ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

Non-negative decomposition of geophysical dynamics

Manuel López-Radcenco[1], Abdeldjalil Aïssa-El-Bey[1],
Pierre Ailliot[1,2], Ronan Fablet[1] *

1- Institut Mines-Télécom, Télécom Bretagne,
UMR CNRS 6285 Lab-STICC,
Technopôle Brest-Iroise CS83818, 29238 Brest Cedex 3, France

2- Université de Brest, UMR CNRS 6205 LMBA,
6, Avenue Victor Le Gorgeu, B.P. 809, 29285 Brest Cedex, France

Abstract. The decomposition of geophysical processes into relevant modes is a key issue for characterization, forecasting and reconstruction problems. The blind separation of contributions from different sources is a well-studied problem in signal and image processing. Recently, significant advances have been reported with the introduction of non-negative and sparse formulations. In this work, we address an extension to the blind decomposition of linear operators or transfer functions between variables of interest with an emphasis on a non-negative setting. As illustrated here, such decompositions are of key interest for the analysis of geophysical dynamics and the relationships between different geophysical variables.

1 Introduction

The decomposition of geophysical processes into relevant and more interpretable modes, most notably by orthogonal decomposition approaches such as empirical orthogonal functions (EOF) [1], has been used extensively in analysis and forecasting applications in oceanography and meteorology. Besides EOF-based decompositions, the blind separation of the contributions from different sources or processes has been extensively studied in signal and image processing [2]. Significant advancements have been reported with the introduction of non-negative [3] and sparse [4, 5] formulations. Here, we address the blind decomposition of spatially-varying or temporally-varying operators or transfer functions. Potential applications include the decomposition of dynamical processes as well as the analysis of the relationships between different variables of interest (See Section 3 for illustrations). In a previous work [6], we introduced a non-negative formulation that generalizes latent class regression models [7] and allows for the blind characterization of the relationships between two processes as the superimposition of multiple linear relationships. We further generalize this previous work and address a blind dictionary-based decomposition of local linear operators. This paper is organized as follows. Section 2 presents the proposed model and the associated calibration and estimation algorithms. In Section 3 we report applications to the analysis of geophysical dynamics. Finally, our concluding remarks and future work perspectives are presented in Section 4.

*This work was supported by ANR (Agence Nationale de la Recherche, grant ANR-13-MONU-0014) and Cominlabs Project SEACS.

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

2 Model formulation

2.1 Blind dictionary-based decomposition of linear operators

Let us consider a multivariate observation dataset $\{\mathbf{x}, \mathbf{y}\}_n$, where $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$ denote the n^{th} observation couple. Assuming that a linear operator relates $\mathbf{x}_n$ to $\mathbf{y}_n$, our goal is to decompose this linear operator as the non-negative superposition of multiple linear relationships. Following our previous work [6], we consider the following model:

$$\begin{aligned} \mathbf{y}_n &= \sum_{k=1}^{K} \alpha_{nk} \boldsymbol{\beta}_k \mathbf{x}_n + \omega_n \\ \text{Subject to } & \begin{cases} \alpha_{nk} \geq 0, & \forall\, k \in [\![1, K]\!],\, \forall n \in [\![1, N]\!] \\ ||\boldsymbol{\beta}_k||_F = 1, & \forall\, k \in [\![1, K]\!] \end{cases} \end{aligned} \tag{1}$$

where $\alpha_{nk} \in \mathbb{R}^+$ are non-negative mixing coefficients that model the contribution of each linear mode to the reconstruction of $\mathbf{y}_n$ given $\mathbf{x}_n$, $\boldsymbol{\beta}_k \in \mathbb{R}^{I \times J}$ is the regression matrix associated with mode k, $||\cdot||_F$ is the Frobenius norm and $\omega_n \in \mathbb{R}^I$ is a noise process, typically a Gaussian noise. N and K denote, respectively, the total number of observations and modes, while $k \in [\![1, K]\!]$ and $n \in [\![1, N]\!]$ indicate, respectively, the current mode and observation.
In [6] we focus on case-studies where I, the number of dimensions of variable $\mathbf{y}_n$, is large enough with respect to the number of modes to guarantee the identifiability of the decomposition. We here extend model (1) with no specific constraint on I and propose a blind dictionary-based approach. Let us denote by $\boldsymbol{\Theta}_n$ an estimation of the linear operator relating $\mathbf{x}_n$ and $\mathbf{y}_n$. Model (1) then resorts to the blind non-negative decomposition of linear operators $\boldsymbol{\Theta}_n$:

$$\begin{cases} \left[\hat{\alpha}_{nk}, \hat{\boldsymbol{\beta}}_k\right] = \underset{\alpha_{nk}, \boldsymbol{\beta}_k}{\text{argmin}} & \sum\limits_{n=1}^{N} \left(\left|\left| \boldsymbol{\Theta}_n - \sum\limits_{k=1}^{K} \alpha_{nk} \boldsymbol{\beta}_k \right|\right|_F^2 \right) \\ \alpha_{nk} \geq 0, & \forall n \in [\![1, N]\!], \forall k \in [\![1, K]\!] \\ ||\boldsymbol{\beta}_k||_2 = 1, & \forall k \in [\![1, K]\!] \end{cases} \tag{2}$$

By considering the set $\{\boldsymbol{\Theta}\}_n$ of all N local linear operators and vectorizing them, the constrained minimization problem in (2) can be rewritten as the following blind dictionary-based decomposition issue:

$$\begin{cases} \left[\hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{D}}\right] = \underset{\boldsymbol{\Gamma}, \boldsymbol{D}}{\text{argmin}} & ||\boldsymbol{\Phi} - \boldsymbol{D}\boldsymbol{\Gamma}||_F^2 \\ \boldsymbol{\Gamma}_{kn} \geq 0, & \forall k \in [\![1, K]\!], \forall n \in [\![1, N]\!] \\ ||\boldsymbol{D}_{:k}||_2 = 1, & \forall k \in [\![1, K]\!] \end{cases} \tag{3}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{IJ \times N}$ is the concatenation of vectorized operators $\boldsymbol{\theta_n^v} = vec(\boldsymbol{\Theta_n})$ (i.e. $\boldsymbol{\Phi} = [\boldsymbol{\theta_1^v}|...|\boldsymbol{\theta_N^v}]$), lines of $\boldsymbol{\Gamma} \in \mathbb{R}^{K \times N}$ contain mixing coefficients α_{nk} for each mode and columns of $\boldsymbol{D} \in \mathbb{R}^{IJ \times K}$ (noted as $\boldsymbol{D_{:k}}$) contain vectorized versions of modal linear regression matrices $\boldsymbol{\beta}_k$.

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1. Available from http://www.i6doc.com/en/.

2.2 Dictionary learning

Compared with [6], the dictionary-based formulation has greater flexibility and adaptability, as it allows for the use of any blind dictionary-based decomposition approach to solve constrained minimization (3) (e.g. NMF [3], K-SVD [8], etc). Hence, model constraints can be changed seamlessly simply by changing the considered blind dictionary-based decomposition approach. Here, for a given dataset of operators $\{\boldsymbol{\Theta}\}_n$, we solve for minimization (3) using a proximal splitting method to account for non-negativity constraints. It comes to iterate two steps until convergence, namely the least-square estimation of dictionary matrix $\boldsymbol{D}$ with normalization constraints $||\boldsymbol{D}_{:k}||_2 = 1$ and the estimation of the mixing coefficients $\boldsymbol{\Gamma}$ using a proximal operator [9] to enforce non-negativity.

2.3 Application to the decomposition of linear operators

Given a trained dictionary of operators $\{\hat{\boldsymbol{\beta}}\}_k$ (matrix $\hat{\boldsymbol{D}}$ in formulation (3)), we can apply the proposed decomposition to any new observation dataset $\{\mathbf{x}^*, \mathbf{y}^*\}_n$ to estimate the associated mixing coefficients $\hat{\alpha}_{nk}$ (matrix $\hat{\boldsymbol{\Gamma}}$ in formulation (3)). Two approaches may be considered. Similarly to the training step, we can first estimate the linear operators $\{\boldsymbol{\Theta}^*\}_n$ (see Section 3 for details on the estimation for each case-study) and estimate the mixing coefficients as the projection of these operators onto the trained dictionary with a non-negativity constraint. A second approach comes to the direct estimation of the mixing coefficients according to a least-squares criterion derived from model (1). This scheme does not require the prior estimation of linear operators $\{\boldsymbol{\Theta}^*\}_n$. Both approaches can exploit either proximal operators, as in the model training step, or classical non-negative least-squares solvers [10].

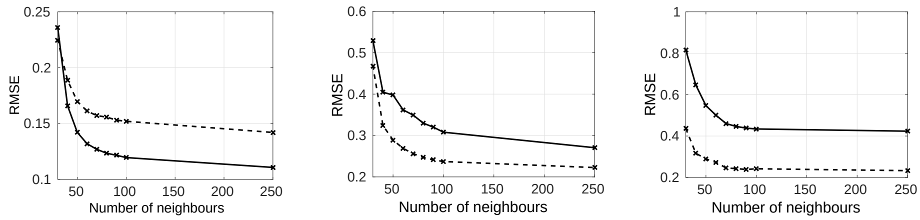
3 Application to geophysical dynamics

3.1 Application to analog forecasting

We first illustrate the interest of the proposed approach for the forecasting of dynamical systems. We apply the proposed blind non-negative decomposition to Lorenz '96 dynamics, which have been extensively studied in the assimilation and forecasting literature, since they are representative of chaotic geophysical dynamical systems (e.g., the atmosphere). We let the reader refer to [11] for a detailed presentation of Lorenz '96 model.

Here, we simulate Lorenz'96 40-dimensional time series with forcing parameter $F = 8$ and time step $\partial t = 0.05$. We build training and test datasets from independent time series corresponding respectively to 2×10^5 and 200 consecutive time steps. Following [12], we consider a locally-linear analog model. It comes to fitting a multivariate linear regression $\Theta_n \in \mathbb{R}^{I \times J} : \mathbf{y}_n = \Theta_n \mathbf{x}_n$, where, for given time series S and variable index $l^* \in [\![1, 40]\!]$, variable $\mathbf{x}_n$ is the 21-dimensional vector $\{S(t^*, l^* - 10), ..., S(t^*, l^* + 10)\}$ and variable $\mathbf{y}_n$ is the 3-dimensional vector $\{S(t^* + \partial t, l^* - 1), ..., S(t^* + \partial t, l^* + 1)\}$. Local linear operators Θ_n are

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.



(a) $N = 2 \times 10^5$. No noise (b) $N = 2 \times 10^5$. $\sigma^2_{noise} = 0.1$ (c) $N = 2 \times 10^3$. No noise

Fig. 1: Forecasting performance of linear analog operators: we depict the normalized forecasting RMSE vs. the number of analogs used to estimate the locally-linear analog forecasting operator for Lorenz '96 dynamics for: locally-linear operators (*Full line*), the proposed non-negative decomposition of locally-linear operators with $K = 4$ modes (*Dashed line*). We consider three experimental settings: a noise-free scenario with a large catalog of 2×10^5 exemplars, a noisy scenario with a large catalog of 2×10^5 exemplars, a noise-free scenario with a smaller catalog of 2×10^3 exemplars.

estimated for each observation pair n under the hypothesis that the nearest observation pairs in the training dataset, referred to as analogs, share the same Θ_n. Numerically, the estimation of Θ_n resorts to a weighted least-square estimate from the dataset formed by the analogs of pair $(\mathbf{y}_n, \mathbf{x}_n)$, the weights accounting for relative similarities.

We illustrate here the application of the proposed non-negative decomposition with $K = 4$ modes. Given the dictionary $\{\hat{\boldsymbol{\beta}}\}_k$ learnt from the training dataset, we state the analog forecasting operators of the test dataset according to model (2). It resorts to the estimation of mixing coefficients in model (2). More precisely, for the current observation pair $(\mathbf{y}_n^*, \mathbf{x}_n^*)$ in the test dataset, the associated analog forecasting operator involves the non-negative projection of the previously estimated locally-linear operator Θ_n^* onto the manifold spanned by the estimated modes $\{\hat{\boldsymbol{\beta}}\}_k$. The forecasting for the next time step simply amounts to the application of the projected linear operator to the current state.

Figure 1 illustrates the performance of the analog forecasting for different parameter settings with a focus on variable index $l^* = 20$ of the Lorenz'96 state. We report the normalized root mean square error (RMSE) of the forecasting as a function of the number of analogs used, for the locally-linear analog forecasting with no decomposition (full line) and the proposed analog forecasting using a non-negative decomposition (dashed line). Three scenarios are simulated: i) a noise-free scenario with a large catalog ($N = 2 \times 10^5$ exemplars) (Fig.1a), ii) a noisy scenario with a large catalog and noise variance $\sigma^2_{noise} = 0.1$ (Fig.1b), iii) a noise-free scenario with a small catalog ($N = 2 \times 10^3$ exemplars) (Fig.1c). The non-negative decomposition of local operators in an analog-based prediction scheme clearly reduces forecasting errors when the analogs are sampled from a noisy or reduced catalog. This decomposition can be seen as a projection of the original operator into a lower-dimensional space, what makes the identification of the model feasible for small datasets and improves robustness to noise.

3.2 Decomposition of upper-ocean dynamics

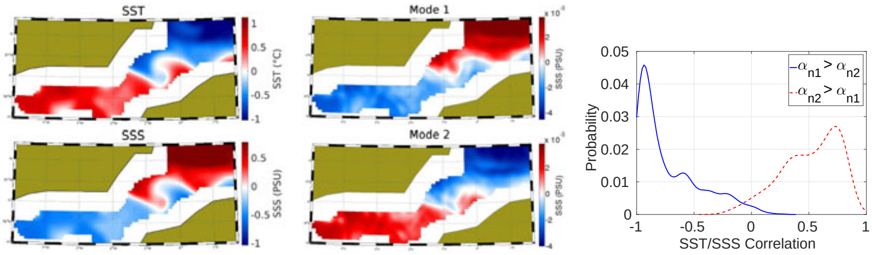
We illustrate a second application of the proposed non-negative decomposition to the characterization of upper ocean dynamics from the synergy exhibited by different sea surface geophysical fields, namely sea surface temperature (SST) and sea surface salinity (SSS). As illustrated in Figure 2a, we analyze the relationships between SST and SSS in the Alboran Sea ($35°N$ - $38°N$,$0°W$ - $5°W$). This region involves strong seasonal patterns associated with the intake of cold Atlantic water through the Gibraltar strait, which strongly affects the SST signature in the Alborean Sea and results in a shift from positive to negative correlations between SST and SSS fields. We expect the proposed non-negative decomposition to capture this seasonal patterns.
For the reported experiments, we exploit $1/16°$ operational ROMS simulations of WMOP model [13] from 2009 to 2012. We analyse the relationships between daily SST and SSS images (Figure 2a) using a convolutional model for 3×3 image patches, which comes to considering SSS pixel values can be approximated daily by a single shared linear regression Θ_n on values from the corresponding 3×3 SST patch. Daily estimated linear operators are decomposed into $K = 2$ modes using the proposed non-negative operator decomposition. Given the learnt dictionary, mixing coefficients are re-estimated directly from SST-SSS observations. Figure 2b presents the SSS fields predicted by each mode. The first mode clearly captures an inversion of the SST, while the second mode captures a sign-coherent SST-SSS relationship. This is further illustrated in Figure 2c by the SST-SSS correlation probability density functions computed independently, via Gaussian-kernel estimation, for dates when either one of the estimated modes are dominant. These results suggest that the proposed decomposition is capable of accurately separating the inverted and coherent SST-SSS relationships, thus providing a useful tool for the analysis of the physical processes behind them.

4 Conclusion

We extended the observation-based non-negative decomposition of linear operators introduced in [6] and explored applications to the characterization and forecasting of geophysical dynamics. The proposed formulation exploits a dictionary-based formulation from locally estimated linear operators and has increased flexibility and adaptability. Especially, besides non-negativity constraints, this formulation may involve other sparsity priors by exploiting adequate dictionary learning techniques [8]. Future work will explore and evaluate such alternative blind dictionary-based approaches as well as the integration of the proposed approach in analog data assimilation methods [12].

References

[1] A Hannachi, IT Jolliffe, and DB Stephenson. Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology*, 27(9):1119–1152, 2007.



(a) Simulated fields (b) Modal SSS predictions (c) SST-SSS correlation PDF

Fig. 2: Non-negative decomposition of SST-SSS relationships in the Alboran Sea: (a) SST and SSS fields on March 22^{nd}, 2011, (b) SST-derived predictions of the SSS fields for each mode of the considered 2-mode decomposition, (c) distribution of SST-SSS correlation when mode 1 (resp. mode 2) dominates, i.e. $\alpha_{n1} > \alpha_{n2}$ (resp. $\alpha_{n2} > \alpha_{n1}$).

[2] P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, 1st edition, 2010.

[3] D.D. Lee and H.S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, October 1999.

[4] F. Abrard and Y. Deville. A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources. *Signal Process.*, 85(7):1389–1403, July 2005.

[5] O. Yilmaz and S. Rickard. Blind separation of speech mixtures via time-frequency masking. *IEEE Transactions on signal processing*, 52(7):1830–1847, 2004.

[6] M. Lopez-Radcenco, A. Aissa-El-Bey, P. Ailliot, R. Fablet, and P. Tandeo. Non-negative decomposition of linear relationships: application to multi-source ocean remote sensing data. In *ICASSP 2016 : 41st IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4179–4183, 2016.

[7] W.S. DeSarbo and W.L. Cron. A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, 5(2):249–282, 1988.

[8] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, Nov 2006.

[9] P. L. Combettes and J.C. Pesquet. Proximal Splitting Methods in Signal Processing. In H.H. Bauschke, R.S. Burachik, P.L. Combettes, V. Elser, D.R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011.

[10] C.L. Lawson and R.J. Hanson. *Solving least squares problems*, volume 15, chapter 23, page 161. SIAM, 1995.

[11] E.N. Lorenz. Predictability: a problem partly solved. In *Seminar on Predictability, 4-8 September 1995*, volume 1, pages 1–18, Shinfield Park, Reading, 1995. ECMWF, ECMWF.

[12] R. Lguensat, P. Tandéo, P. Ailliot, Pulido M., and R. Fablet. The analog data assimilation. In *Monthly Weather Review*, Submitted.

[13] M. Juza, B. Mourre, L. Renault, S. Gómara, K. Sebastián, S. Lora, J. P. Beltran, B. Frontera, B. Garau, C. Troupin, M. Torner, E. Heslop, B. Casas, R. Escudier, G. Vizoso, and J. Tintoré. Socib operational ocean forecasting system and multi-platform validation in the western mediterranean sea. *Journal of Operational Oceanography*, 9(sup1):s155–s166, 2016.