NON-NEGATIVE DECOMPOSITION OF LINEAR RELATIONSHIPS: APPLICATION TO MULTI-SOURCE OCEAN REMOTE SENSING DATA

Manuel López-Radcenco[†]   *Abdeldjalil Aïssa-El-Bey*[†]   *Pierre Ailliot*[⋆]   *Pierre Tandéo* [†]   *Ronan Fablet*[†]

[†] Institut Mines-Télécom, Télécom Bretagne, UMR CNRS 6285 Lab-STICC
Technopôle Brest-Iroise CS83818, 29238 Brest Cedex 3, France
[⋆] Laboratoire de Mathématiques de Bretagne Atlantique, UMR 6205, Université de Brest
6, Avenue Victor Le Gorgeu, B.P. 809, 29285 Brest Cedex, France

ABSTRACT

The identification and separation of contributions associated with different sources or processes is a general problem in signal and image processing. Here, we focus on the decomposition of multiple linear relationships and introduce a non-negative formulation. The proposed models can be viewed as generalizations of latent class regression models and account for possibly varying magnitudes of the linear transfer functions. Along with these models, we present model calibration algorithms. We first demonstrate their performance on simulated data. We also report an application to the analysis of upper ocean dynamics from remote sensing data (namely, satellite-derived Sea Surface Height (SSH) and Sea Surface temperature (SST) image series). This application further stresses the proposed formulation's relevance compared to state-of-the-art regression models.

Index Terms— Linear relationships, non-negativity, latent class model, Expectation-Maximization algorithm, multi-source remote sensing data

1. INTRODUCTION AND PROBLEM STATEMENT

The separation and identification of contributions associated with different types of sources or processes from multiple observations is a general problem in signal and image processing [1]. Significant advances have been reported in the last decade with the introduction of concepts such as non-negative matrix factorization (NN-MF) and sparsity [2–5]. From a mathematical point of view, this issue is generally stated as the decomposition of a signal into K additive components:

$$\mathbf{y} = \sum_{k=1}^{K} \alpha_k \boldsymbol{b}_k + \boldsymbol{\omega} \qquad (1)$$

where $\mathbf{y} \in \mathbb{R}^I$, coefficient $\alpha_k \in \mathbb{R}$ quantifies the contribution of component $\boldsymbol{b}_k \in \mathbb{R}^I$, which corresponds to the k-th reference signal or image and $\boldsymbol{\omega} \in \mathbb{R}^I$ is a multivariate white Gaussian noise process. In this context, non-negativity constraints are used to develop part-based representations that have multiple applications, ranging from learning parts of faces or semantic features of text [2] to audio blind source separation of convolutive mixtures [3]. Sparsity constraints, on the other hand, allow for dimensionality reduction and simpler representations of high-dimensional data [6], thus leading to the development of simpler models and representations that are easier to understand.

In this paper, we address the decomposition of multiple linear relationships between two signals under non-negativity constraints as in model (1). The proposed models generalize classical regression and latent class regression models [7,8]. Compared to these state-of-the-art models, their key features are two-fold: first, they account for possibly varying magnitudes of the linear relationships; second, they explicitly evaluate the relative importance of different linear relationships. This is of wide interest for a variety applications such as regression hypothesis testing, transfer function identification, regime-switching dynamics, etc. As an illustration, we address here the analysis of upper ocean dynamics from the identification and decomposition of the linear relationships exhibited by satellite-derived sea surface geophysical image series.

2. PROPOSED MODELS

2.1. General model

Let us denote by $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$ the nth samples of some multivariate observation dataset. These two variables may account for independent pairs of samples, when addressing classical regression problems, as well as signal or images patches when dealing with transfer functions between two signals, images or images series $\mathbf{X}$ and $\mathbf{Y}$.

Our goal is to identify and decompose multiple linear relationships possibly exhibited by multivariate variables $\mathbf{x}_n$ and $\mathbf{y}_n$. To this end, we propose to extend model (1) as:

$$\mathbf{y}_n = \sum_{k=1}^{K} \alpha_{nk} \boldsymbol{\beta}_k \mathbf{x}_n + \omega_n$$

$$\text{Subject to} \begin{cases} \alpha_{nk} \geq 0, & \forall\, k \in [\![1, K]\!],\, \forall n \in [\![1, N]\!] \\ ||\boldsymbol{\beta}_k||_F = 1, & \forall\, k \in [\![1, K]\!] \end{cases} \qquad (2)$$

where $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$, $\alpha_{nk} \in \mathbb{R}^+$, $\boldsymbol{\beta}_k \in \mathbb{R}^{I \times J}$, $|| \cdot ||_F$ is the Frobenius norm and ω_n is a centered Gaussian noise process with covariance matrix $\boldsymbol{\Sigma}$ that models estimation residuals. N is the total number of observations available, K is the total number of components or modes in the mixture and k and n indicate, respectively, the current observation and mode considered. The first model constraint derives from the desired non-negativity of coefficients α_{nk}, while the second constraint ensures the identifiability of the problem. The non-negativity constraint provides the mean for distinguishing the form of the linear relationships between variables $\mathbf{x}_n$ and $\mathbf{y}_n$ from the magnitude of these relationships.

Model (2) allows us to address decomposition problems involving mixed linear contributions. This model generalizes linear mixture problems involving linear regressions and latent class regression models [7, 8]. The later might be viewed as a simplified version of

our model (2), where mixing coefficients do not depend on index n and only one mode is actually active for each sample pair. Besides, a classical linear regression resorts to model (2) with $K = 1$ and $\alpha_{nk} = \alpha^* \; \forall k \in [\![1, K]\!], \forall n \in [\![1, N]\!]$.

2.2. Latent class model

Under the assumption that for each pair of variables $\mathbf{x}_n$ and $\mathbf{y}_n$ there is a strongly dominant mode, that is to say that for any sample pair $(\mathbf{x}_n, \mathbf{y}_n)$ only one of the mixing coefficients α_{nk} is non-null, we introduce a hidden variable z_n that indicates the mode which operates for each sample pair $(\mathbf{x}_n, \mathbf{y}_n)$, and we derive a simplified latent class formulation for model (2) :

$$
\begin{cases}
P\left(\mathbf{y}_n | z_n = k\right) \sim \mathcal{N}\left(0, \mathbf{\Sigma}_k, r\left(\mathbf{x}_n, \mathbf{y}_n\right)\right) \\
r\left(\mathbf{x}_n, \mathbf{y}_n\right) = \mathbf{y}_n - \alpha_{nk}^* \boldsymbol{\beta}_k \mathbf{x}_n \\
\alpha_{nk}^* = \underset{\alpha_{nk}}{\operatorname{argmin}} ||\mathbf{y}_n - \alpha_{nk} \boldsymbol{\beta}_k \mathbf{x}_n||_{\mathbf{\Sigma}_k}^2
\end{cases}
\tag{3}
$$

where $\mathcal{N}\left(0, \mathbf{\Sigma}_k, \cdot\right)$ is a centered Gaussian distribution with covariance matrix $\mathbf{\Sigma}_k$ and $|| \cdot ||_{\Sigma_k}$ is the mode dependent covariance-weighted distance.

It may be noted that this latent class model still generalizes latent class regression model as used in [8], since non-null coefficients α_{nk} may still vary with respect to sample index n. As detailed in the subsequent section, given its lower computational complexity, this latent class model is of particular interest in the proposed multi-stage calibration process for model (2), and may be used as an initialization.

3. MODEL CALIBRATION AND INVERSION

Given some joint dataset $\{\mathbf{x}, \mathbf{y}\}_n$, we aim at calibrating and solving for model (2). Two different situations may be distinguished: i) the calibration of all model parameters, *i.e.* regression matrices $\boldsymbol{\beta}_k$ and non-negative mixing coefficients α_{nk}, from dataset $\{\mathbf{x}, \mathbf{y}\}_n$, ii) the estimation of non-negative mixing coefficients α_{nk} for dataset $\{\mathbf{x}, \mathbf{y}\}_n$ with known regression matrices $\boldsymbol{\beta}_k$. The later may be regarded as a special case of the first one, as detailed below.

We first describe the proposed alternating least square (ALS) procedure for model (2). We then derive the Expectation-Maximization (EM) procedure for the proposed latent class model (3), as well as a moment-based calibration scheme.

3.1. ALS calibration

The calibration of model (2) according to a Maximum Likelihood criterion resorts to the following least square minimization to infer model parameters α_{nk} and $\boldsymbol{\beta}_k$ from an observation dataset $\{\mathbf{x}, \mathbf{y}\}_n$:

$$
\begin{cases}
\left[\hat{\alpha}_{nk}, \hat{\boldsymbol{\beta}}_k\right] = \underset{\alpha_{nk}, \boldsymbol{\beta}_k}{\operatorname{argmin}} \sum_{n=1}^{N}\left(\left\|\mathbf{y}_n - \sum_{k=1}^{K} \alpha_{nk} \boldsymbol{\beta}_k \mathbf{x}_n\right\|_{\mathbf{\Sigma}}^2\right) \\
\alpha_{nk} \geq 0, \qquad \forall n \in [\![1, N]\!], \forall k \in [\![1, K]\!] \\
||\boldsymbol{\beta}_k||_2 = 1, \qquad \forall k \in [\![1, K]\!]
\end{cases}
\tag{4}
$$

where $||\cdot||_{\mathbf{\Sigma}}$ is the quadratic norm associated with covariance matrix $\mathbf{\Sigma}$.

The minimization problem presented in Eq. (4) is not linear nor convex if we consider joint minimization over α_{nk} and $\boldsymbol{\beta}_k$. It becomes linear and convex, however, if we fix α_{nk} and perform the minimization over $\boldsymbol{\beta}_k$ or vice versa.

A classical alternating least squares (ALS) minimization [9–12], in which we alternate a minimization over $\boldsymbol{\beta}_k$ with fixed α_{nk} ($\boldsymbol{\beta}$-step)

with a minimization over α_{nk} with fixed $\boldsymbol{\beta}_k$ (α-step), leads to the following updates of model parameters α_{nk} and $\boldsymbol{\beta}_k$:

$$
\hat{\alpha}_{nk}^{i+1} = \hat{\alpha}_{nk}^i + \frac{\mathbf{x}_n^T\left(\hat{\boldsymbol{\beta}}_k^i\right)^T \mathbf{\Sigma}^{-1}\left(\mathbf{y}_n - \sum_{p=1}^{K} \hat{\alpha}_{np}^i \hat{\boldsymbol{\beta}}_p^i \mathbf{x}_n\right)}{\mathbf{x}_n^T\left(\hat{\boldsymbol{\beta}}_k^i\right)^T \mathbf{\Sigma}^{-1} \hat{\boldsymbol{\beta}}_k^i \mathbf{x}_n}
\tag{5}
$$

$$
\hat{\boldsymbol{\beta}}_k^{i+1} = \hat{\boldsymbol{\beta}}_k^i + \left[\sum_{n=1}^{N} \hat{\alpha}_{nk}^i \left(\mathbf{y}_n - \sum_{p=1}^{K} \hat{\alpha}_{np}^i \hat{\boldsymbol{\beta}}_p^i \mathbf{x}_n\right) \mathbf{x}_n^T\right] \\
\left[\sum_{n=1}^{N}\left(\hat{\alpha}_{nk}^i\right)^2 \mathbf{x}_n \mathbf{x}_n^T\right]^{-1}
\tag{6}
$$

The non-negativity constrains over α_{nk} and the normalization constraints over $\boldsymbol{\beta}_k$ are forced at the end of each step as follows:

$$
\hat{\boldsymbol{\beta}}_k^{i+1} = \frac{\hat{\boldsymbol{\beta}}_k^{i+1}}{\left\|\hat{\boldsymbol{\beta}}_k^{i+1}\right\|_F}, \quad \forall k \in [\![1, K]\!]
\tag{7}
$$

$$
\hat{\alpha}_{nk}^{i+1} = 0, \quad if \quad \hat{\alpha}_{nk}^{i+1} < 0
\tag{8}
$$

Experimental results suggest that the ALS approach is dependent on initialization, which motivated us to use the maximum likelihood parameters of model (3) as initialization for model (2).

3.2. Latent class model calibration

For latent class model (3), we derive a classical iterative EM procedure [13–15] to infer model parameters $\boldsymbol{\beta}_k$ and α_{nk} according to a ML criterion. At iteration i, the E-step comes to compute the posterior of hidden variables z_n given current model parameters $\boldsymbol{\beta}_k^i$ and α_{nk}:

$$
\tau_{nk}^{i+1} = P\left(z_n = k | \mathbf{x}_n, \mathbf{y}_n\right) = \frac{\pi_k^i \mathcal{N}\left(0, \mathbf{\Sigma}_k, \mathbf{y}_n - \hat{\alpha}_{nk}^i \hat{\boldsymbol{\beta}}_k^i \mathbf{x}_n\right)}{\sum_{l=1}^{K} \pi_l^i \mathcal{N}\left(0, \mathbf{\Sigma}_k, \mathbf{y}_n - \hat{\alpha}_{nl}^i \hat{\boldsymbol{\beta}}_l^i \mathbf{x}_n\right)}
\tag{9}
$$

The M-step then updates model parameters as follows:

$$
\pi_k^{i+1} = P\left(z = k\right) = \frac{1}{N} \sum_{n=1}^{N} \tau_{nk}^{i+1}
\tag{10}
$$

$$
\hat{\boldsymbol{\beta}}_k^{i+1} = \left[\frac{1}{N} \sum_{n=1}^{N} \tau_{nk}^{i+1} \mathbf{y}_n \hat{\alpha}_{nk}^i \mathbf{x}_n^T\right]\left[\frac{1}{N} \sum_{n=1}^{N} \tau_{nk}^{i+1} \hat{\alpha}_{nk}^i \mathbf{x}_n \hat{\alpha}_{nk}^i \mathbf{x}_n^T\right]^{-1}
\tag{11}
$$

$$
\hat{\alpha}_{nk}^{i+1} = \frac{\mathbf{y}_n^T\left(\mathbf{\Sigma}_k^i\right)^{-1} \hat{\boldsymbol{\beta}}_k^{i+1} \mathbf{x}_n}{\mathbf{x}_n^T\left(\hat{\boldsymbol{\beta}}_k^{i+1}\right)^T\left(\mathbf{\Sigma}_k^i\right)^{-1} \hat{\boldsymbol{\beta}}_k^{i+1} \mathbf{x}_n}
\tag{12}
$$

$$
\mathbf{\Sigma}_k^{i+1} = \frac{\sum_{n=1}^{N} \tau_{nk}^{i+1}\left(\mathbf{y}_n - \hat{\alpha}_{nk}^{i+1} \hat{\boldsymbol{\beta}}_k^{i+1} \mathbf{x}_n\right)\left(\mathbf{y}_n - \hat{\alpha}_{nk}^{i+1} \hat{\boldsymbol{\beta}}_k^{i+1} \mathbf{x}_n\right)^T}{\sum_{n=1}^{N} \tau_{nk}^{i+1}}
\tag{13}
$$

The non-negativity constraints over α_{nk} and the normalization constraints over $\boldsymbol{\beta}_k$ are forced after each EM iteration (Eqs. (7) and (8)).

Experimentally, it has been noted that an ill-conditioned or nearly singular covariance matrix may result in numerical instabilities in

the update of scalars α_{nk} (Eq. (12)). Hence, we may consider a variant of the algorithm with a spherical covariance matrix, $\boldsymbol{\Sigma}_k^i = \sigma^2 \mathbf{I}_I$, $\forall k \in [1, K]$. We will refer to this variant of the EM algorithm as the pseudo-EM algorithm.

3.3. Moment-based estimation of regression parameters

Regarding initialization issues, we also derive a moment-based estimation of regression matrices $\boldsymbol{\beta}_k$. Noting that all amplitude information is contained in coefficients α_{nk}, a direct application of the orthogonality principle [16] yields that regression matrix $\boldsymbol{\beta}_k$ that minimizes the root mean square estimation error is the matrix that causes vectors $\boldsymbol{\beta}_k \mathbf{x}_n$ and $\mathbf{y}_n$ to be colinear for all sample pairs $(\mathbf{x}_n, \mathbf{y}_n)$ assigned to mode k. Therefore, assuming that we are provided with some estimates or initial values of posteriors τ_{nk} (Eq. (9)), we can derive the following moment-based estimate for regression matrix $\boldsymbol{\beta}_k$:

$$\hat{\boldsymbol{\beta}}_k = \left[\frac{1}{N} \sum_{n=1}^{N} \tau_{nk} \mathbf{y}_n \mathbf{x}_n^T \right] \left[\frac{1}{N} \sum_{n=1}^{N} \tau_{nk} \mathbf{x}_n \mathbf{x}_n^T \right]^{-1} \quad (14)$$

We then define a moment-based conditional estimation algorithm (MICE) [17–20]. It consists in replacing the update of regression matrices $\boldsymbol{\beta}_k$ in the M-step of the EM procedures (Eq. (11)) by the above moment-based update (Eq. (14)). As detailed below, this moment-based estimation is considered as initialization to the EM iterations, since it does not require knowing initial values for mixing coefficients α_{nk}.

3.4. Non-negative linear decomposition algorithm (NN-LD)

We detail in Algorithm 1 the overall procedure for the calibration of model (2). We adopt a greedy approach to improve convergence properties. As initialization, we perform an initial K class clustering of the data using a simple k-means algorithm [21–23], compute initial binary posteriors τ_{nk}, and exploit the MICE update of regression matrices $\boldsymbol{\beta}_k$ and mixing coefficients α_{nk}. We then iterate the EM (or pseudo-EM) procedure until convergence for latent class model (3) and follow with iterations of the ALS procedure until convergence for general model (2).

Algorithm 1 Non-negative Linear Decomposition

1: Input: $\{\mathbf{x}_n\}, \{\mathbf{y}_n\}$
2: Initialization:
$\quad \tau_{nk}^0 \leftarrow \text{k-means}(\mathbf{x}_n, \mathbf{y}_n)$
$\quad \left[\hat{\alpha}_{nk}^1, \hat{\boldsymbol{\beta}}_k^1, \tau_{nk}^1, \pi_k^1, \boldsymbol{\Sigma}_k^1 \right] \leftarrow \text{MICE}(\mathbf{x}_n, \mathbf{y}_n, \tau_{nk}^0)$
3: **repeat**
$\quad \left[\hat{\alpha}_{nk}^{i+1}, \hat{\boldsymbol{\beta}}_k^{i+1}, \tau_{nk}^{i+1}, \pi_k^{i+1}, \boldsymbol{\Sigma}_k^{i+1} \right] \leftarrow$
$\quad\quad \text{EM}\left(\mathbf{x}_n, \mathbf{y}_n, \hat{\alpha}_{nk}^i, \hat{\boldsymbol{\beta}}_k^i, \tau_{nk}^i, \pi_k^i, \boldsymbol{\Sigma}_k^i \right)$
$\quad$ **or**
$\quad \left[\hat{\alpha}_{nk}^{i+1}, \hat{\boldsymbol{\beta}}_k^{i+1}, \tau_{nk}^{i+1}, \pi_k^{i+1}, \boldsymbol{\Sigma}_k^{i+1} \right] \leftarrow$
$\quad\quad \text{pseudo-EM}\left(\mathbf{x}_n, \mathbf{y}_n, \hat{\alpha}_{nk}^i, \hat{\boldsymbol{\beta}}_k^i, \tau_{nk}^i, \pi_k^i, \boldsymbol{\Sigma}_k^i \right)$
4: **until** convergence
5: **repeat**
$\quad \left[\hat{\alpha}_{nk}^{i+1}, \hat{\boldsymbol{\beta}}_k^{i+1} \right] \leftarrow \text{ALS}\left(\mathbf{x}_n, \mathbf{y}_n, \hat{\alpha}_{nk}^i, \hat{\boldsymbol{\beta}}_k^i \right)$
6: **until** convergence

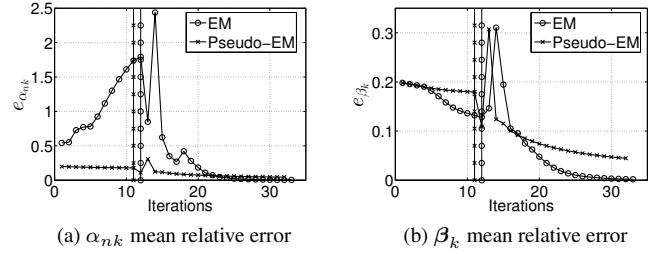(a) α_{nk} mean relative error   (b) $\boldsymbol{\beta}_k$ mean relative error

Fig. 1: Mean relative parameter estimation errors. *Vertical lines indicate the iteration number at which the change between the EM (pseudo-EM) procedure for the latent class model (3) and the ALS procedure for general model (2) occurs*

Table 1: Estimation features for the groundtruthed synthetic datasets. *Results are presented for both variants of the NN-LD algorithm (Algorithm 1)*

	EM variant	pseudo-EM variant
$e_{\alpha_{nk}}$	$0.19\,\%$	$8.02\,\%$
$e_{\boldsymbol{\beta}_k}$	$0.40\,\%$	$4.47\,\%$
$RRMSE$	$0.07\,\%$	$0.48\,\%$
$\mathcal{L}(\alpha_{nk}, \boldsymbol{\beta}_k)$	-8.8305×10^5	-1.4104×10^6

4. EXPERIMENTS

4.1. Evaluation on synthetic datasets

We first evaluate the proposed model calibration procedures on synthetic datasets. We simulate data according to model (2), considering $I = 3$, $J = 2$, $N = 100000$, $K = 2$ and randomly generated mixing coefficients α_{nk}^* and regression matrices $\boldsymbol{\beta}_k^*$.

Both variants of the NN-LD algorithm (EM and pseudo-EM) are then applied to the generated dataset to estimate its mixing coefficients α_{nk} and regression matrices $\boldsymbol{\beta}_k$.

With a view to characterizing calibration performance, we analyze the evolution of the mean relative error between our parameter estimates $\hat{\boldsymbol{\beta}}_k$ and $\hat{\alpha}_{nk}$ and the real parameter values $\boldsymbol{\beta}_k^*$ and α_{nk}^* ($e_{\alpha_{nk}}$ and $e_{\boldsymbol{\beta}_k}$) for both variants of the NN-LD algorithm. The obtained results are presented in Fig. 1.

We also compute the log-likelihood ($\mathcal{L}(\alpha_{nk}, \boldsymbol{\beta}_k)$) and the relative root mean square error ($RRMSE$) for the reconstruction of variables $\{\mathbf{y}_n\}$. The results are summarized in Table 1.

The proposed algorithms accurately estimate model parameters with reasonable error statistics (Table 1).

The EM variant of the algorithm outperforms the pseudo-EM variant, since it reaches lower error values and higher log-likelihood values (Table 1). This result is explained by the fact that the pseudo-EM variant does not take the covariance structure of estimation residuals into account for the estimation of model parameters.

As far as reconstruction is concerned, the relative root mean square error values (Table 1) indicate that both variants have good reconstruction performance, with a greater accuracy for the EM variant.

We conclude that the EM variant of the algorithm should be used when possible, but in cases where numerical instability problems may arise, the pseudo-EM variant could be a viable choice. Choosing between these two algorithm variants implies making a compromise between accuracy and numerical stability and will ultimately
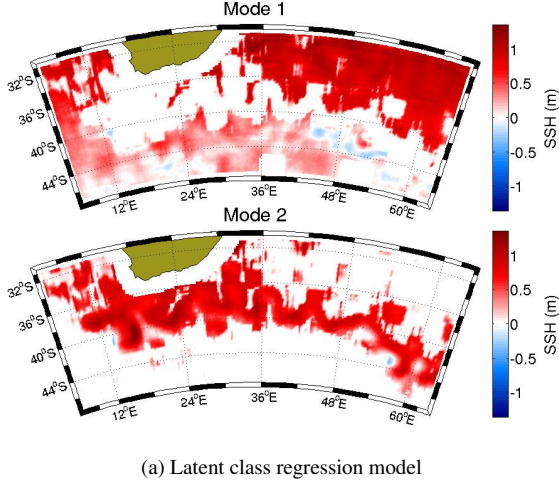
(a) Latent class regression model
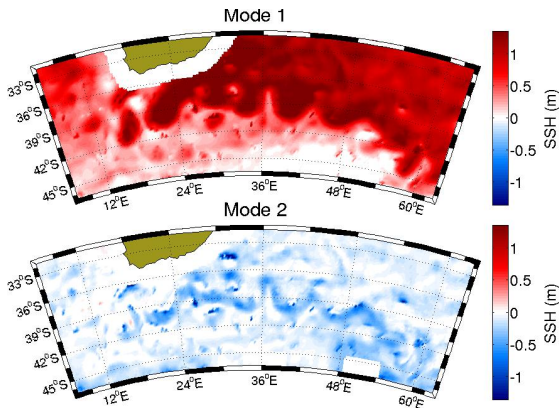
(b) Proposed model

Fig. 2: Ocean surface dynamics segmentation. *Predicted SSH per mode for the considered models (K = 2)*

depend on the nature of each specific problem and dataset.

4.2. Application to ocean remote sensing data

We applied the NN-LD algorithm in its EM variant to real ocean remote sensing microwave temperature and altimetry data. The main objective is to illustrate the relevance of the proposed model to investigate and decompose upper ocean dynamics from multi-sensor data synergies, namely sea surface temperature (SST) and sea surface height (SSH). From a theoretical point of view [24, 25], SST and SSH images are expected to be related through linear transfer functions. As case-study, we focus on the Agulhas region, a dynamical ocean region between longitudes $5°E$ to $65°E$ and latitudes $30°S$ to $48°S$, during the year 2004. The dataset used is the same as in [8]. SST and SSH images are interpolated on the same $1/4° \times 1/4°$ Mercator grid. Restating linear transfer functions as matricial operators applied to image patches, we apply model (2) to vectorized versions of 3×3 SSH patches (noted $\mathbf{y}_n$, with $I = 9$) and 9×9 SST patches (noted $\mathbf{x}_n$, with $J = 81$). For model calibration, we randomly sample a dataset of 10^5 patch pairs. We also consider a test dataset of $\sim 1.3 \times 10^4$ patch pairs randomly sampled from the 1st January, 2004.

We consider a two-mode version of model (2) (*i.e.*, $K = 2$). We cal-

Table 2: Relative root mean square error for the different models considered

	RRMSE
Linear regression	28.69 %
Two-class regression model [8]	27.22 %
Proposed model with $K = 1$	1.67 %
Proposed model with $K = 2$	0.08 %

ibrate model parameters using the NN-LD algorithm applied to the training dataset. The retrieved regression matrices $\boldsymbol{\beta}_k$ are then used as the true regression matrices for the test dataset, and the NN-LD algorithm is used with fixed $\boldsymbol{\beta}_k$ to estimate mixing coefficients α_{nk} for the test dataset.

For comparison purposes, we also apply a classical regression model, a latent class regression model with two classes [8], and a one-mode version of model (2) (*i.e.*, $K = 1$) to the same datasets. We illustrate the intrinsic difference between the latent class regression model ($K = 2$) and the proposed model ($K = 2$) through the contribution of each mode to the reconstruction of the SSH field (Fig. 2). Whereas the latent class regression model performs some binary spatial segmentation, the proposed model decomposes the SSH field according to two additive components: the first one proving a coarse prediction of the SSH field, the second one consisting of a negative anomaly along the Agulhas current (the region depicting the greater current values). From a quantitative point view, we report in Table 2 the relative root mean square error (RRMSE) of the reconstruction of the SSH image for each model.

These results clearly demonstrate the relevance of the non-negative decomposition. It allows us to distinguish the form of the retrieved linear relationships between the SST and SSH fields, *i.e.* the normalized regression matrices $\boldsymbol{\beta}_k$, from the local magnitude of these linear relationships, *i.e.* the mixing coefficients α_{nk}. This is a key feature, as the local magnitude of the retrieved linear relationships is expected to vary in space and time, especially in relation to the depth of the mixing layer. Our model also provides the means to test for the relative relevance of one-mode against two-mode models. The two-mode model leads to a significant relative improvement by a factor of 20 (Table 2), whose geophysical relevance needs to be further analyzed in future work.

5. CONCLUSIONS AND FUTURE WORK

We have proposed new models to characterize linear relationships between variables, along with the associated calibration algorithms to estimate the model parameters from data. Good experimental convergence properties were reported on synthetic datasets. We have demonstrated the relevance of the proposed non-negative models for the modal decomposition of upper ocean dynamics from SSH-SST synergies.

Future work includes extensions to non-Gaussian noise processes and non-linear kernel-based decompositions [26, 27]. Additional constraints on the regression matrices, for instance sparsity constraints [5]) may be of particular interest for the interpretation of the identified regression matrices. Future work also involves further development of the application to joint SSH-SST analysis and reconstruction of upper ocean dynamics. Other application fields, such as the blind identification of MIMO systems that can be applied, for example, to the multipath problem in communication systems [28–30], might also be explored.

6. REFERENCES

[1] M. Pal, R. Roy, J. Basu, and M.S. Bepari, "Blind source separation: A review and analysis," in *Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), 2013 International Conference*, Nov 2013, pp. 1–5.

[2] D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[3] C. Févotte, N. Bertin, and J.L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural Comput.*, vol. 21, no. 3, pp. 793–830, Mar. 2009.

[4] A. Ozerov and C. Fevotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 3, pp. 550–563, March 2010.

[5] K. Stadlthanner, F.J. Theis, C.G. Puntonet, and E.W. Lang, "Extended sparse nonnegative matrix factorization," in *Computational Intelligence and Bioinspired Systems*, J. Cabestany, A. Prieto, and F. Sandoval, Eds., vol. 3512 of *Lecture Notes in Computer Science*, pp. 249–256. Springer Berlin Heidelberg, 2005.

[6] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[7] W.S. DeSarbo and W.L. Cron, "A maximum likelihood methodology for clusterwise linear regression," *Journal of Classification*, vol. 5, no. 2, pp. 249–282, 1988.

[8] P. Tandeo, B. Chapron, S. Ba, E. Autret, and R. Fablet, "Segmentation of mesoscale ocean surface dynamics using satellite sst and ssh observations," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 7, pp. 4227–4235, July 2014.

[9] T. Hastie, R. Mazumder, J. Lee, and R. Zadeh, "Matrix Completion and Low-Rank SVD via Fast Alternating Least Squares," *ArXiv e-prints*, Oct. 2014.

[10] R. Zadeh, H. Li, B. He, M. Lublin, and Y. Perez, "Cme 323: Distributed algorithms and optimization, spring 2015," University Lecture, 2015.

[11] G. Takács and D. Tikk, "Alternating least squares for personalized ranking," in *Proceedings of the Sixth ACM Conference on Recommender Systems*, New York, NY, USA, 2012, RecSys '12, pp. 83–90, ACM.

[12] D. Zachariah, M. Sundin, M. Jansson, and S. Chatterjee, "Alternating Least-Squares for Low-Rank Matrix Reconstruction," *IEEE Signal Processing Letters*, vol. 19, pp. 231–234, Apr. 2012.

[13] J. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," Tech. Rep. TR-97-021, ICSI, 1997.

[14] F. Dellaert, "The expectation maximization algorithm," Tech. Rep., Georgia Institute of Technology, 2002.

[15] P. Smyth, "Finite mixture models and the em algorithm," University Lecture, 2015.

[16] T.K. Moon and W.C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice Hall, 2000.

[17] W. Pieczynski, "Parameter estimation in the case of hidden data," in *"Proceedings 16th Biennal Symposium on Communications", Kingston, Canada*, 1992.

[18] W. Pieczynski, "Convergence of the iterative conditional estimation and application to mixture proportion identification," in *Statistical Signal Processing, 2007. SSP '07. IEEE/SP 14th Workshop on*, Aug 2007, pp. 49–53.

[19] W. Pieczynski, "Champs de markov cachés et estimation conditionnelle itérative," 1994.

[20] O. Aliagnat, J.M. Boucher, D.C. He, W. Pieczynski, and Universitc De Sherbrooke, "Hidden markov fields and unsupervised segmentation of images," in *IAPR 11th International Conference on Pattern Recognition (ICPR 92)*, 1992.

[21] T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, and A.Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 881–892, Jul 2002.

[22] J. B. Macqueen, "Some Methods for classification and analysis of multivariate observations," in *Procedings of the Fifth Berkeley Symposium on Math, Statistics, and Probability*. 1967, vol. 1, pp. 281–297, University of California Press.

[23] S. Na, L. Xumin, and G. Yong, "Research on k-means clustering algorithm: An improved k-means clustering algorithm," in *Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on*, April 2010, pp. 63–67.

[24] G. Lapeyre and P. Klein, "Dynamics of the upper oceanic layers in terms of surface quasigeostrophy theory," 2006.

[25] P. Klein, B.L. Hua, G. Lapeyre, et al., "Upper ocean turbulence from high-resolution 3d simulations," 2008.

[26] B. Schölkopf, A. Smola, E. Smola, and K.R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299–1319, 1998.

[27] Q. Wang, "Kernel principal component analysis and its applications in face recognition and active shape models," *CoRR*, vol. abs/1207.3538, 2012.

[28] A.G. Burr, "The multipath problem: an overview," in *Multipath Countermeasures, IEE Colloquium on*, May 1996, pp. 1/1–1/7.

[29] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *Information Theory, IEEE Transactions on*, vol. 43, no. 2, pp. 499–511, Mar 1997.

[30] E. Eidinger and A. Yeredor, "Blind mimo identification using the second characteristic function," *Signal Processing, IEEE Transactions on*, vol. 53, no. 11, pp. 4067–4079, Nov 2005.