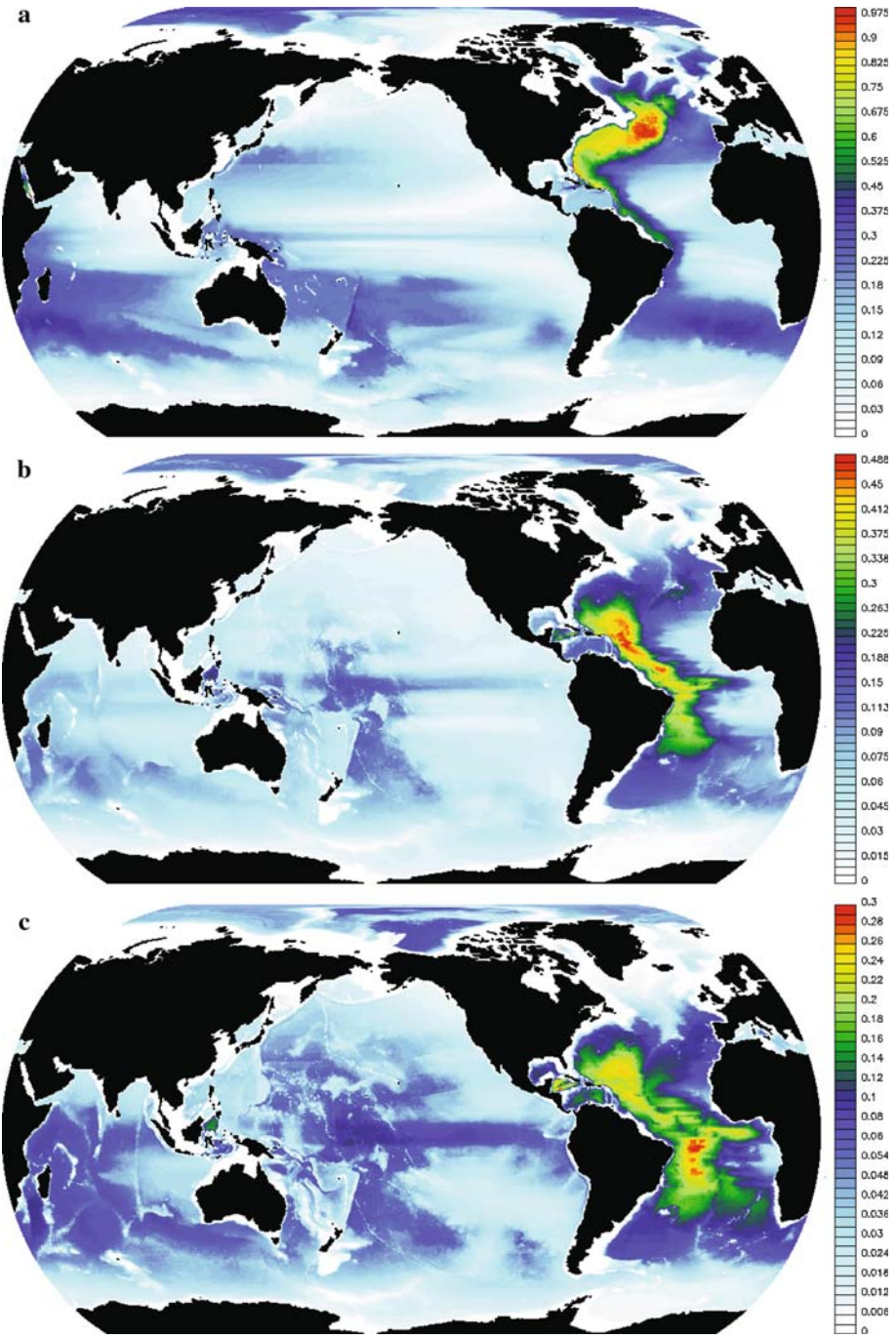


Fig. 3 Vertical integrals (weighted by the grid cell area) of the ensemble average annual mean global BIR for years 10 (five ensemble members), 50 (four ensemble members), and 100 (three ensemble members). Note the change in color scales between panels

also seen along the equator in the central and eastern Pacific, this being a reflection of rapid advection along shallow zonal currents in the equatorial Pacific basin.

At a time of 100 years (Fig. 3c), the highest column inventories are in the western half of the Atlantic, and AABW has spread through much of the abyss in both the Pacific and Indian basins. In PM06 (their figure 1), the DWBC signal has diffused across the entire North Atlantic basin, and the AABW has only begun to cross the equator by this time. It should be noted that it is not possible to distinguish from this figure alone between tracer added by thermocline ventilation and that added by deep ventilation; the inferences above were made by analyzing the distribution of the BIR on individual depth layers.

As described above, HZWH08 prove that TTDs and BIRs share the same statistics. An interesting question is thus how different the individual BIR timeseries are from their ensemble averages. Figure 4 shows the time dependent behavior of the simulated BIRs at several locations for each ensemble member (colored curves) and the ensemble mean (thick black curve). All show the greatest variability at early times, followed by remarkable conformity for times later than the peak in the distribution, (i.e. for the most probable transit time). At the Indian Ocean site, the response is dominated by the mixed layer processes related to the annual cycle (though there is significant eddy variability originating in the Leeuwin Current). In the northeast Pacific thermocline, eddies and differences in arrival time dominate the variability. The Southern Ocean location is exceptional in that it is bimodal, with an early signal from bottom water formed in the Ross Sea, followed by other Southern Ocean bottom water sources (a similar bimodal distribution was also reported in PM06 for most deep Southern Ocean locations). The most notable outlier in Fig. 4 is BIR5 (magenta curve) at the Arctic location. Recall that BIR5 was introduced into the simulation 30 years after BIR4, so this behaviour reflects a significant shift in the Arctic circulation during those three decades. This demonstrates that the flow cannot be considered stationary (at decadal time scales), thus the ensemble averaged BIRs are not a reliable estimate of the TTD in this particular region.

In addition to results from this simulation, a time series of the global BIR from PM06 can be seen in Fig. 4c. As noted above in reference to Fig. 3a, the signal arrives much later and has a broader leading edge in the low resolution (order 3°) case due to sluggish transport velocities and stronger diffusion. As one moves southward in the DWBC, the ratio of the coarse to fine grid arrival time remains fairly constant around a value of 5. For example, at 10°S, the BIRs from the current simulation arrive after about 10 years, compared to 50 years for the PM06 run (not shown).

4.2 Convergence to complete TTD

Because ocean ventilation times are thousands of years for much of the deep ocean, and many hundreds of years for the thermocline (PM06), it is important to determine over how much of the ocean the simulated BIR tracers have captured the full breadth of the TTD within the 100-year integration. To asses this, we define two criteria for locating areas of (near) convergence. Such a metric should prove useful in obtaining a global view of where the moments computed from the simulated TTD, including the mean age, are meaningful. We here show results using two different definitions of convergence:

(1) A *TTD normalization threshold*, which uses a criterion based on the TTD itself. By definition, TTDs must integrate to unity,

$$I(\mathbf{x}) = \lim_{\tau \to \infty} \int_0^{\tau} TTD(\mathbf{x}, t')dt' = 1, \tag{4}$$

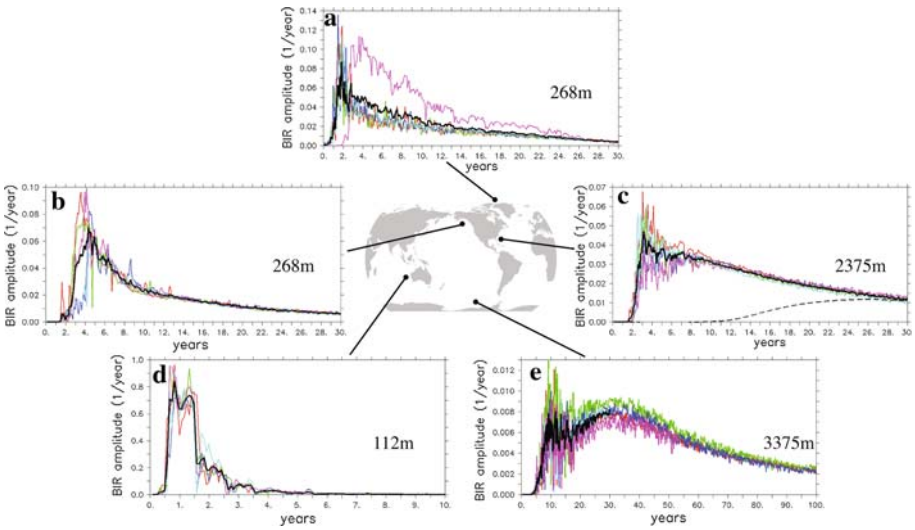


Fig. 4 Time series of monthly mean global BIRs at five locations. BIR1 through BIR5 are denoted by *red*, *green*, *blue*, *cyan*, and *magenta*, respectively, and the five-member ensemble average by the thick *black line*. Note that time axis varies in all of the plots, and that the ensemble mean is only defined for the first 30 years since that is maximum duration of the shortest time series (BIR4). The DWBC location (**c**) also includes a time series (*dashed curve*) from the 3° simulation of PM06. Note that the time axis varies in all of the plots

so a threshold of 0.95 for this integral has been chosen to define convergence. It is important to note that this relation only holds for BIRs in an ensemble average, that is, any given realization of the BIR may integrate to a value somewhat greater than or less than unity (HZWH08). The extent to which the average BIR integral exceeds unity may also be an indicator of how robust the estimate is for a given ensemble size, with highest values corresponding to lowest confidence. However, other factors such as numerical dispersion of the advection scheme or the finite length of the initial tracer impulse could contribute, and we are not able to distinguish between these possibilities with the available model output.

Figure 5 shows TTD integral, $I(\mathbf{x})$, after 100 years at 268 m depth. The colors in Fig. 5 change to yellow at a value of 0.95 (the normalization threshold), so areas that are not blue–green may be close to convergence. In general, the subtropical gyres are relatively close to convergence since they are regions with large amounts of rapidly ventilated mode waters. The North Atlantic subpolar gyre is also close to convergence due to annual wintertime convection reaching this depth. Areas that experience upwelling, such as eastern boundaries and equatorial regions, or that are strongly stably stratified, such as the northern Indian and Weddell Sea, are furthest from converging after 100 years of model integration.

(2) An *age-trend threshold* For direct comparison with ocean data, it is the *age* of a water parcel that is of primary interest, rather than the full BIR or TTD. As noted above, TTDs represent PDFs of transit times at a particular location, so by definition the first moment of the TTD is the mean age of the water mass:

$$Age(\mathbf{x}) = \lim_{\tau\to\infty} Age(\mathbf{x},\tau) = \lim_{\tau\to\infty} \int_0^\tau TTD(\mathbf{x},t')t'dt'. \tag{5}$$

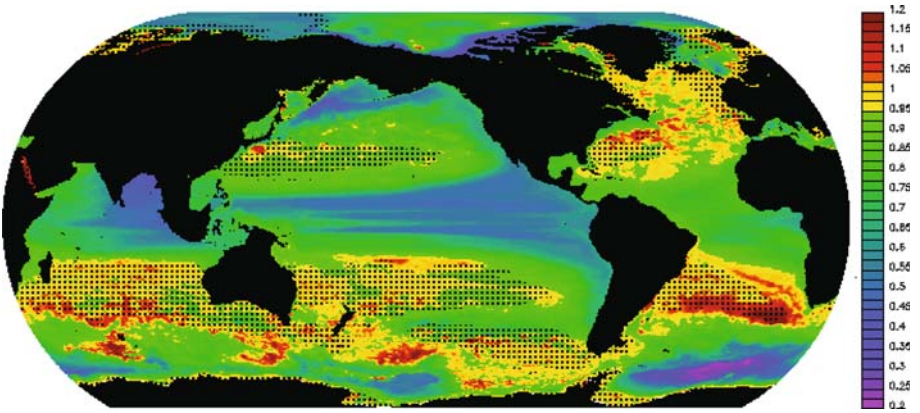


Fig. 5 Ensemble averaged BIR normalization integral (Eq. 4) at 268 m after 100 years of model integration (*colors*). Stipling shows regions of convergence using an age trend threshold (Eq. 5) of 1 year/decade

Because TTDs and BIRs share the same statistics, the first moment of the BIRs can be interpreted as an estimate of the true mean age of the water-mass. Figure 6 shows the running integral of the mean age, calculated using the BIRs at the same locations as in Fig. 4. Note that at each of the selected locations, the mean age of the water-mass changes rapidly with time, before asymptotically approaching the "true" mean age. Running integrals are thus useful as an indicator of the completeness of our century long TTDs, i.e, the degree to which they sample the complete range of transit times. Even after 100 years of model integration, only those locations with an age of less than 10 years (such as the Indian Ocean site in Fig. 4) appear to be have complete TTDs. The age-trend threshold is defined as a trend of 1 year of age increase per decade of simulation. The stipled regions in Fig. 5 show areas that meet this criterion. The same general patterns emerge using the normalization threshold and age-trend threshold metrics.

4.3 Spatial distribution of the first moment

Figure 7 shows the ensemble average, standard deviation, and relative standard deviation of mean age at 268 m for the three-member (i.e., each longer than a century) ensemble after 100 years of model integration. The standard deviation of mean age is typically only a few years at this depth, resulting in relative variation of only a few percent over most of the globe. In the regions where the age is close to its asymptotic value (Fig. 5), the standard deviation of age tends to be very small, indicating that the average BIR age is likely to be a good approximation to the TTD age in these regions.

There are several areas that stand out in Fig. 7 that indicate where the three member ensemble average BIR may not be a reliable estimate of the TTD. For example, in the south-eastern Pacific, the relatively high variability in the standard-deviation of mean age is due to BIR3 having a somewhat higher amplitude and wider distribution than either BIR1 or BIR2, and thus a greater age. This could be the result of an additional water mass contribution at this location, but without more flow pathway information (from regional BIRs or lagrangian particles, for example) it is difficult to be conclusive. The Weddell Sea, Sea of Japan, and Baffin Bay are all showing a sensitivity to the depth and timing of convection.

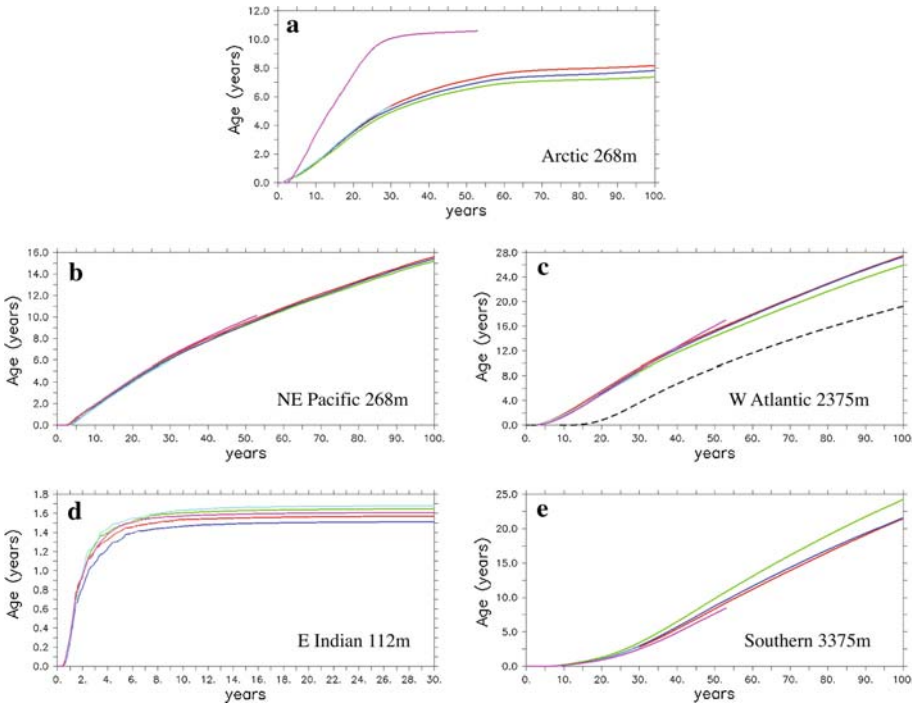


Fig. 6 Time series of the first moment ($Age(\mathbf{x}, \tau)$) of global BIR1 through BIR5, at the same five locations as in Fig. 4. Again, the individual realizations are denoted by *red*, *green*, *blue*, *cyan*, and *magenta*, respectively, and the five-member ensemble average by the *thick black line*. The *dashed curve* in **c** is from the 3° simulation of PM06

Figure 8a–c shows meridional sections of the mean age derived from the three-member ensemble average BIRs, again with the standard deviation and relative variation, at 55°W in the North Atlantic. The youngest water is found above the thermocline and in Labrador Current water making its way around the Grand Banks at the northern end of the section. The average BIR integral contours in Fig. 8 clearly show the more rapidly ventilated (younger) NADW in both the north and south centered around 2,500 m. Even though the deepest water (below 4,000 m) appears to be younger than at mid-depths, the BIR here is incomplete, so it is very likely that it would become the oldest water if the simulation was run for a substantially longer time. The variability along this section shows a relative minimum centered around 1,000 m associated with low salinity intermediate water from the south.

In the Pacific along 155°W (Fig. 8d), the youngest water is again found in the upper thermocline. The apparently low age water that occupies much of the range between 1,000 m and the bottom is actually a result of the fact that no signal has reached these depths after 100 years, consistent with the absence of large scale deep water formation in the north Pacific. The signature of AABW can clearly be seen originating from the south and flowing into the middle northern latitudes in the abyss.

One important question is that of whether increased ensemble size leads to an increase in our confidence in the statistics. Similar statistics to those in Fig. 7 may be obtained using the five-member ensemble, though due to the shorter duration it is necessary to move higher up in the water column to find BIRs that are complete. It was found that, at 112 m depth, after

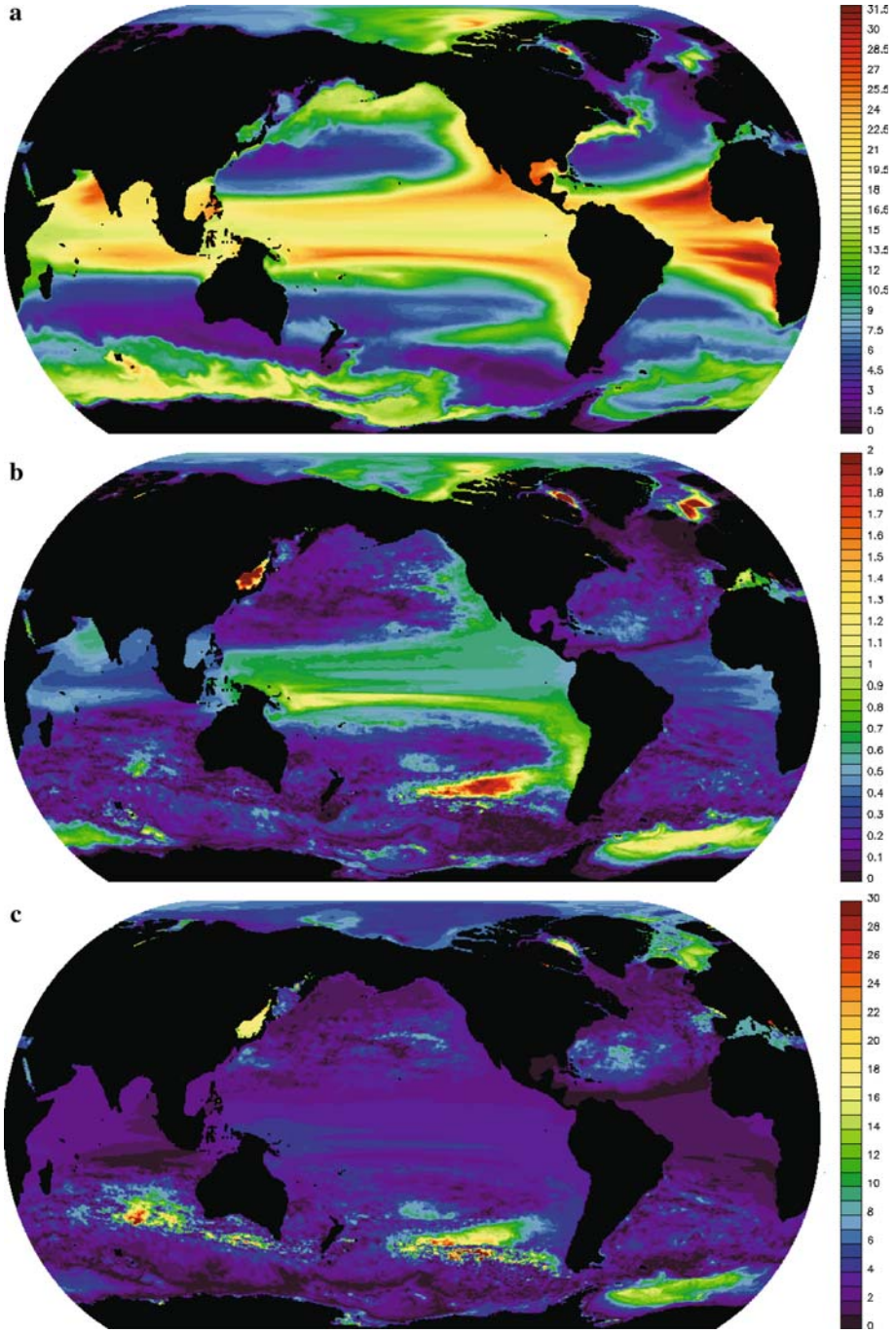


Fig. 7 **a** Ensemble average age (years), **b** standard deviation of age (years), and **c** relative standard deviation (standard deviation divided by the average, in percent) calculated from global BIR1, BIR2, and BIR3 at 268 m after 100 years

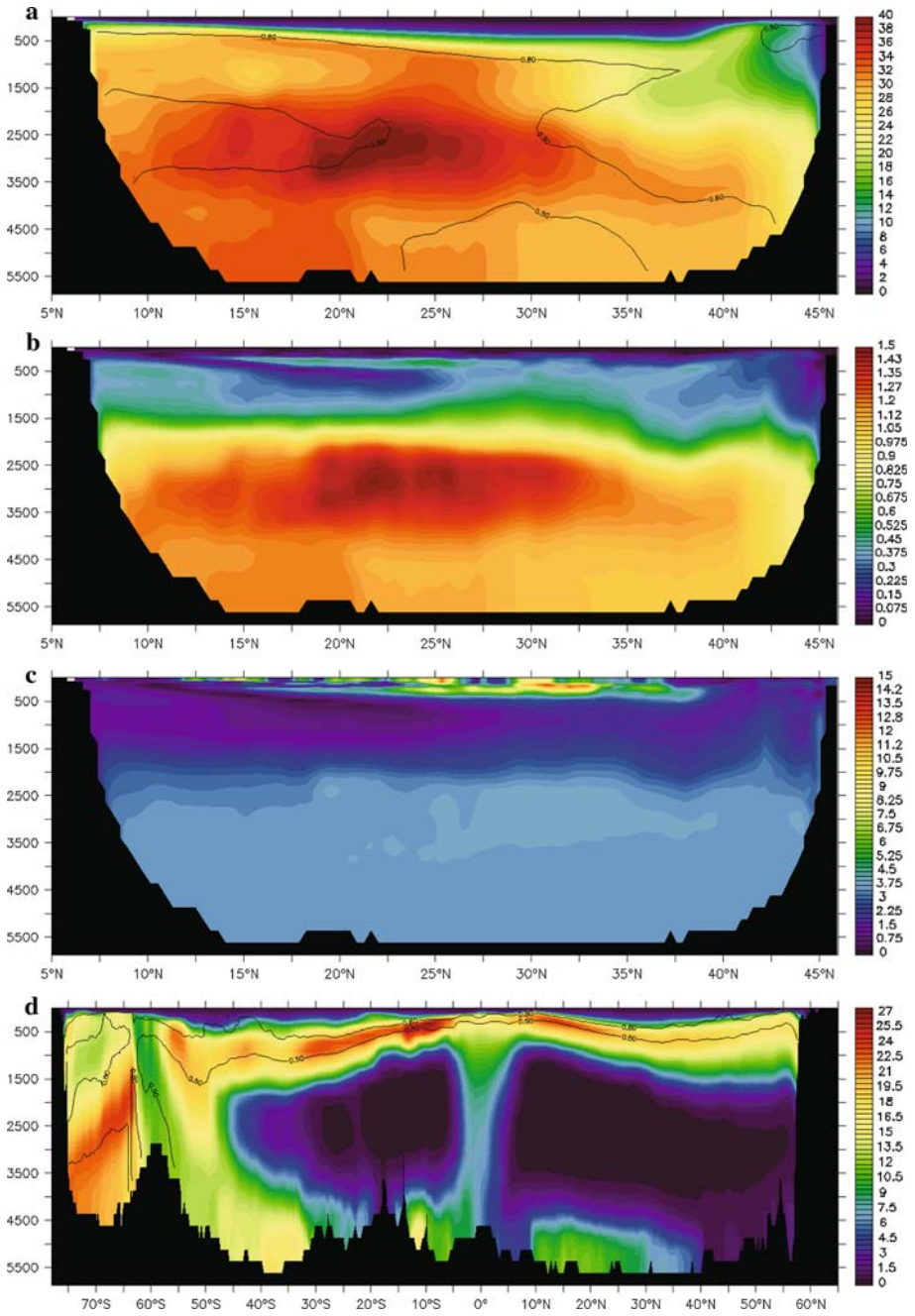


Fig. 8 **a** Ensemble average age (years), **b** standard deviation of age (years), and **c** relative standard deviation (percent) calculated from global BIR1, BIR2, and BIR3 at 55 W after 100 years. The average BIR integral is contoured with values of 0.5 and 0.8 in **a** and **d** to indicate which areas are relatively close to convergence. **d** Ensemble average age (years) at 150 W with same overlying contours as in **a**

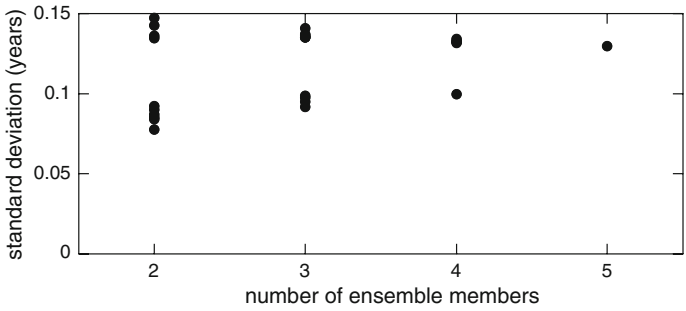


Fig. 9 Global area-averaged ensemble standard deviation of age (in years) at 112 m after 30 years using all possible unique combinations of between two and five ensemble members. The clusters of points with higher values each contain member BIR5 (see text)

30 years, the ensemble average, standard deviation, and relative standard deviation of age computed using five ensemble members was very similar to that seen in Fig. 7, with lowest mean age in the subtropical gyres and convection regions, and relative standard deviation of only a few percent in most areas.

Figure 9 shows the global area-averaged ensemble standard deviation of mean-age at 112 m as a function of ensemble size, using all possible combinations of ensemble members. For example, the ensemble size of three on the x-axis includes the ten possible unique combinations from (BIR1, BIR2, BIR3), (BIR1, BIR2, BIR4), etc. Since the spread in values decreases with the ensemble size, we conclude that a larger ensemble *does* improve our confidence, but that even an ensemble size of three can give a fairly good estimate of the statistical behavior, at least for this definition of a metric. It is worth noting that there are two distinct groupings of points in Fig. 9 for each ensemble size. The higher values all include contributions from BIR5 which has anomalously large departures from the ensemble mean in the Arctic (Fig. 4). If the Arctic is excluded from the calculation, the higher values drop by about 10%. This is an indication that the ideal of statistical stationarity of the flow field was not completely achieved over multi-decadal timescales.

4.4 Regional BIRs

As noted in Sect. 3.2, the simulation also included six regional BIRs for which the surface region Ω was not the entire ocean surface, but a specific area of interest (Fig. 10a). Regional BIRs can further elucidate the water mass properties at a given location by quantifying the relative importance of the source region on the parcel composition. These regional BIRs were simulated for 20 years, and only a single realization was performed for each source region.

Figure 10b–d shows the percentage of the global BIR3 integral (4) due to contributions from the regional BIRs. The global BIR3 integral is used as a reference since it was initiated at the same time as the regional BIRs (Fig. 1). By definition, the sum of all of the regional BIRs add up (within numerical error) to be identical to BIR3 in areas where there are no contributions from undefined regions, thus creating well-defined percentages. The colors map directly to the regions denoted in Fig. 10a (with yellow being the contribution from all regions *except* these), and the color saturation reflects the percentage. For example, Fig. 10b shows which source regions contribute the most to the total BIR3 integral after 20 years at 2,125 m depth. The NADW at this level is dominated (typically 70–90%) by contributions from the Labrador Sea (blue); only to the southeast of the Reyjknes Ridge is Irminger Sea

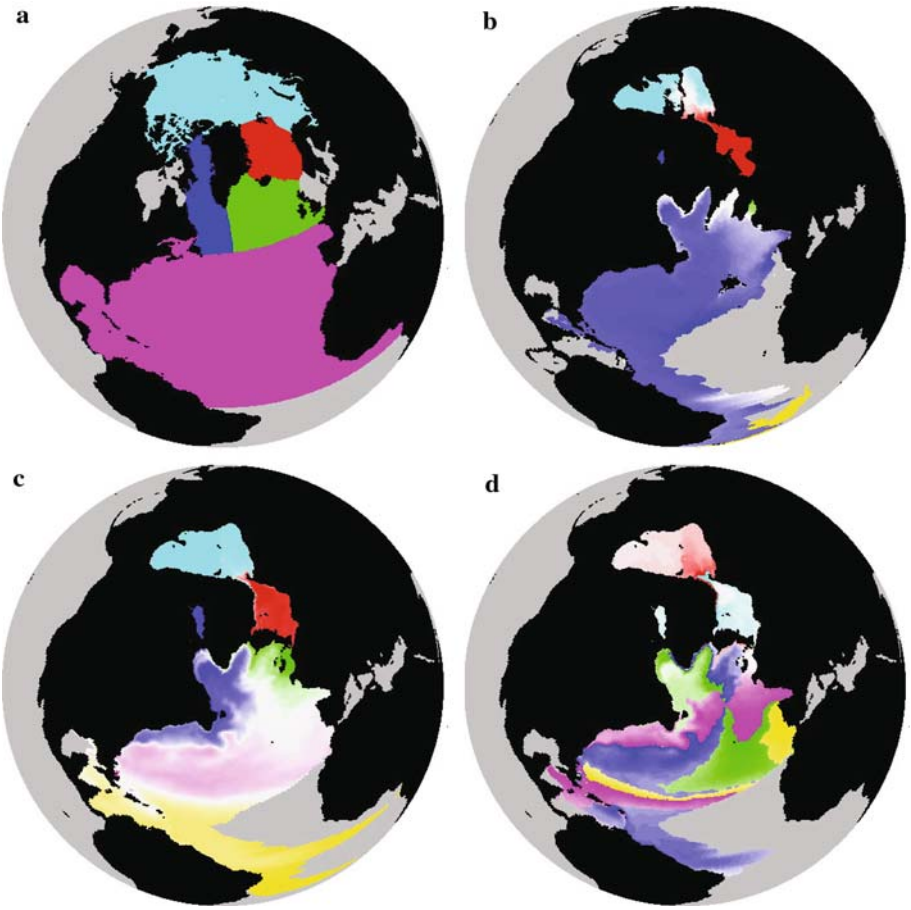


Fig. 10 **a** Locations of the surface ocean area which define the regional BIRs. Note that there is a sixth regional BIR that includes much of the Southern Ocean which is not considered here. **b** Primary contributions (color coded by the surface region seen in **a**) to the global BIR3 integral, $I(\mathbf{x})$, at 2,125 m depth after 20 years from each of the regional BIRs. The most saturated colors denote 100% of the total due to a particular regional BIR, and the least saturated (i.e., *white*) denote 50%. Areas that are grey have no BIR value, indicating that no ventilation has occurred yet. **c** Primary contributions to the global BIR3 integral at 918 m depth after 20 years from each of the regional BIRs. The most saturated colors denote 100% of the total, and the least saturated denote 40%. **d** Secondary contributions to the global BIR3 integral at 918 m depth after 20 years from each of the regional BIRs. The most saturated colors denote 40% of the total, and the least saturated denote 2%

water (green) the most plentiful component. If we consider which regional BIRs are the second most important in the total BIR3 integral, the patterns remain the same as in Fig. 10b, but the colors have switched (blue with green and cyan with red, not shown). The percentages have also dropped (since these are the secondary contributions), showing that Irminger Sea water accounts for 10–20% of the NADW at this depth. When considering the third most important contributors, we find a 5–15% contribution to the NADW from GIN Seas water (red, not shown).

At shallower depths (918 m), the primary contributions to BIR3 more directly correspond to the locations of the surface regions (Fig. 10c). In particular, the Subtropical water

(magenta) has been ventilated to the south of the Gulf Stream, with Labrador Sea water making its way south along the North American coast. Waters from the south (yellow) have penetrated through the equatorial current system into the Caribbean. The secondary contributions show a much more complex arrangement (Fig. 10d). Advection of Irminger Sea (green) and Subtropical (magenta) water through the gyre from the northeast can be seen. Labrador Sea water (blue) has been captured from the DWBC via the equatorial currents and is returning northward along the coast of South America. Mediterranean Sea outflow (yellow) can be seen clearly and is an important constituent (35%) in the eastern basin. The striped pattern to the east of Florida indicates the intersection of the different water masses, each associated with a range of densities, with this depth.

5 Discussion

In a highly turbulent flow, the use of ensemble averaging is fundamental. This is especially true here, where the statistics of the field we want to know (the TTD) are the same as for the field that is being simulated (the BIR). Due to the large computational burden imposed by the high horizontal resolution grid, a tradeoff between ensemble size and length of simulation was required a priori, with minor modifications being made during the run. Similarly, the length of the spinup run was chosen to be 15 years to allow for the maximum possible duration of the tracers. While this certainly is not enough time for the complete circulation to reach equilibrium, it is likely sufficient for the upper ocean processes, such as ventilation and formation of boundary undercurrents, that are of interest here (see, e.g., figure 1 of MM05).

Although performing statistics on an ensemble size of 5 (or 3) is far from ideal, it appears that we can obtain valuable information from them. Figure 9 suggests that any larger ensemble that could have practically been simulated (say, 10 or 15 members) would not have increased our confidence by a tremendous amount. Further, the 40-member simulations performed by Haine et al. [8] clearly show that a much smaller number of ensemble members are necessary at times sufficiently greater than the peak. On the other hand, Fig. 9 is only one metric at one model depth. The need to employ more statistical techniques, especially those specific to small ensemble sizes, is a path for future work.

As can be seen in Fig. 5, it appears that only a small percentage of the ocean has complete/converged BIRs. Figure 11 shows the depth dependence of convergence for both the TTD normalization and age-trend thresholds. These criteria are very consistent except at depth where the normalization threshold picks up the Atlantic DWBC, while the age-trend does not because the tails of the BIRs fall off relatively slowly with time (Fig. 4). Globally integrated, the TTD normalization threshold shows that 4.7% of the ocean volume has converged, while the age-trend threshold results in a somewhat lower value of 3.2% by the end of the 100 year simulation.

However, BIR convergence is likely to be too stringent a criterion for determining whether BIRs can be used as an estimate of the TTD. As seen in Fig. 4, the variability of the BIRs (both in time and among the ensemble members) decreases dramatically during their decay phase. It may then be possible to extrapolate the ensemble averaged BIR to much greater times, thus increasing the number of complete TTDs. For example, Fig. 12a shows the estimated TTD at a single location, showing that it exhibits a power law behavior for times following the peak of the distribution. Using this form in the first moment integral (Eq. 5), it is possible to estimate the asymptotic value of the age (Fig. 12b). This method likely provides an upper bound on age, since theory predicts that the BIRs decay exponentially at very large times [7], thus asymptoting faster than for any power law.

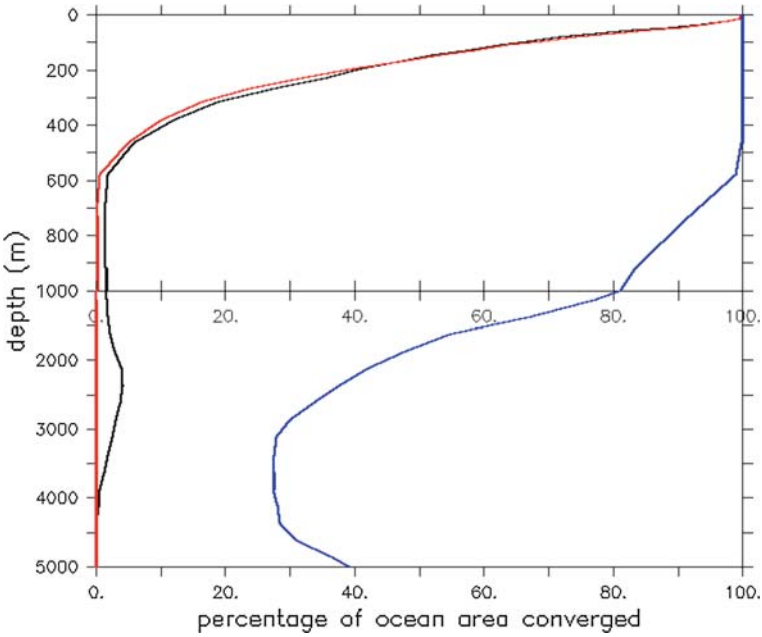


Fig. 11 Percentage of ocean area that has converged at each model depth using the normalization integral threshold (Eq. 4, *black*) and age-trend (Eq. 5, *red*). *Blue curve* shows the percentage of area at each depth that has passed the peak in the TTD by year 75

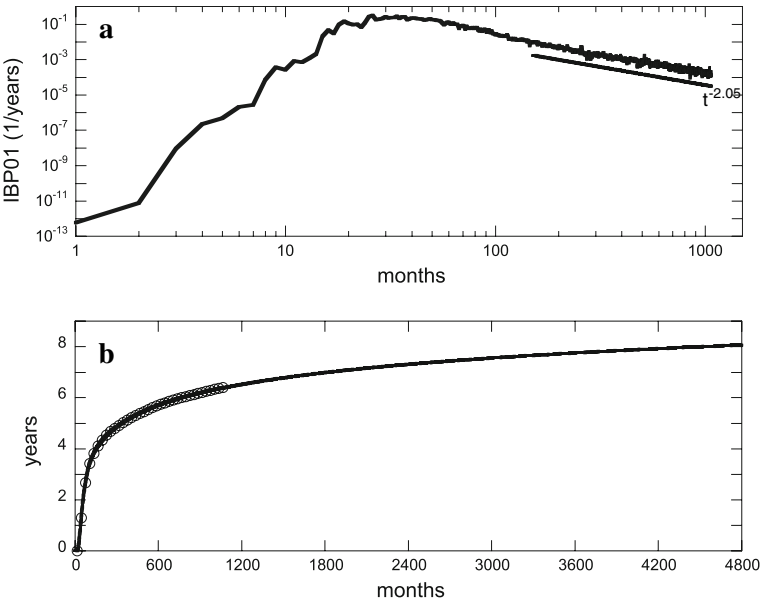


Fig. 12 **a** Time series of BIR1 at 21 N, 55 W, and 230 m depth in the subtropical Atlantic. The best least-squares fit to the slope for times greater than 12 years is −2.05. **b** First moment integral of the same BIR where *circles* denote the actual values and the *solid line* is the extrapolation based on the $t^{-2.05}$ power law fit

Although extrapolating all of the TTDs is beyond the scope of this paper, it is possible to estimate how many distributions may be amenable to extrapolation. The blue curve in Fig. 11 shows the percentage of TTDs whose peaks are reached by year 75, leaving 25 years of decay to guide the extrapolation. This results in a major increase in the volume of the ocean with potentially complete TTDs from 3.2 (using the age-trend threshold) to 50%. Note that essentially the entire ocean above 500 m has passed the peak of the TTD, and that 35–40% of even the deep ocean volume has achieved this threshold.

A second possible limitation of our experiments relates to the assumption of statistically stationary flow underlying the connection between the ensemble mean BIR and the TTD. While the simulation was forced with repeating annual cycle atmospheric conditions, and in many metrics (e.g. global mean kinetic energy) the flow is statistically stationary, there are signatures of low-frequency, multi-decaldal trends in some aspects of the circulation. These can be due to either model drift or intrinsic low frequency modes of variability arising in the turbulent solution. Evidence of this behavior was noted in the relatively large variability (as high as 40–50%) in the Arctic seen in Fig. 4. This may provide some insight into an upper bound, since it is due to a shift over the three decade period between the release of BIR4 and BIR5. Where the flow remains statistically steady, we would therefore expect the age variability to be much lower, perhaps on the order of 10%. We are planning a companion simulation forced with interannually varying atmospheric fields to quantify its effect.

6 Conclusions

We have presented results from a century-long simulation of the global ocean using a 1/10° grid that allows for spontaneous formation of mesoscale eddies. A suite of passive tracers, including an ensemble of five global BIRs were introduced in the simulation for the purpose of estimating distributions of water mass ventilation timescales. These provide the first estimate of the ocean TTD where the stirring and mixing effects of eddies have been accounted for directly rather than parameterized.

Consistent with the idealized basin study of HZWH08, one of our primary results is that the variance across the ensemble collapses dramatically for times following the peak in the TTD. At earlier times, significant variations due to chaotic processes can be seen in time series of the BIR for any given ensemble member (Fig. 4), but indications are that the modest size ensemble used here (five members for transit times less than 30 years) may be sufficient to obtain a robust estimate of the TTD. The TTD has important advantages over the individual BIRs, such as representing a true probability distribution function of transit times, and being able to be used in the convolution integral (Eq. 3) that propagates surface values of a tracer into the interior.

We also have found that the first moment (or age) of the BIRs is typically dominated by the long tail of the distribution where the variability is relatively small, so in those regions that have reached convergence the variations in the estimated age are typically only a few percent of the mean. Further, it appears that the long-time behavior of the BIRs may be amenable to extrapolation, resulting in even a more accurate assessment of water mass age.

While the variability across the ensemble arising from eddies is relatively small for intermediate and long transit times, there are fundamental differences in the character of the TTDs obtained here as compared to coarser resolution simulations. This is an indication of the deficiencies in the representation of sub-gridscale transport processes in the coarse resolution models. This result suggests that the TTD obtained from this simulation may provide a useful metric against which to judge eddy mixing parameterizations used in climate models.

Acknowledgements M. M. was supported by the Department of Energy Office of Science Climate Change Prediction Program. Participation of F. B. and S. P. was supported by the National Science Foundation by its sponsorship of the National Center for Atmospheric Research. The simulation was performed at the National Center for Computational Sciences at Oak Ridge National Laboratory with computer time awarded under the INCITE program, and at the National Center for Atmospheric Research Computational and Information Systems Laboratory.

References

1. Adcroft A, Hill C, Marshall J (1997) Representation of topography by shaved cells in a height coordinate ocean model. Mon Weather Rev 125:2293–2315
2. Adcroft A, Campin J-M, Heimbach P, Hill C, Marshall J (2005) MITgcm manual. http://mitgcm.org/sealion/online_documents/manual.html
3. Dukowicz JK, Smith RD (1994) Implicit free-surface method for the Bryan–Cox–Semtner ocean model. J Geophys Res 99:7991–8014
4. England M (1995) The age of water and ventilation timescales in a global ocean model. J Phys Oceanogr 25:2756–2777
5. Fekete BM, Vörösmarty CJ, Grabs W (2000) Global, composite runoff fields based on observed river discharge and simulated water balances. Technical Report, Global Runoff Data Centre.
6. Gouretski VV, Koltermann KP (2004) Woce global hydrographic climatology. A Technical Report. Tech. Rep. 35, Berichte des Bundesamtes für Seeschiffahrt un Hydrographi
7. Haine T, Hall T (2002) A generalized transport theory: water-mass composition and age. J Phys Oceanogr 32:1932–1946
8. Haine T, Zhang H, Waugh DW, Holzer M (2008) On transit-time distributions in unsteady circulation models. Ocean Model 21:35–45
9. Hall T, Haine T (2002) On ocean transport diagnostics: the idealized age tracer and the age spectrum. J Phys Oceanogr 32:1987–1991
10. Holzer M, Hall T (2000) Transit-time and tracer-age distributions in geophysical flows. J Atmos Sci 57:3539–3558
11. Holzer M, Primeau F (2006) The diffusive ocean converyor. Geophys Rev Lett 33:L14618
12. Holzer M, Primeau F (2008) The path density distribution of ocean surface-to-surface transport. J Geophys Res 113:C01018
13. Hurrell JW, Hack JJ, Shea D, Caron JM, Rosinski J (2008) A new sea surface temperature and sea ice boundary data set for the community atmosphere model. J Clim 21:5145–5153
14. Khatiwala S, Visbeck M, Schlosser P (2001) Age tracers in an ocean GCM. Deep Sea Res 48:1423–1441
15. Large WG, Pond S (1982) Sensible and latent heat flux measurements over the oceans. J Phys Oceanogr 12:464–482
16. Large WG, McWilliams JC, Doney SC (1994) Oceanic vertical mixing–a review and a model with a nonlocal boundary layer parameterization. Rev Geophys 32:363–403
17. Large WG, Yeager SG (2004) Diurnal to decadal global forcing for ocean and sea-ice models: the datasets and flux climatologies. NCAR Technical Note TN-460+STR, National Center for Atmospheric Research
18. Maltrud ME, McClean JL (2005) An eddy-resolving global 1/10° ocean simulation. Ocean Model 8: 31–54
19. Maltrud ME, Smith RD, Semtner AJ, Malone RC (1998) Global eddy-resolving ocean simulations driven by 1985–1995 atmospheric fields. J Geophys Res 103:30825–30853
20. Maltrud ME, Peacock SL, Bryan FO (2009) A century-long global eddying ocean simulation (in preparation)
21. McClean JL, Maltrud ME, Bryan FO (2006) Measures of the fidelity of eddying ocean models. Oceanography 19:104–117
22. McDougall T, Jackett D, Wright D, Feistel R (2003) Accurate and computationally efficient algorithms for potential temperature and density of seawater. J Atmos Ocean Technol 20:730–741
23. Murray RJ (1996) Explicit generation of orthogonal grids for ocean models. J Comput Phys 126:251–273

24. Peacock S, Maltrud ME (2006) Transit-time distributions in a global ocean model. J Phys Oceanogr 36:474–495
25. Primeau F (2005) Characterizing transport between the surface mixed layer and the ocean interior with a forward and adjoint global ocean transport model. J Phys Oceanogr 35:545–564
26. Waugh D, Hall T, Haine T (2003) Relationships among tracer ages. J Geophys Res 108:1–16