NONLINEAR COMPLEX PCA
FOR SPATIO-TEMPORAL ANALYSIS OF GLOBAL SOIL MOISTURE

Diego Bueso, Maria Piles, Gustau Camps-Valls

Image Processing Lab (IPL), Universitat de València, València, Spain

ABSTRACT

Soil moisture (SM) is a key state variable of the hydrological cycle, needed to monitor the effects of a changing climate on natural resources. Soil moisture is highly variable in space and time, presenting seasonalities, anomalies and long-term trends, but also, and important nonlinear behaviours. Here, we introduce a novel fast and nonlinear complex PCA method to analyze the spatio-temporal patterns of the Earth's surface SM. We use global SM estimates acquired during the period 2010-2017 by ESA's SMOS mission. Our approach unveils both time and space modes, trends and periodicities unlike standard PCA decompositions. Results show the distribution of the total SM variance among its different components, and indicate the dominant modes of temporal variability in surface soil moisture for different regions. The relationship of the derived SM spatio-temporal patterns with El Niño Southern Oscillation (ENSO) conditions is also explored.

Index Terms— PCA, spatio-temporal data, kernel methods, promax, SMOS, soil moisture

1. INTRODUCTION

Soil moisture exerts a strong control not only on the water cycle but also in ecosystem functioning. In this regard, monitoring changes in soil moisture is critical, since they have a direct impact on agricultural productivity, forestry, and ecosystem health [1]. In this work, we present a novel methodology ,the rotated kernel complex Principal component analysis (ROCK PCA) that allows characterizing global SM spatio-temporal dynamics. To do so, we use the first seven years (2010-2017) of soil moisture estimates obtained from the ESA-led Soil Moisture and Ocean Salinity (SMOS) mission.

Principal component analysis, also known in geophysics as Empirical Orthogonal Functions (EOFs), is widely used to obtain compact representations of the signal [2], and has been widely exploited to obtain spatio-temporal features in climatological studies [3–5]. However, PCA is computationally demanding with a large number of features, and at the same time the spatio-temporal components are not disentangled in the

feature space. In this work, we propose a method that alleviates both problems. The complex PCA is first applied to spatial and temporal components through the Hilbert transform, thus leading to spatial and temporally explicit eigendecompositions easy to analyze. It exploits the fact that the eigendecomposition of the covariance and the Gram matrix return the same results, and hence the computational cost can be drastically reduced from quadratic $\mathcal{O}(n^2 t)$ to linear in the number of pixels $\mathcal{O}(t^2 n)$ [6]. The use of the Gram matrix is not incidental, and allows further improvements, such as the direct kernelization of the method [7], thus allowing to derive a fast nonlinear spatio-temporal PCA. An extra rotation, the promax oblique method, is also introduced for extra versatily as it allows us to find the optimal parameters for the decomposition and make the principal components physically interpretable.

The rest of the paper is outlined as follows. Section 2 introduces our method for spatio-temporal data decomposition: we first fix the notation, review the standard approaches in the literature, and introduce our nonlinear (kernel-based) complex PCA and the rotation approach. Section 3 presents the experimental results on SMOS data to uncover explicit temporal and spatial patterns of soil moisture, which are then related to ENSO events. Section 4 concludes with some remarks and future work.

2. NONLINEAR SPATIO-TEMPORAL DATA ANALYSIS

2.1. Notation

Let us define a spatio-temporal data cube $\mathbf{C} \in \mathbb{R}^{r \times c \times t}$, defined in a $r \times c$ spatial grid and t temporal observations. The cube can be reshaped into matrix form as $\tilde{\mathbf{X}} = [x_1, \cdots, x_n] \in \mathbb{R}^{t \times n}$, where the tilde indicates the column-wise centering operation for each time series $\mathbf{x}_i \in \mathbb{R}^{t \times 1}, i = 1, \cdots, n$. A standard way to analyze the feature relations contained in the data is to diagonalize its covariance matrix, $\mathbf{C} = \frac{1}{t}\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}} \in \mathbb{R}^{n \times n}$. However, given the high number of pixel observations, n, obtaining the eigenvalues and eigenvectors involves a high computational cost. A simple efficient alternative is to eigedecompose the Gram matrix: $\mathbf{G} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top \in \mathbb{R}^{t \times t}$, which returns exactly the same solution up to a scaling factor. This

is known as the primal-dual solution of the PCA.

2.2. Complex PCA

The main problem by following the standard PCA approach, is that interpretability of the components is hardly accessible. Eigenvectors and eigenvalues do not longer have a clear, physically meaningful interpretation in terms of spatial and temporal coordinates in the projection space [8]. The decomposition of the datacube into spatial and temporal eigecomponents is a traditional methodology commonly known as *complex* PCA [9]. The complex modality of PCA returns a more accurate decomposition and interpretable eigenvectors for geophysical data analysis than the plain PCA version. Formally, the complex PCA essentially applies the Hilbert transform to a signal $\mathbf{x}_i(t)$:

$$H(x_i(t)) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x_i(\tau)}{t - \tau} \, d\tau, i = 1, \cdots, n.$$

The Hilbert-transformed point is now expressed as $\mathbf{x}_{i,h} = \mathbf{x}_i + \mathbb{i}H(\mathbf{x}_i)$, and hence the centered Hilbert-transformed data matrix becomes:

$$\tilde{\mathbf{X}}_h = \tilde{\mathbf{X}} + \mathbb{i}H(\tilde{\mathbf{X}}),$$

where we define the Hermitian of $\mathbf{X}_h \in \mathbb{C}$ as $\mathbf{X}_h^{\mathbb{H}} \in \mathbb{C}$, and orthogonality holds, $\mathbf{x}_i \perp H(\mathbf{x}_i)$. Now, one can easily demonstrate that the covariance matrix of Hilbert-transformed data reduces to

$$\mathbf{C}_h = \frac{1}{t}\tilde{\mathbf{X}}_h^{\mathbb{H}}\tilde{\mathbf{X}}_h = \mathbf{C}_R + \mathbb{i}\mathbf{C}_I \in \mathbb{C}^{n \times n},$$

and the corresponding, and more convenient, centered Gram matrix is:

$$\mathbf{G}_h = \tilde{\mathbf{X}}_h\tilde{\mathbf{X}}_h^{\mathbb{H}} = \mathbf{G}_R + \mathbb{i}\mathbf{G}_I \in \mathbb{C}^{t \times t},$$

which is then used for the sake of a fast eigendecomposition. Note that now we retrieve separate spatial and temporal eigenvalues and eigenvectors, which allows us to study the signal in those terms. Making the eigendecomposition of the Gram matrix (classically, the covariance matrix), we obtain the eigenvalues ($\boldsymbol{\lambda}_h \in \mathbb{R}^{t \times 1}$) which are the explained variance for each Principal Component (PC) and the eigenvectors or PCs ($\boldsymbol{V}_h \in \mathbb{C}^{t \times t}$) which in Gram matrix case will be the estimate time series. The spatial features will be the projection onto the PCs ($\boldsymbol{Xp}_h = \boldsymbol{V}_h^{\mathbb{H}}\tilde{\mathbf{X}}_h \in \mathbb{C}^{t \times n}$).

2.3. Kernelized Complex PCA

Working with Gram matrices instead of covariances allows us to directly derive a nonlinear version of the previous complex PCA method by means of the kernel trick [2, 7]. Working with a nonlinear version can avoid adopting orthogonality constraints [8], which has a traditional solution with the rotated PCA [10]. The kernel PCA [2] could also achieve good representation of the data features but the spatial and temporal components would still become intertwined.

Alternatively, we derive here a kernelized version of the complex PCA. Let us define a feature map into a Hilbert space, $\phi : \mathbf{x}_i \mapsto \phi(\mathbf{x}_i) \in \mathcal{H}$, which is endorsed with a dot, product called *kernel function*, $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j) \in \mathbb{R}$. The (centered) kernel matrix groups all dot products into a matrix defined as $\mathbf{K} = \tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}^\top \in \mathbb{R}^{t \times t}$. The kernel function essentially computes similarities between feature vectors. Similarly, a kernel feature vector contains all similarities between a test point $\mathbf{x}_*$ and all the points in the datasets, and is defined as $\mathbf{k}_* := [k(\mathbf{x}_*, \mathbf{x}_1), \ldots, k(\mathbf{x}_*, \mathbf{x}_n)]^\top \in \mathbb{R}^{n \times 1}$. Then, it simply follows from the application of the Hilbert transform and the orthogonality property that the corresponding kernelized complex PCA reduces to eigendecompose

$$\mathbf{K}_h = \tilde{\mathbf{\Phi}}_h\tilde{\mathbf{\Phi}}_h^{\mathbb{H}} = \mathbf{K}_R + \mathbb{i}\mathbf{K}_I \in \mathbb{C}^{t \times t},$$

and thus we can analyze the signal in nonlinear terms, and separate components in space and time as a classic kernel PCA [11]. Taking into account the circularity of our complex data from Hilbert transform we use a complex RBF (Radial Basis Function) kernel function introduced by [12] defined as $k(\mathbf{x}_i, \mathbf{x}_j) = exp(-(\mathbf{x}_i - \mathbf{x}_j^*)(\mathbf{x}_i - \mathbf{x}_j^*)^T/2\sigma^2)$ which hold the circular properties of input data and is defined as Hermitian and semi-positive definite. Our nonlinear decomposition is defined by the kernel parameter σ and the number of PCs. This reduces our problem to compute the eigendecomposition of the kernel matrix $\mathbf{K}$ to obtain $\boldsymbol{V}_h$. In the next section we show the way to fit the σ hyperparameter using a rotated approach while making an oblique rotation of the PCs to create a more amenable physically-meaningful representation of data.

2.4. Promax Rotation

Using Kernel functions allows to finding nonlinear representations of the data, but frequently cannot tackle the orthogonality constraint since PCA methods find the basis which represents the maximum variance of data, but not necessarily a physically-meaningful representation of the variables [8]. Using an oblique rotation onto PCs, the orthogonality problem can be addressed [10]. The promax rotation [13] is based on the varimax orthonormal criterion which finds a rotation of a subset of PCs to maximize variance [14]. Optimizing the varimax parameter will also allow us to find a proper σ value and a number of PCs which makes the PCs more physically interpretable.

2.5. Illustrative example

We illustrate the benefits of the proposed ROCK PCA in a simple toy example. A theoretical spatio-temporal dataset has

been generated following the function

$$F(x, y, t) = e^{-|t|}cos(w_x x) + sin(w_t t)cos(w_y y)$$

with two independent time-series and spatial dependence, fixing the time range between $[-2\pi : 2\pi]$ with $w_t = 2\pi f = 8$. Optimizing the varimax function, we obtain four PCs and $\sigma = 793$, but only two of them represent the theoretical data (first and third in the rotated case). Note that in the case of rotation, one frequently finds two identical PCs, since the oblique rotation allows correlated PCs. As we show in Fig. 1 the kernel method find a close decomposition from the theoretical time-series but the rotation is needed to obtain the closest estimation.
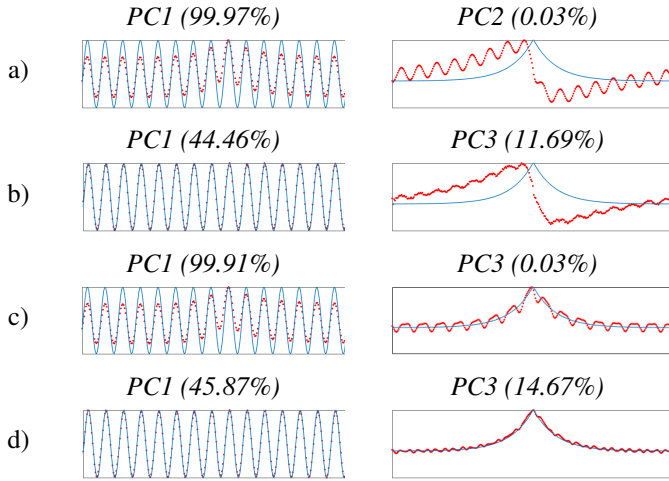
Fig. 1. Time series: We show the real part from the estimated principal components (red) and his theoretical function (blue) with his explained variance on top. a) Complex PCA and b) Rotated Complex PCA represent the linear methods, and the c) Complex Kernel PCA and d) ROCK-PCA the nonlinear approaches.

3. EXPERIMENTAL RESULTS

This section presents the application of the proposed ROCK PCA in the analysis of the spatial and temporal relations in 7 years of Earth Observation (EO) SM data.

3.1. SMOS data

Global SMOS soil moisture data from the Barcelona Expert Center has been used in this study (http://bec.icm.csic.es/). Since its launch in 2009, SMOS is providing global maps of the Earth's surface soil moisture (top 5 cm) every 3-days with a spatial resolution of ∼50 km and a target accuracy of 0.04 m^3/m^3. For this work, we chose the first seven years of SMOS observations, after its commissioning phase (from May 2010 to May 2017). Note that this period encloses the 2015-2016 El Niño event –the so-called *el Niño Godzilla*, which will be studied.

3.2. SM spatio-temporal decomposition

The ROCK PCA analysis has been applied to the SM maps to obtain a multimodal decomposition. The explained variance for the top four components account for 98.2% of the total variance: 70.6 %(PC1), 13.5% (PC2), 10.2 % (PC3) and 4.3% (PC4). The first component returns an annual oscillation with a 12-month period and a global spatial distribution, representing the principal SM global trend. The second and fourth components are cast as 'resonances' of the annual SM cycle with six- and four-months periods which represent subcycles of global SM as ITCZ (Intertropical Convergence Zona) oscillation (PC2) or seasonal transition (PC4). Interestingly, the third component (PC3) shows a non-seasonal oscillation with a principal period of 4.5 years trend as ENSO anomalies (3−7 years) which can be interpreted as a long-term change in the global SM distribution.
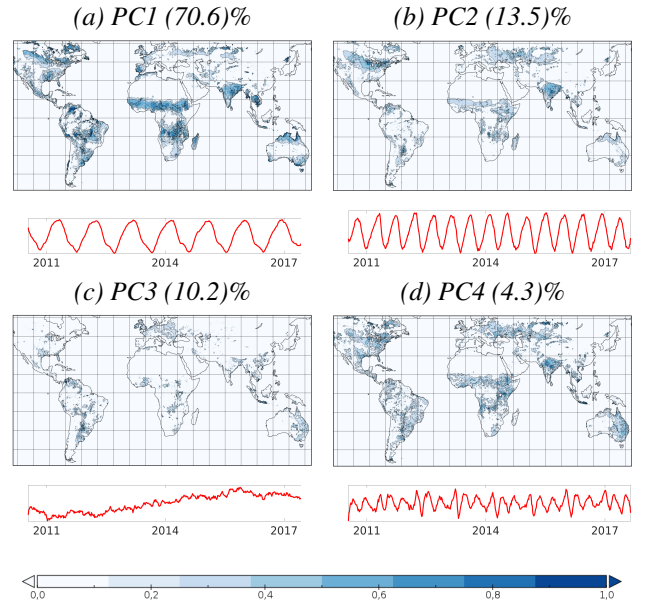
Fig. 2. Normalized Spatial distribution of amplitude (top) and his respective time series (bottom) for a) PC1, b) PC2, c) PC3 and d) PC4 with his explained variance respectively.

Figure 2 shows the (normalized) amplitudes of the spatial maps, which are the absolute value of the spatial projection of the PCs representing the spatial distribution of each time series. PC3 has a restricted spatial distribution with an homogeneous spatial phase that represent SM change which is strongly linked to the ENSO.

3.3. Connection between the ENSO and the SM trend

In this analysis, the identified SM cycles and long-term trends have been compared to the ENSO4 index, which represent the SST (Sea Surface Temperature) anomalies from western Pacific Ocean. To do so, we computed the Spearman correlation coefficient between the ENSO4 index and the SM long-trend

with the complex spatial phase added to show the spatial connection of the variables. Results are shown in Figure 3. It can be seen that there is a significant association between the SM long-trend and the ENSO4 index with a maximum absolute correlation of 0.8 with a p-value below of 10^{-16}.

The observed relationship shows the spatial distribution and strength of ENSO influence in wet-dry patterns, as previous studies [15] but using only direct EO data and also a wider spatial wet-dry pattern. Figure 3 presents the spatial distribution of positive and negative correlation coefficients: an increasing SM trend can be observed for instance in eastern Europe, southeastern South America and Southern US whereas a decreasing SM trend is visible in Australia's eastern cost, northern South America, northern gulf of Guinea. This connection shows an intuitive and direct relation between the ENSO4 dynamics and the extracted SM long term trends.
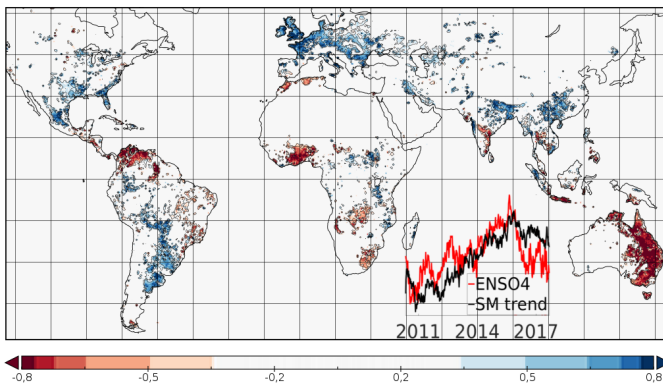
Fig. 3. Correlation map between SM long term trend and ENSO4 time series. Spatial dry-wet patterns can be identified. The inset figure shows the temporal evolution of PC3 and the ENSO4 index during the study period.

4. CONCLUSION

In this paper we propose the ROCK PCA method for spatial-temporal analysis and its improvement compared to classical PCA methods. Our approach unveiled both time and space modes, trends and periodicities unlike standard PCA decompositions. We illustrated the performance with a toy example with nonlinear time and space feature relations showing his ability for nonlinear data decomposition. we then showed a preliminary study of its applications to global satellite SM estimates that allowed associate its long-term component to ENSO. This method is a powerful tool for the study of global dynamics of EO data and can ultimately bring more information about climate teleconnections.

5. REFERENCES

[1] Sonia I. Seneviratne and et al., "Investigating soil moisture-climate interactions in a changing climate: A review," *Earth-Science Reviews*, vol. 99, no. 3-4.

[2] J. Arenas-Garca, K.B. Petersen, G. Camps-Valls, and L.K. Hansen, "Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 16–29, 2013.

[3] Bernhard Bauer-Marschallinger, Wouter A. Dorigo, and et al., "How oceanic oscillation drives soil moisture variations over mainland australia: An analysis of 32 years of satellite observations," *Journal of Climate*, vol. 26, no. 24.

[4] Denis L. Volkov, "Do the north atlantic winds drive the non-seasonal variability of the arctic ocean sea level?," *Geophysical Research Letters*, vol. 41, no. 6, 2014.

[5] E. Forootan, Khandu, and et al., "Quantifying the impacts of ENSO and IOD on rain gauge and remotely sensed precipitation products over australia," *Remote Sensing of Environment*, vol. 172.

[6] Alok Sharma and Kuldip K. Paliwal, "Fast principal component analysis using fixed-point algorithm," *Pattern Recognition Letters*, vol. 28, no. 10, 2007.

[7] G. Camps-Valls and L. Bruzzone, Eds., *Kernel methods for remote sensing data analysis*, Wiley & Sons, UK, Dec 2009.

[8] Thomas R. Karl, Albert J. Koscielny, and Henry F. Diaz, "Potential errors in the application of principal component (eigenvector) analysis to geophysical data," *Journal of Applied Meteorology*, vol. 21, no. 8, 1982.

[9] J. D. Horel, "Complex principal component analysis: Theory and examples," *Journal of Climate and Applied Meteorology*, vol. 23, no. 12, 1984.

[10] Tao Lian and Dake Chen, "An evaluation of rotated EOF analysis and its application to tropical pacific SST variability," *Journal of Climate*, vol. 25, no. 15, 2012.

[11] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 12991319, 1998.

[12] "Extension of wirtingers calculus to reproducing kernel hilbert spaces and the complex kernel LMS, volume=59, number=3, journal=IEEE Transactions on Signal Processing, author=Bouboulis, Pantelis and Theodoridis, Sergios, year=2011, pages=964978,".

[13] Alan E. Hendrickson and Paul Owen White, "Promax: A quick method for rotation to oblique simple structure," *British Journal of Statistical Psychology*, vol. 17, no. 1, pp. 6570, 1964.

[14] Henry F. Kaiser, "The varimax criterion for analytic rotation in factor analysis," *Psychometrika*, vol. 23, no. 3, pp. 187200, 1958.

[15] Diego G. Miralles, Martinus J. Van Den Berg, and et al., "El Niño-la Niña cycle and recent trends in continental evaporation," *Nature Climate Change*, vol. 4, no. 2.