

Projection of time series with periodicity on a sphere

Victor Onclinx^{1,2}, Michel Verleysen¹ and Vincent Wertz^{1,2 *}

1- Université catholique de Louvain - Machine Learning Group
Place du Levant, 3, 1348 Louvain-la-Neuve - Belgium

2- Université catholique de Louvain - Department of Applied Mathematics
Avenue Georges Lemaître, 4, 1348 Louvain-la-Neuve - Belgium

Abstract. Predicting time series necessitates choosing adequate regressors. For this purpose, prior knowledge of the data is required. By projecting the series on a low-dimensional space, the visualization of the regressors helps to extract relevant information. However, when the series includes some periodicity, the structure of the time series is better projected on a sphere than on an Euclidean space. This paper shows how to project time series regressors on a sphere. A user-defined parameter is introduced in a pairwise distance criterion to control the trade-off between trustworthiness and continuity. Moreover, the theory of optimization on manifolds is used to minimize this criterion on a sphere.

1 Introduction

Time series forecasting is an important topic in many application domains. Conceptually, traditional methods [1, 2, 3] use the past values of a time series to predict future ones; these methods fit a linear or a nonlinear model between the vectors that gather the past values of the series, the regressors, and the values that have to be predicted. Note that exogenous variables and prediction errors may be used as inputs to the model too.

A first difficulty encountered by these methods is the choice of a suitable regressor size. Indeed, the regressors have to contain the useful information to allow a good prediction [4]. If the regressor size is too small, the information contained in the vector yields a poor prediction. Conversely, with oversized regressors, there can be redundancies such that the methods will overfit and predict the noise of the series.

For this reason and many other ones, including the choice of the model itself, it is useful to visualize the data (here the regressors) for a preliminary understanding before using them for prediction. This can be achieved by data projection methods [5, 6, 7, 8] which are aimed at representing high-dimensional data in a lower dimensional space. The projection of the regressors makes, for example, easier the visualization of some peculiarity in the time series.

*V. Onclinx is funded by a grant from the Belgium F.R.I.A. Part of this work presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its author(s). The authors thank Prof. Pierre-Antoine Absil for his suggestions on the theory of optimization on manifolds.

Moreover, assuming that data projection methods minimize the loss of information between the initial regressors and the projected ones, the forecasting of a time series can be achieved by using the projected regressors instead of the original ones, expecting that the smoothing resulting from the projection will help increasing the prediction performance.

In a first step, oversized regressors are projected to remove their potential redundancies and to reduce the noise. Most distance-based projection methods define the loss of information by the preservation of the pairwise distances. However, projection methods have to deal with a trade-off between trustworthiness and continuity [9], respectively the risk of flattening and tearing the projection. To control these types of behaviour, a user-defined parameter is introduced in the criterion [10] that implements the trade-off and that allows its control.

Furthermore, when time series have a periodic behaviour, it is difficult to embed them in an Euclidean space because of their complex structure [11]. Indeed, let us assume that the oversized regressors are lying close to an unknown manifold embedded in a high-dimensional space. Since the series is periodic, the manifold probably intercepts itself. In this context, the choice of a suitable projection manifold is motivated by its ability to keep the loops observed in the original space; the quality of the projection relies on its ability to preserve the global topology underlying the data distribution. The constraint of preserving loops is widely used in the context of topology-based projection methods, as the Self-organizing maps, where spheres [12, 13] and tori [14] are often used as projection manifolds; this paper presents a distance-based projection method on a sphere, a manifold that allows loops in the projection space.

The projection is achieved by the minimization of the pairwise distance criterion presented in Section 2. Since the projection space is non-Euclidean, Section 3 presents an adequate optimization procedure. Next to a brief introduction of the theory of optimization on manifolds [15], the theory is adapted to project data on a sphere. The projection of a sea temperature series on a sphere is presented in Section 4.1.

In order to take into consideration the advantages of the projection on manifolds, the forecasting methods should be adapted such that the prediction of time series can be based on the projected regressors. Section 4.2 is dedicated to the prediction of time series. By projecting the regressors on a sphere, a new *projected* time series is defined on the sphere; this series can easily be predicted using the Optimal-Pruned Learning Machine method [16]. Following these first results, the original time series is predicted with the projected regressors; the results of the forecasting are compared with the prediction of the series based on a 52-dimensional oversized regressors.

2 Projection criterion

This section aims at defining a projection criterion. As previously mentioned, data projection methods have to deal with a trade-off between trustworthiness and continuity. Two illustrative examples of the projection of a cylinder on \mathbb{R}^2

comment the trustworthiness and the continuity of a projection. Having in mind the compromise to reach these two objectives, a pairwise criterion can then be defined without restriction on the structure of the manifold.

Assuming that data close to a cylinder must be projected on the two-dimensional Euclidean space, a first option is to cut the cylinder along a generating line and to unfold it on the \mathbb{R}^2 Euclidean space. The resulting projection is trustworthy since two data that are close in the projected space (\mathbb{R}^2) are also close in the original space (the cylinder). However, because the cylinder has been torn, the projection cannot be continuous.

A second option is to flatten the cylinder to preserve the continuity. Actually, two data that are close in the original space, the cylinder, remain close in the projected one; the projection is thus continuous. Nevertheless, this projection is no more trustworthy since data coming from opposite part of the cylinder may be projected close from each other.

By counting the points that are close in one space but not in the other space, the trustworthiness and the continuity quality measures [9] are intuitively defined. Nevertheless, these measures are discrete and the optimization of these criteria is therefore difficult. To bypass this problem, distance-based projection methods minimize some weighted mean square errors between the original distance D_{ij} and the distance δ_{ij} on the projection manifold; the distances D_{ij} and δ_{ij} are defined between points i and j in their corresponding space with $1 \leq i, j \leq N$, N being the number of data.

The minimization of the unweighted cost function

$$f \equiv \sum_{i=1}^{N-1} \sum_{j>i}^N (D_{ij} - \delta_{ij})^2$$

cannot yield good results since large distances increase the cost function. In the projection context, this situation is against the intuition; one prefers to preserve the pairwise distances between close data rather than minimizing f .

By dividing each term of the cost function by the original distance D_{ij} , the minimization of the tearing error favours the continuity of the projection. Indeed, if it happens that two original data are close despite they are faraway in the projected space, they will dominate. Therefore, the minimization of the following cost function tends to make these data closer in the projected space:

$$\text{Tearing error} \equiv \sum_{i=1}^{N-1} \sum_{j>i}^N \frac{(D_{ij} - \delta_{ij})^2}{D_{ij}}.$$

Conversely, by weighting each term with the corresponding distance δ_{ij} in the projected space, the trustworthiness of the projection is favoured:

$$\text{Flattening error} \equiv \sum_{i=1}^{N-1} \sum_{j>i}^N \frac{(D_{ij} - \delta_{ij})^2}{\delta_{ij}}.$$

The flattening error expresses that points that are close in the projected space while they are not in the original space (small δ_{ij} and large D_{ij}) have to move faraway from each other during the optimization procedure.

Finally, to implement a trade-off between the trustworthiness and the continuity, a user-defined parameter $\lambda \in [0, 1]$ is introduced:

$$f \equiv \sum_{i=1}^{N-1} \sum_{j>i}^N \left(\lambda \frac{(D_{ij} - \delta_{ij})^2}{D_{ij}} + (1 - \lambda) \frac{(D_{ij} - \delta_{ij})^2}{\delta_{ij}} \right). \quad (1)$$

3 Optimization on manifolds

This section shows how to minimize the pairwise distance criterion (1). Because the projected points have to lie on a manifold, traditional optimization procedures cannot be used; the theory of optimization on manifolds proposes a powerful alternative. After an introduction to the topics from the theory of optimization on manifolds, adaptations to project data on a sphere are presented.

One could argue that to perform an optimization while keeping the projected points on a sphere, it is possible to perform a standard optimization in the spherical coordinate space. Unfortunately, this is not true since there are singularities in the two poles of the sphere. Actually, these two points are represented by two segments in the spherical coordinate space. Moreover, because the search space is limited to $\{(\phi, \theta) \in [0, 2\pi[\times [\frac{-\pi}{2}, \frac{\pi}{2}]\}$ and because it is not an Euclidean space anymore, traditional optimization methods cannot be used.

To circumvent these difficulties, the theory of optimization on manifolds proposes to consider the problem as an unconstrained minimization problem but by taking in mind that each point has to stay on the manifold all along the optimization procedure [15].

Working on a manifold does not allow movements through straight lines, as it is the case in the steepest descent gradient method; the curves of the manifold can however replace these straight directions since they include the curvature of the manifold and its global topology.

Searching for a minimum of a cost function f can be achieved by adapted line-search algorithm. Let us assume that the algorithm has successfully performed the k first iterations and that it has found the vector $\mathbf{y}(k) = (\mathbf{y}_1(k), \dots, \mathbf{y}_N(k))$ where $\mathbf{y}_i(k)$ is the location of data i on the projection manifold after iteration k . Moreover, let us denote the vector $\nu(k)$ that gathers the parameters of the manifold; since the optimal projection manifold cannot be determined *a priori*, this vector has to be optimized too. For example, in the case of the sphere, $\nu(k)$ will denote the radius of the sphere (which is unknown *a priori*).

First the gradient $-\nabla f(\mathbf{y}_1(k), \dots, \mathbf{y}_N(k), \nu(k))$ is evaluated. Nevertheless, this direction may point faraway from the manifold. To take into consideration the manifold constraint and its curvature, the gradient $-\nabla f$ is projected on the tangent space $\mathcal{T}_{\mathbf{y}}\mathcal{M}$. In this way, the new direction $-\nabla' f(\mathbf{y}_1(k), \dots, \mathbf{y}_N(k), \nu(k))$ is tangent to some curve $\gamma : \mathbb{R} \mapsto \mathcal{M} : t \mapsto \gamma(t)$ and therefore close to the manifold.

By searching in this direction with a step size α , a new location $\mathbf{y}'(k)$ can be found on the tangent space $\mathcal{T}_{\mathbf{y}}\mathcal{M}$. However, this location is not *on* the manifold; it has then to be retracted on the latter. The retraction, which is a kind of deterministic projection from the tangent space to the manifold, has to be chosen such that the new candidate location $\mathbf{y}(k+1)$ belongs to the curve γ determined by the direction $-\nabla' f$. The step size α is chosen under the Armijo condition [15] that ensures a sufficient decrease of the cost function. This means that the decrease of the cost function must be larger than the expected decrease of the first order approximation of the cost function f with a smaller step size $\sigma\alpha$ where $\sigma \in [0, 1]$. In other words, if the Armijo condition

$$f(\mathbf{y}(k)) - f(\mathbf{y}(k+1)) \geq \sigma\alpha\|\nabla' f\|^2 \quad (2)$$

is satisfied, the cost function has sufficiently decreased.

For details of the propose line-search algorithm see [15]. Fig. 1 shows the different steps of a single iteration.

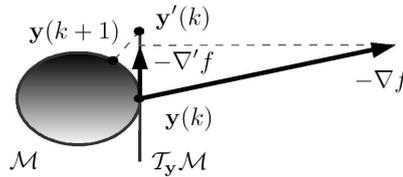


Fig. 1: Optimization iteration

After this brief introduction to the theory of optimization on manifolds, the latter is adapted to the problem of minimizing criterion (1) on a sphere. First, one has to define the manifold \mathcal{M} and the tangent space $\mathcal{T}_{\mathbf{y}}\mathcal{M}$. In addition to the spherical form of the manifold, one has also to add its radius R . The value of the radius is a scaling factor; this means that the radius R is considered as a parameter of the manifold because the adequate sphere is not known *a priori*. As each vector on the sphere has to have the same norm, the definition of the manifold can be expressed by:

$$\mathcal{M} \equiv \{(\mathbf{y}_1, \dots, \mathbf{y}_N, R) \in S_R^3 \times \dots \times S_R^3 \times \mathbb{R}^+ | \mathbf{y}_i^T \mathbf{y}_i - R^2 = 0, 1 \leq i \leq N\}.$$

By differentiating the set of constraints, the tangent space $\mathcal{T}_{\mathbf{y}}\mathcal{M}$ is defined by:

$$\mathcal{T}_{\mathbf{y}}\mathcal{M} \equiv \{(\mathbf{u}_1, \dots, \mathbf{u}_N, u_R) \in \mathbb{R}^3 \times \dots \times \mathbb{R}^3 \times \mathbb{R} | \mathbf{y}_i^T \mathbf{u}_i - Ru_R = 0, 1 \leq i \leq N\}.$$

Finally, if the angle between the vectors \mathbf{y}_i and \mathbf{y}_j is known, the product between the radius and this angle defines the distance between \mathbf{y}_i and \mathbf{y}_j . In order to evaluate this angle, the geodesic distance between \mathbf{y}_i and \mathbf{y}_j on the sphere is defined by the expression $\delta_{ij} \equiv R \arccos \frac{\mathbf{y}_i^T \mathbf{y}_j}{\|\mathbf{y}_i\| \|\mathbf{y}_j\|}$. Concerning the distance in the high-dimensional space, the geodesic distance is approximated

by the construction of a graph through the data where the edges are weighted by the Euclidean distances. The distance D_{ij} is evaluated by a shortest path algorithm [17, 18] such as Dijkstra's one. At the end, the evaluation of the gradient $-\nabla f$ is defined by the partial derivatives with respect to the locations y_i and the radius R .

4 Experiments

In this section, the data projection method is illustrated on the ESTSP2007 competition dataset of the weekly evolution of the sea temperature. The series is represented in Fig. 2 where the colour varies with the temperature. The series contains 875 temperature measures; a yearly periodicity can easily be observed.

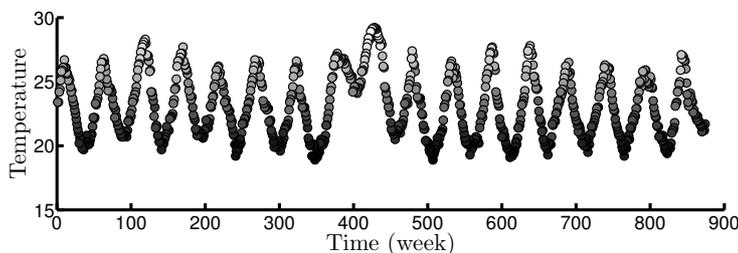


Fig. 2: Weekly evolution of the sea temperature

The methodology to forecast a periodic time series, as proposed in this paper, begins by building oversized regressors. The size of the regressors is chosen experimentally with respect to the length of a single period: 52-dimensional oversized regressors are built. Even if they probably contain all useful information for the prediction, these regressors are noisy and they certainly contain redundancies. The regressors are thus projected on a sphere according to the above methodology. The forecasting of the time series is, at the end, based on the projected regressors.

Section 4.1 shows the results of the projection; hence, the projected regressors define a curve on the optimal sphere. Section 4.2 first studies the forecasting of this new time series on the sphere to show the accuracy of the projection and of the methodology. Finally, the prediction of the original time series is performed and evaluated. Both the prediction of the projected time series on the sphere, and the prediction of the original time series based on the projected regressors, use the OPELM method [16].

4.1 Projection of the sea temperature series

The intrinsic dimension of the 52-dimensional oversized regressors is much lower than the embedding Euclidean space. For example, by projecting the data with Principal Component Analysis [19] in order to reduce the dimensionality to

the 10 principal components, the residual variance is less than 1 percent; this motivates the idea of projecting the regressors on a low-dimensional manifold.

The geodesic distance in the high-dimensional space D_{ij} is approximated by the shortest path in the graph built through the 50 closest neighbours [17, 18].



Fig. 3: 52-regressors projected on the sphere with $\lambda = 0.9$

The result of the projection on the sphere is shown in Fig. 3 where the colour varies smoothly with respect to the value of $\mathbf{y}(t)$. The colours used are the same as in Fig. 2; it can be easily seen that similar values of the original time series, thus similar colours, are close on the sphere. The additional curve in Fig. 3 joins points that are consecutive in time to illustrate the path of the projected time series on the manifold. The projected time series turns around the sphere such that the sphere keeps the periodicity of the time series. Furthermore, the isolated part of the projected data in the upper left region of the sphere in Fig. 3 corresponds to the irregularities of the time series observed between times $t = 380$ and $t = 420$ in Fig. 2.

In Fig. 4, the corresponding result in the spherical coordinate space is represented in order to visualize all the data; the *glyph* in the center of the figure corresponds to the above-mentioned irregularities. According to both Fig. 3 and 4, the projection of the times series makes it possible to isolate its irregularities in a visual way.

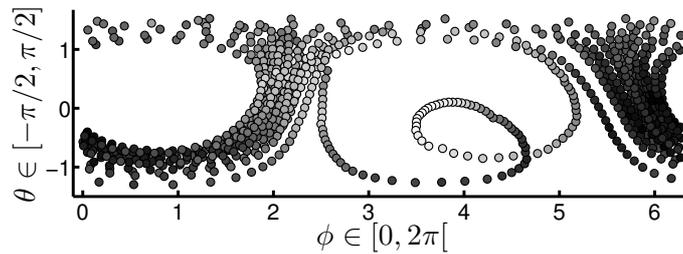


Fig. 4: 52-regressors projected on the sphere, in the spherical coordinate space

4.2 Prediction of the sea temperature series using the projected regressors

Besides the visualization applications, the projection of the time series defines new regressors where redundancies are removed and noise is probably reduced. This subsection shows how the projected regressors can be used.

Let us consider the projected time series defined by the locations $\mathbf{y}(t)$ on the sphere, with t between 1 and N . To test the quality of the projected time series, a model $\hat{\mathbf{y}}(t+1) = f(\mathbf{y}(t), \mathbf{y}(t-1), \theta)$ is built with the Optimal-Pruned Learning Machine method [16]. OPELM is a two-layer regression model, where the first layer is chosen randomly among a set of possible activation functions and kernels, and the second layer is optimized with linear tools. The speed of optimizing such models makes it possible to test a large number of them, among which the best according to some validation criterion is selected. θ represents the parameters of the method, more specifically the number and the types of kernels or functions; both Gaussian and sigmoidal functions are used. The learning and validation errors are estimated according to the following definitions:

$$\begin{aligned} \text{Learning error} &\equiv \frac{\sum_{t=1}^{N_1} \|\hat{\mathbf{y}}(t) - \mathbf{y}(t)\|^2}{N_1} \\ \text{Validation error} &\equiv \frac{\sum_{t=1}^{N_2} \|\hat{\mathbf{y}}(t) - \mathbf{y}(t)\|^2}{N_2}, \end{aligned}$$

where N_1 and N_2 represent respectively the size of the learning and of the validation sets. The learning set is randomly built with 66 percent of the initial set; 10000 simulations are performed in order to estimate the learning and the validation errors as average over all the 5000 experiments. The results are shown in Fig. 5 with respect to the number of kernels/functions used in the OPELM tool.

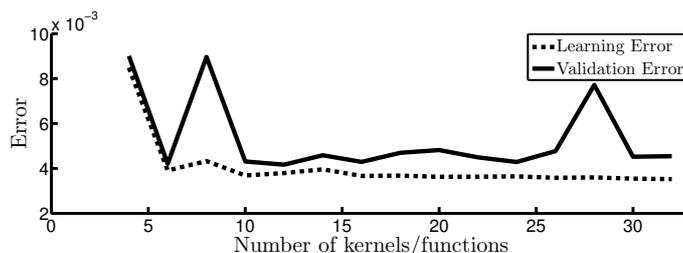


Fig. 5: Learning and validation errors of the normalized projected time series versus the number of kernels/functions used

Fig. 5 shows that the projected time series on the sphere can easily be predicted. However, this result does not mean that the original series can be easily predicted too. As a first attempt in this direction, we propose to build another prediction model based on the projected regressors. Assuming that the locations

$\mathbf{y}(t)$ on the sphere are known, they define reduced regressors such that it can be used to forecast the original time series $x(t)$. In [20], the authors define new regressors by concatenating the projected regressors with the corresponding value $x(t)$. Here, we use an alternative idea, which consists in predicting the variations in the time series using the projected regressors. The model is thus defined by:

$$x(t+1) = x(t) + \tilde{f}(\mathbf{y}(t), \theta). \quad (3)$$

The quality of the prediction is close to the forecasting with the 52-dimensional regressors as shown in Fig. 6. In this figure the learning error of the prediction based on the projected regressors is higher than the learning error based on the 52-dimensional initial regressors, but the validation error is lower when using the projection. This is likely to be due to overfitting of the model based on the 52-dimensional regressors.

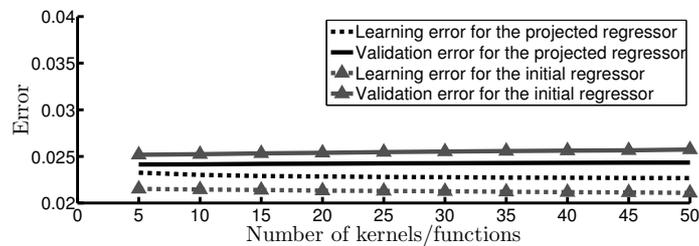


Fig. 6: Learning and validation errors for the prediction of the normalized time series with the initial regressors and the projection on the sphere

5 Conclusion

This paper presents a nonlinear method aimed at projecting the regressors of a time series on a sphere such that redundancies are removed and noise is reduced. The method minimizes a pairwise distance cost function where the trade-off between trustworthiness and continuity is controlled by a user-defined parameter. The projection on a sphere is aimed at embedding the periodicity of time series using a dedicated optimization method. The quality of the projection is assessed through the trustworthiness and the continuity quality measures and is compared to the same measures obtained after projecting on Euclidean spaces.

The projected regressors can be used to forecast the original time series. First results are shown using the OPELM algorithm. Nevertheless, the OPELM prediction method is not specifically adapted to spherical data for which the manifold contains another part of useful information. This will be studied in future work.

References

- [1] G.E.P. Box and G. Jenkins. *Time Series Analysis : Forecasting and Control*. Holden-Day, Incorporated, 1990.

- [2] L. Ljung. *System Identification, Theory for the user*. Prentice Hall Information and System Sciences Series, 1987.
- [3] C. Chatfield and A.S. Weigend. Time series prediction: Forecasting the future and understanding the past. *International Journal of Forecasting*, 10(1):161–163, June 1994.
- [4] F. Takens. On the numerical determination of the dimension of an attractor. In *Dynamical Systems and Bifurcations*. Groningen, 1984.
- [5] J.A. Lee and M. Verleysen. *Nonlinear Dimensionality Reduction*. Springer Science+Business Media, LLC, 2007.
- [6] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15(6):1373–1396, 2003.
- [7] A. Brun, C.-F. Westin, M. Herberthson, and H. Knutsson. Fast manifold learning based on riemannian normal coordinates. In Heikki Kälviäinen, Jussi Parkkinen, and Arto Kaarna, editors, *SCIA*, volume 3540 of *Lecture Notes in Computer Science*, pages 920–929. Springer, 2005.
- [8] J.A. Lee and M. Verleysen. Nonlinear projection with the isotop method. In J. R. Dorronsoro ed., editor, *Artificial Neural Networks*, Lecture Notes in Computer Science 2415, pages 933–938, London, UK, Augustus 2002. ICANN, Springer-Verlag.
- [9] J. Venna and S. Kaski. Neighborhood preservation in nonlinear projection methods: An experimental study. In *ICANN '01: Proceedings of the International Conference on Artificial Neural Networks*, pages 485–491, London, UK, August 21-25 2001. Springer-Verlag.
- [10] J. Venna and S. Kaski. Local multidimensional scaling with controlled tradeoff between trustworthiness and continuity. In *Proceedings of WSOM'05, 5th workshop on self-organizing maps*, pages 695–702. WSOM, September 5-8 2005.
- [11] V. Onclinx, V. Wertz, and M. Verleysen. Nonlinear data projection on a sphere with a controlled trade-off between trustworthiness and continuity. In *ESANN 2008, European Symposium on Artificial Neural Networks*, pages 43–48, Bruges (Belgium), April 23-25 2008. ESANN, d-side publi.
- [12] H. Ritter. Self-organizing maps on non-euclidean spaces. In S. Oja and E. Kaski, editors, *Kohonen Maps*, pages 97–108. Elsevier, Amsterdam, 1999.
- [13] H. Nishio, Md. Altaf-Ul-Amin, K. Kurokawa, K. Minato, and S. Kanaya. Spherical som with arbitrary number of neurons and measure of suitability. In *Proceedings of WSOM'05, 5th workshop on self-organizing maps*, pages 323–330, September 5-8 2005.
- [14] J.X. Li. Visualization of high-dimensional data with relational perspective map. *Information Visualization*, 3(1):49–59, 2004.
- [15] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, January 2008.
- [16] Y. Miche, P. Bas, C. Jutten, O. Simula, and A. Lendasse. A methodology for building regression models using extreme learning machine: Op-elm. In *European Symposium on Artificial Neural Networks (ESANN)*. d-side publi., April 23-25 2008.
- [17] J.A. Lee, A. Lendasse, and M. Verleysen. Curvilinear distance analysis versus isomap. In *ESANN 2002, European Symposium on Artificial Neural Networks*, pages 185–192, Bruges (Belgium), April 22-24 2002. ESANN, d-side publi.
- [18] J.A. Lee, A. Lendasse, N. Donckers, and M. Verleysen. A robust nonlinear projection method. In *ESANN 2000, European Symposium on Artificial Neural Networks*, pages 13–20, Bruges (Belgium), April 28-28 2000. ESANN, D-Facto public.
- [19] K. Pearson. Analysis of a complex statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.
- [20] A. Lendasse, J. Lee, V. Wertz, and M. Verleysen. Forecasting electricity consumption using nonlinear projection and self-organizing maps. *Neurocomputing*, 48, 2002.