CENTRE DE RECERCA MATEMÀTICA

Title:	*Lagged teleconnections of climate variables identified via complex rotated Maximum Covariance Analysis*
Journal Information:	*Journal of Climate*
Author(s):	Rieger N., Corral Á., Olmedo E., Turiel A..
Volume, pages:	*9861-9878*, DOI:[www.doi.org/10.1175/JCLI-D-21-0244.1]

Lagged teleconnections of climate variables identified via complex rotated Maximum Covariance Analysis

Niclas Rieger[1,2], Álvaro Corral[1], Estrella Olmedo[2], Antonio Turiel[2]

[1]Centre for Mathematical Research (CRM), Barcelona, Spain
[2]Institute of Marine Sciences (ICM-CSIC) & Barcelona Expert Center, Barcelona, Spain

March 16, 2021

A proper description of ocean-atmosphere interactions is key for a correct understanding of climate evolution. The interplay among the different variables acting over the climate is complex, often leading to correlations across long spatial distances (teleconnections). In some occasions, those teleconnections occur with quite significant temporal shifts that are fundamental for the understanding of the underlying phenomena but which are poorly captured by standard methods. Applying orthogonal decomposition such as Maximum Covariance Analysis (MCA) to geophysical data sets allows to filter out common dominant patterns between two different variables, but generally suffers from (i) the nonphysical orthogonal constraint as well as (ii) the consideration of simple correlations, whereby temporally offset signals are not detected. Here we propose an extension, complex rotated MCA, to address both limitations. We transform our signals using the Hilbert transform and perform the orthogonal decomposition in complex space, allowing us to correctly correlate out-of-phase signals. Subsequent Varimax rotation removes the orthogonal constraints, leading to more physically meaningful modes of geophysical variability. As an example of application, we have employed this method on sea surface temperature and continental precipitation; our method successfully captures the temporally and spatially interactions between these two variables, namely for (i) the seasonal cycle, (ii) canonical ENSO, (iii) the global warming trend, (iv) the Pacific Decadal Oscillation, (v) ENSO Modoki and finally (vi) the Atlantic Meridional Modulation. The complex modes of MCA provide information on the regional amplitude, and under certain conditions, the regional time lag between changes on ocean temperature and land precipitation. This approach can be useful especially, but not limited to, studies on the predictability of continental precipitation by other climate variables. The method is freely available as an easy-to-use `Python` package.

1 INTRODUCTION

The Earth's climate system is extremely complicated and deciphering the web of interdependencies and influences of different climate subsystems is an involved challenge. With the increase of Earth observations thanks to the progresses in remote sensing, the temporal and spatial resolutions of the observational data have increased along with the amount of data to be processed. Data-driven dimensionality reduction methods are therefore essential for climate studies, as they allow high-dimensional spatio-temporally resolved signals to be disaggregated into the dominant patterns, while still capturing the subtle details of higher resolution data. As such, Principal Component Analysis (PCA), or Empirical Orthogonal Functions (EOF) analysis as it is often referred to in climate science, allows to identify the dominant internal structure of a climate variable, with a variety of different available versions of PCA proving the popularity of such methods in climate science [1].

Climate phenomena with different expression in oceanic and atmospheric variables, such as the El Niño-Southern Oscillation (ENSO), however, require the simultaneous analysis of several variables for a more compre-

hensive description. In principle, multivariate PCA [2] makes possible to extract the patterns of co-variability of more than one variable. However, multivariate PCA accumulates the variance and the covariance of variables with very different variability in the same quantities. In consequence this may mask co-varying patterns as low-variability patterns of one variable can be erroneously accumulated in very dominant structures of one of the other, large-variability variables [3].

Maximum Covariance Analysis (MCA)[1] avoids this masking by taking into account only the covariance between two sets of variables. As such, it bears similarity to Canonical Correlation Analysis [4] which aims at maximising the temporal correlation between both variables. When the number of grid points (i.e. number of time series) is higher than the number of observations (i.e. number of time steps) and the data exhibits multicollinearity, as it is often the case for climate data, CCA fails as it requires the individual variance matrices to be non-singular unless regularised [5, 6]. In case the two fields of variables are identical, MCA reduces to PCA, the former thus being a natural generalisation of PCA.

Yet the methods discussed above maximise instantaneous correlation and do not consider time-delayed signals. To gain a deeper understanding of the dynamics of climate phenomena, however, it is necessary to systematically investigate time lags. A typical approach to tackle with this problem is to consider one variable set with a time lag defined *a priori* followed by a MCA [7]. However, this requires knowledge of the time lag which may vary from one location to another.

In this paper, we propose *complex rotated MCA* to systematically investigate the phase shift of two variables. We generate complex time series where the real and imaginary parts are related to each other by the Hilbert transform, known as the analytical signal, and decompose the covariance matrix in complex space, in analogy to complex PCA [8, 9]. We also effectively reduce spectral leakage inherent in the Hilbert transform of non-cyclic signals by a proper extrapolation of the signal beyond its boundaries. Finally, to relax the orthogonality constraint of the obtained solutions, we apply Varimax rotation to the spatial patterns, which leads to more localised solutions and thus facilitates their physical interpretation [10, 11].

To make the method readily accessible as a tool, we provide it as a Python package, called `pymca`, at `https://github.com/nicrie/pycca`. Due to the power and popularity of `numpy` and `xarray`, both classes form the basis of `pymca`, so that their typical data format can be used directly as input for analysis. The package is modularised in a way that provides the user free choice whether normal, complex, rotated or complex rotated MCA is to be performed. The user can also choose between Varimax orthogonal rotation as well as Promax oblique rotation. Further, if desired, standardisation of the input data is computed on the fly. The different flavours work in the same way for PCA, if one instead of two fields is provided as input.

The remainder of the article is structured as follows. Section 2 introduces the methodology, where we briefly discuss MCA (Sec. 2.1), complex MCA (Sec. 2.2) and rotated MCA (Sec. 2.3). Section 3 describes the data used to test the method using first synthetic data (Sec. 3.1) and then climatic variables (Sec. 3.2). Section 4 presents the results of both the synthetic (Sec. 4.1) and real-world experiments (Sec. 4.2). We conclude our study and provide directions for future research in Section 5.

2 METHODS

2.1 Maximum Covariance Analysis

Let us consider two spatio-temporal data fields $X_A \in \mathbb{R}^{m \times n_A}$ and $X_B \in \mathbb{R}^{m \times n_B}$ representing two different geophysical fields $s = \{A, B\}$, both having temporal dimension m and spatial dimensions n_A and n_B, respectively. In the following, we will refer to the temporal dimensions as the number of observations while we denote the

1 Sometimes referred to as Singular Value Decomposition (SVD) analysis. This name is unfortunate and should not be confused with the actual factorisation technique of a real/complex matrix.

spatial dimensions by the number of grid points. Assuming each time series to have zero mean, MCA then aims at maximising

$$v_A^T C v_B, \quad \text{s.t.} \quad v_A^T v_A = v_B^T v_B = 1 \tag{1}$$

where C denotes the temporal covariance matrix and v_A, v_B the *spatial patterns*, of both fields, respectively. Mathematically, this can be achieved by applying the singular value decomposition (SVD) to the covariance matrix,

$$C \propto X_A^T X_B = V_A \Sigma V_B^T, \tag{2}$$

with the columns of the obtained singular vector matrices $\{V_s\}_{s=A,B}$ representing pairs of spatial patterns describing the maximum amount of temporal covariance between both sets of variables. The entries of $\Sigma \in \mathbb{R}^{n_A \times n_B}$ along the main diagonal, the singular values σ_k, represent the covariance of each spatial pattern pair k, providing a mean of estimating the relative importance of each pair via the covariance fraction[2] γ_k

$$\gamma_k = \sigma_k \left(\sum_{j=1}^{\min(n_A, n_B)} \sigma_j \right)^{-1}. \tag{3}$$

By projecting the data fields on their respective singular vectors, we obtain the corresponding temporal evolution for each spatial pattern given by the columns of $X_s V_s$. Since the singular vectors are orthogonal, i.e. $V_s^T V_s = \mathbb{1}$, $s = A, B$, the corresponding projections are uncorrelated. In this paper, we will refer to the spatial patterns and their corresponding projections as *empirical orthogonal functions* (EOFs) and *principal components* (PCs), respectively, according to the usual convention in climate science. The EOFs and the PCs associated with a specific singular value σ_k is denoted as *mode* k.

2.2 Complex MCA

Propagating features or lagged signals could be detected by using a complex representation of the input fields. In analogy to complex PCA [12, 13, 8, 9], we complexify the real input fields via the Hilbert transform to construct the *analytical signal* $\hat{X}_s$ defined as

$$\hat{X}_s = X_s + i\mathcal{H}(X_s), \quad s = A, B \tag{4}$$

where $\mathcal{H}(\cdot)$ denotes the column-wise applied Hilbert transform. The analytical signal constructed in that way is a unique complex representation of the real signal, but whether or not it also represents a physical reality depends on the frequency spectrum of the analysed signal. By construction, the frequency components of the Hilbert transform are phase shifted by $-\pi/2$ with respect to those of the original signal. Therefore, for narrow-bandwidth signals, the Hilbert transform has a simple physical interpretation i.e., it represents a signal which arrives with a lag of one fourth of the typical period. If the signal consists of multiple dominating frequencies, however, the interpretation of the phase is more elusive, as it cannot be simply associated to a single frequency. Thus, the more narrow is the signal bandwidth, the more directly we can relate the phase to specific timings of its Hilbert transform [14].

A fundamental issue in the computation of the Hilbert transform arises when non-stationary or drifting signals are processed. Such signals are non-cyclic, and therefore, when the Fourier coefficients are calculated, strong boundary effects can occur due to spectral leakage [14] (Fig. 1). This problem can be circumvented by detrending the time series and considering only integer cycles as well as by applying window functions to the time series

2 Typically the squared covariance fraction defined as $\gamma_k^* = \sigma_k^2 \left(\sum_{j=1}^{\min(n_A, n_B)} \sigma_j^2 \right)^{-1}$ is considered for the relative importance of each mode for MCA. However, we opt for the non-squared covariance fraction since it is conserved under rotation. Furthermore, this measure is comparable to the solutions obtained by PCA, and in fact it is equivalent when $X_A = X_B$, for which MCA reduces to PCA and the singular values equal the eigenvalues in PCA.

(e.g. Hanning, Hamming etc.). However, this comes at the cost of information loss. Besides, such mitigation techniques are particularly ill suited to deal with intensification effects associated to climate change (that include not only noticeable trends on the mean level, but also increases in the amplitude of some periodic phenomena). Therefore, it seems important to introduce techniques capable to dealing with non-stationary dynamics.

To mitigate spectral leakage across the boundaries of the time series, we extrapolate the time series at both boundaries, to the past and to the future, using the optimised Theta model [15, 16], a special case of an autoregressive integrated moving average model with drift, ARIMA(0,1,1)[17]. The Theta model is a relatively simple yet well performing extrapolation method. While the forecast in itself actually works with non-cyclic signals, seasonal features are considered via multiplicative classical decomposition, thus allowing cyclic and non-cyclic signals to be extrapolated. This approach of handling the seasonal structure of a time series requires the user to specify the dominant period of the signal, T_s, beforehand. For discrete time series T_s represents the number of time steps needed to complete one cycle, that is, e.g. 365 for daily data considering an annual cycle, or 24 for hourly data with a daily cycle (for more details we refer the reader to [16]). We then apply the Hilbert transform to the extended series, so the spectral leakage is only important on the backward and forward extensions of it. Finally, we extract the central part (removing the parts corresponding to the extension), that correctly corresponds to the Hilbert transform of the original series. Using this approach, we effectively reduce the edge effects of the Hilbert transform compared to a non-processed time series (Fig. 1). Notice that it is not necessary that the extrapolation faithfully reproduces the characteristics of the original series; it just suffices for our purpose that the extrapolated time series approximately continues the cyclic structure at the original time series boundaries in order to reduce spectral leakage. Apart from tracking trends, the exact extrapolation beyond the boundaries is not essential since its effects on the central part of the Hilbert transform are very marginal at most.

After the described complexification of the original time series, we follow the steps of standard MCA, with the difference that the transpose $^\mathsf{T}$ incorporates the complex conjugate * and the obtained EOFs and projections PCs are complex and unitary. This allows us to calculate the spatial amplitude A_s and phase function Θ_s for both fields, $s = A, B$,

$$A_s = (V_s \odot V_s^*)^{\odot -\frac{1}{2}}, \tag{5}$$

$$\Theta_s = \arctan2\left(\mathrm{Im}(V_s) \oslash \mathrm{Re}(V_s)\right), \tag{6}$$

where $\odot$ and $\oslash$ denote the element-wise multiplication/exponentiation and division, and arctan2 refers to the two-argument arctangent (Appendix 6). Although this matrix notation seems somewhat cumbersome compared to the more direct expression through scalar fields, it allows us to be coherent with the rest of the paper. The phase function can be interpreted directly as a time lag if the corresponding (real) PC has a narrow-band spectrum with just one dominant frequency. If the spectrum is rather broad-band or has several dominant frequencies, an interpretation of the phase function is usually not straightforward. We note that the EOFs derived from the SVD are only defined up to the sign. Therefore, the phase given by Equation (6) is also only defined up to a factor of π. However, a change in sign will always be reflected in the projected PCs, so that taking into account both, PC and phase function, the results are unambiguous.

2.3 Rotated MCA

While orthogonality is often a mathematically desirable property, it does not make a lot of sense from a purely geophysical standpoint. Therefore, standard EOFs are difficult to interpret in the case of geophysical data. The two major drawbacks of EOFs due to orthogonality are (i) they are dependent on the selected domain shape and (ii) they tend to split certain geophysically meaningful patterns across several consecutive modes [10]. To relax the orthogonality constraint to better accommodate the geophysical reality, the EOFs can be *rotated*, which implies a linear transformation of the first r loaded[3] EOFs $L_{s,r}$. This concept, which was originally developed

3 Loaded EOFs are weighted by the square root of the corresponding singular value.

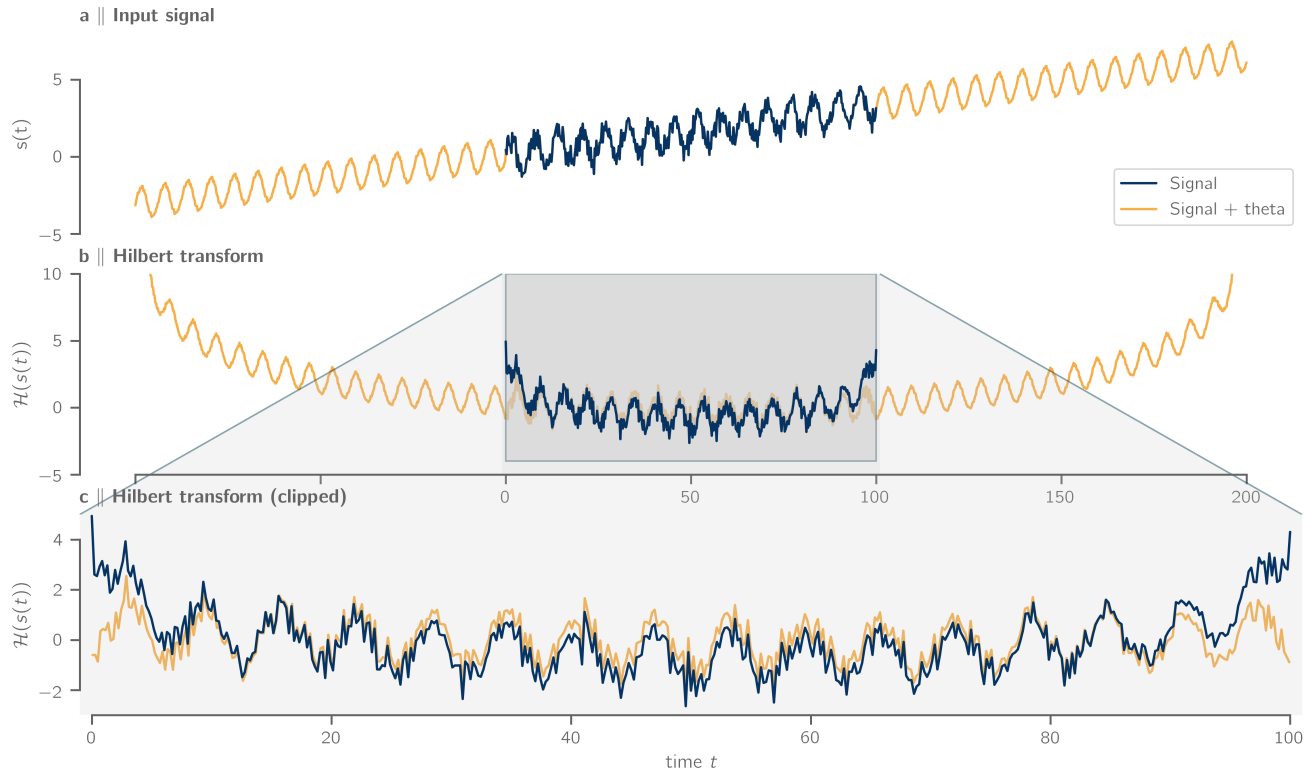


Figure 1: Example illustrating the Hilbert transform using the Theta extension. **(a)** Input signal $s(t) = \sin(t) + ct + \epsilon$ using an arbitrary constant c and Gaussian white noise ϵ (blue) as well as the extended time series via forecasting/backcasting using the optimised Theta model [16]. **(b)** Hilbert transform $\mathcal{H}(\cdot)$ of the original and extended signal, respectively. **(c)** Comparison of both Hilbert transforms over the domain of the original series.

in the context of PCA, can also be applied to MCA [11]. For this, we apply the rotation matrix $R \in \mathbb{C}^{r\times r}$ to the *loading matrix* $L_r \in \mathbb{C}^{n\times r}$, $n = n_A + n_B$,

$$L_r = \begin{pmatrix} L_{A,r} \\ L_{B,r} \end{pmatrix} = \begin{pmatrix} V_{A,r} \\ V_{B,r} \end{pmatrix} \Sigma_r^{1/2}, \tag{7}$$

where r reflects the respective submatrices containing only the first r columns.

There are a number of different criteria for defining the rotation matrix R (see e.g. [10]), including the Varimax orthogonal rotation [18] and the Promax oblique rotation [19], whose general aim is to regroup the obtained patterns by approximating *simple structures* [20, 10]. Mathematically, Varimax rotation seeks to maximise the summed variances of squared loadings which is achieved by (i) restricting rotated EOFs to be composed by only a few numbers of grid points with high loadings while the remaining grid points exhibit near-zero loadings and by (ii) limiting each grid point to contribute to only one rotated EOF while having near-zero loadings for the other EOFs. This basically transforms the usually dense subspace obtained from MCA into a more sparse solution leading to spatially compact structures which allow a clearer interpretation. Promax oblique rotation builds upon the Varimax solution by raising the rotated, normalised EOFs to the power $p \geqslant 1$ while retaining the original sign, thus further reducing low loading compared to high loading of the EOFs. Promax can be understood as an oblique generalisation, with $p = 1$ yielding a Varimax orthogonal solution. Reference [19] provides a value for p, which the authors consider appropriate for most applications ($p = 4$). In the extensive review in Ref. [10] points out, however, that the Promax rotation using $p = 2$ consistently performs better, which is what we will use in this paper. In order to keep the paper self-contained, we provide a brief summary of both rotation criteria in Appendix 7.

The main difference between both rotation types is, that Promax allows rotated PCs to be correlated with each other, with higher values of p typically leading to stronger correlations. In contrast, Varimax solutions yield always uncorrelated PCs. For both, Varimax orthogonal and Promax oblique rotation, the obtained EOFs are no longer orthogonal. The question which rotation method is the most suitable for a given analysis remains unsettled in the literature. In reality, we do not expect geophysical signals to be perfectly uncorrelated, which generally argues for applying an oblique rotation. Nevertheless, Ref. [21] has shown that Varimax orthogonal and Promax oblique solutions perform similarly, in particular when the PCs obtained by the oblique solution exhibit low linear Pearson correlations (e.g. < 0.15). In the presence of simple structures, however, Promax oblique rotation performs better by effectively reducing the number of grid points that contribute to each mode, hence further simplifying the EOFs and increasing correlations among the PCs [21]. Therefore, the decision on how many EOFs to rotate and which rotation type to perform is a choice to be taken case-by-case and which we will explore in Sec. 4.2.

3 DATA

To test our method, we apply it to artificial and real climate data. For the artificial data sets, we consider complex MCA without rotation, as studies already exist that demonstrate the better interpretability and lower sensitivity to sampling errors of the Varimax-rotated solutions compared to the unrotated EOFs. [22, 10, 23]. By means of two synthetic experiments we seek to illustrate the advantages and caveats of complex MCA. In a first experiment (Experiment I), we test the performance of complex MCA compared to standard MCA considering time-lagged signals. In a second experiment (Experiment II), we investigate how the Theta extension can improve the response of complex MCA to non-stationary processes. Finally, we apply complex MCA with rotation to climatic data that we expect to have intrinsic geophysical cycles but are also affected by the non-stationarity of climate change.

Table 1: Models for data generation of X_s for experiment I and II. In order to obtain a 3D data cube for time t, longitude λ and latitude φ, we stack 10 of the obtained (t, λ)-fields. For our final model, we consider the coordinate ranges $t = [0, 6\pi]$ and $\lambda = [0, 2\pi]$ with $\lambda_{max} = 2\pi$. Gaussian white noise is denoted by $\epsilon(t, \lambda) = \mathcal{N}(0, 1)$.

				Experiment I		Experiment II
Field s	$X_s(t, \lambda)$	$c_s(\lambda)$	a_s	$\zeta_s(t, \lambda)$	$\tau_s(\lambda)$	$\zeta_s(t, \lambda)$
A	$c_A(\lambda)\zeta_A(t, \lambda) + \epsilon(t, \lambda)$	$\sin^2 a_A\lambda$	$3.8 * (2\pi)/\lambda_{max}$	$-\cos(t + \tau_A)$	$2\pi\lambda/\lambda_{max}$	$(-2\lambda/\lambda_{max} + 1)\, t$
B	$c_B(\lambda)\zeta_B(t, \lambda) + \epsilon(t, \lambda)$	$\sin^2 a_B\lambda$	$2.4 * (2\pi)/\lambda_{max}$	$-\cos(t + \tau_B)$	$\tau_A + \pi/2$	$(+2\lambda/\lambda_{max} - 1)\, t$

3.1 Synthetic Data

We create two 3D spatio-temporal data fields X_A, X_B with dimensions representing longitude λ and time t adding the third spatial dimension by stacking multiple (λ, t) layers. The data generation model follows

$$X_s(t, \lambda) = c_s(\lambda)\zeta_s(t, \lambda) + \epsilon(t, \lambda), \quad s = A, B \tag{8}$$

with a "zonal" modulation factor $c_s(\lambda) = \sin^2 a_s\lambda$, where a_s can be interpreted as a spatial decorrelation factor of field s, a temporal signal propagating in the $-\lambda$ ("zonal") direction $\zeta_s(t, \lambda)$, and Gaussian white noise $\epsilon(t, \lambda)$ with zero mean and variance of 1.
The idea of our first experiment is to highlight the advantage of using complexified fields compared to normal MCA in the presence of moving patterns and phase-shifted, stationary fields. Therefore, we define both signals as travelling waves $\zeta_s(t, \lambda) = -\cos(t + \tau_s(\lambda))$ assigning a different phase shift $\tau_A = 2\pi\lambda/\lambda_{max}$ and $\tau_B = \tau_A + \pi/2$ to the first and second field, where $\lambda_{max} = \max\lambda$ (Fig.2a).
In our second experiment, we investigate the response of complex MCA to non-cyclical, non-stationary signals. Since MCA seeks to maximise covariance through a new set of linear combinations, we restrict ourselves here to linear trends given by $\zeta_{1/2}(t, \lambda) = \pm(-2\lambda/\lambda_{max} + 1)\, t$. For field A, the signal basically represents a positive trend for low values of λ, gradually decreasing towards a negative trend for high values of λ. For field B, the trend structure is reversed with a negative trend for low λ and a positive trend for high λ (Fig. 3a). For a better overview, the signals of both experiments are summarised in Tab. 1.

3.2 Climate Data

We analyse monthly global sea surface temperature (SST) and continental precipitation using the extended ERA5 data set from 1950-2019 [24] provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) as a state-of-the-art replacement of the ERA-Interim reanalysis [25]. In general, trends and low frequency variability of surface temperature and humidity perform well for 1979-2019 [24, 26]. In contrast, the period from 1950 to 1978 suffers from less abundant observational data potentially reducing data quality. From data inspection, we found an unusual increase in precipitation over East Africa for the early 50s, which was found to be related to issues in the model initialisation process of soil moisture in some deep layers in that time period [26]. Since the start and end point of the time series are of essential significance, in that they form the starting point for the Theta extensions, we remove the first 10 years, providing us data from 1960 to 2019. Furthermore, the SVD of the covariance matrix is a rather memory intensive numerical operation, which is why we limit the domain of interest from 40° S to 60° N with a 1°x1° spatial resolution, guaranteeing most of the continents to be included into the analysis. In total, the fields of SST (excl. continents) and precipitation (excl. sea) cover $n_A = 23\,500$ and $n_B = 12\,860$ time series, respectively. To align the different temporal scales of highly variable precipitation and slow-varying SST and to filter out the high frequency signals, we smooth both data sets with a 6-month moving average, for each month taking into account the 3 preceding and the 2 following months. We choose this particular time window in favour to an typical odd-numbered smoothing window to effectively filter out the biannual cycles present in the geophysical observations without .We normalise both sets by dividing each grid point by its temporal standard deviation to give equal spatial importance. In a non-standardised MCA analysis, regions of high rainfall, like the tropics, would dominate the covariance patterns,

since annual rainfall in the mid-latitudes and subtropics is typically much lower. After normalisation, we weight the data points located on the regular 1°x1° grid according to their associated area on a sphere [27] by multiplying each grid point with $\sqrt{\cos(\varphi_j)}$, φ_j being the latitude at grid point j.

Monthly climate indices, used in this study for the sake of comparison, are downloaded from the websites of the National Oceanic and Atmospheric Administration (NOAA; https://psl.noaa.gov/data/climateindices/list/) and of the Japan Agency for Marine-Earth Science and Technology (JAMSTEC; http://www.jamstec.go.jp/virtualearth/general/en/index.html). All indices are smoothed using a 6-month moving average in alignment with our data pre-processing. We further center and max-normalise all indices for better comparison with our obtained PCs.

4 RESULTS

In the following, we discuss the results from the synthetic experiments before investigating the results of the analysis of SST and continental precipitation.

4.1 Synthetic Data

EXPERIMENT 1: LAGGED SIGNALS The first mode of standard MCA (denoted by subscript std) describes 14.7% of the shared covariance, followed by the second mode with 14.4% (Fig. 2b). We recognise that the spatial patterns described by EOF 1 represent only part of the travelling waves of field A and B, with another part of it given by EOF 2 (Fig. 2a,b). However, this is not surprising: (i) Standard MCA seeks to maximise temporal correlation between field A and B. Considering that a phase shift of $\pm\pi/2$ relates the trigonometric functions via[4] $\cos\left(\alpha \pm \frac{\pi}{2}\right) = \mp \sin\alpha$, the two functions $\cos\alpha$ and $\cos\left(\alpha \pm \frac{\pi}{2}\right)$ are uncorrelated over full cycles since both form two orthogonal base vectors. In such a situation, standard MCA will be unable to extract the physically meaningful signals because the sine and cosine modes are uncorrelated. (ii) In our experiment, both fields contain phase shifts covering one entire period of 2π. Consequently, for each time-lagged signal in field A, there is a correlating signal in field B, such that the entire range of signals can be detected by standard MCA. Orthogonality of the singular vectors and respective uncorrelation of the PCs forces the patterns to be separated into two distinct modes following a sine and a cosine, which are perfectly uncorrelated for a full number of cycles and thus orthogonal. The remaining modes do not show any more distinct patterns, basically representing noise.

In comparison, complex MCA (denoted by subscript com) reliably shows the underlying (lagged) cosine for both amplitude fields of A and B (Fig. 2c). We have not used the Theta model extension here, since the signals in this experiment consist of integer periodic time series. The respective phase functions both indicate decreasing phase shifts with increasing longitude, clearly indicating a moving pattern. The regions which correlate with the real part of the PC (Fig. 2c; center) are depicted in blue while regions exhibiting an anti-correlating signal (phase shifted by $\pm\pi$) are shown as red. In that sense, the "colours" white and black represent signals with a phase shift of $\pm\pi/2$. The described correlation of mode 1 is 22.8%. It makes more sense though, to compare the actual singular values against each other, since they provide an absolute measure of the described covariance. We note that $\sigma_{com,1} > \sigma_{std,1} + \sigma_{std,2}$ (Tab. 2) indicating that the complex mode 1 accounts for more correlation than the combined mode 1 and 2 of the standard solution. This makes sense, since EOF_s 1 and EOF_s 2 are reduced in regions which are phase shifted with respect to the associated PC_s 1 and PC_s 2. Although in general the singular values of complex MCA tend to be larger than their non-complex counterparts even for purely Gaussian white noise signals, we stress that the difference between $\sigma_{com,1}$ and $\sigma_{std,1} + \sigma_{std,2}$ is much larger than the observed increase for white noise (Tab. 2). As a pitfall of complex MCA, however, it should be noted that regions whose spatial amplitude function is very low, tend to have noisy phase function values. In our experiment, this is obvious for the different interspaces of both fields where $c_s(\lambda) \approx 0$ and the associated phase

4 This is a special case of $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$.

Table 2: Singular values of MCA (Standard), complex MCA (Complex) and Theta extended complex MCA (Complex + Theta) obtained for purely white noise random fields (Noise) and the synthetic experiments described in Section 4.1 (Experiment I and II). Covariance fraction γ_k for mode k (in % shown within parenthesis) is calculated according to Eq. (3). Some exemplary modes (**bold**) are depicted in Fig. 2 and 3. Total cumulated covariance is shown in the last row.

	Noise		Experiment I		Experiment II		
Mode	Standard	Complex	Standard	Complex	Standard	Complex	Complex + Theta
σ_1	13.7 (3.6)	41.6 (4.1)	**73.0** (14.7)	**288.4** (22.8)	2852.8 (88.7)	**5712.6** (85.4)	**2925.8** (74.9)
σ_2	12.4 (3.3)	37.9 (3.7)	**71.2** (14.4)	38.6 (3.1)	13.4 (0.4)	39.7 (0.6)	39.8 (1.0)
σ_3	12.4 (3.3)	36.6 (3.6)	13.2 (2.7)	37.6 (3.0)	12.7 (0.4)	38.9 (0.6)	38.9 (1.0)
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$\sum_j \sigma_j$	376.1	1015.1	495.4	1265.0	3215.9	6692.7	3905.4

function does no longer correspond to the global, migrating tendency. In general, it is therefore advisable not to consider regions with low amplitudes.

EXPERIMENT 2: TRENDS AND NON-CYCLIC SIGNALS As discussed in Sec. 2.2, non-periodic behaviour usually leads to spectral leakage at the boundaries. In our experiment, this effect is clearly evident for both PCs of the complex MCA (denoted by subscript com) (Fig. 3b). The Theta extended complex MCA (denoted by subscript thc), on the other hand, successfully mitigates the boundary effects of the PCs (Fig. 3c), where we have set the Theta period $T_s = 1$ as the time series have no seasonality. Nevertheless, the spatial patterns are similar for both methods and can hardly be distinguished visually. The amplitudes in the left and right edges of the EOFs show the strongest signal, which weakens and disappears towards the centre, where the signal can not be distinguished from the noise $\epsilon(t,\lambda)$ since $\zeta_s(t,\lambda \approx \lambda_{max}/2) \approx 0$ for all $t \in [0, 6\pi]$ (Fig. 3a). In our example, the phase functions clearly show correlating (phase shifted by 0; blueish) and anti-correlating (phase shifted by $\pm\pi$; redish) regions represented by $PC_s(\theta = 0)$ and $PC_s(\theta = \pi)$ (Fig. 3a) where θ denotes the phase shift applied to the complex PC. Although the Examining the singular values, we notice that $\sigma_{com,1} >> \sigma_{thc,1} \approx \sigma_{std,1}$ (Tab. 2), indicating the increased covariance due to the boundary effects of the non-treated Hilbert transform . Notice that this implies that the boundary effect created by the Hilbert transforms have a strong influence on the existing correlations and, depending on their magnitude, can be reflected in one of the first modes, thus completely distorting the obtained modes and misleading the interpretation of the results.

More generally, any trend has a broadband frequency spectrum and thus the analytical signal constructed by the Hilbert transform has not a physical interpretation in terms of characteristic frequencies. Therefore, the phase function cannot be interpreted in terms of a physical phase shift. The only exception is for $\theta = 0; \pm\pi$ (corresponding to a phase function depicted as blue/red) since relative phase shifts of that order basically represent correlating and anti-correlating signals disregarding the mathematical nature of the phase. A fundamental consequence of this is that for modes whose PC is broadband (e.g. a trend), only correlating ($\theta \approx 0$) as well as anti-correlating ($\theta \approx \pm\pi$) patterns should be considered. Furthermore, this aspect can also lead to an (non-meaningful) increase of the singular value and thus of the relative importance of the mode if the proportion of non-integer phase-shifted signals of this mode is high.

4.2 Climate Data: SST & Continental Precipitation

We apply Theta extended, complex MCA to SST and continental precipitation using a Theta period $T_s = 12$ to account for the seasonal cycle. The dimensionality of the problem can be greatly reduced with the first 72 modes describing 99% of the existent (lagged) covariance (Fig. 41). In order to simplify the obtained patterns and to increase the physical meaning of the results (Sec. 2.1), we rotate the first 150 modes representing 99.82 % of the existing lagged cross-covariance. Our decision to rotate 150 modes is motivated by the idea of retaining

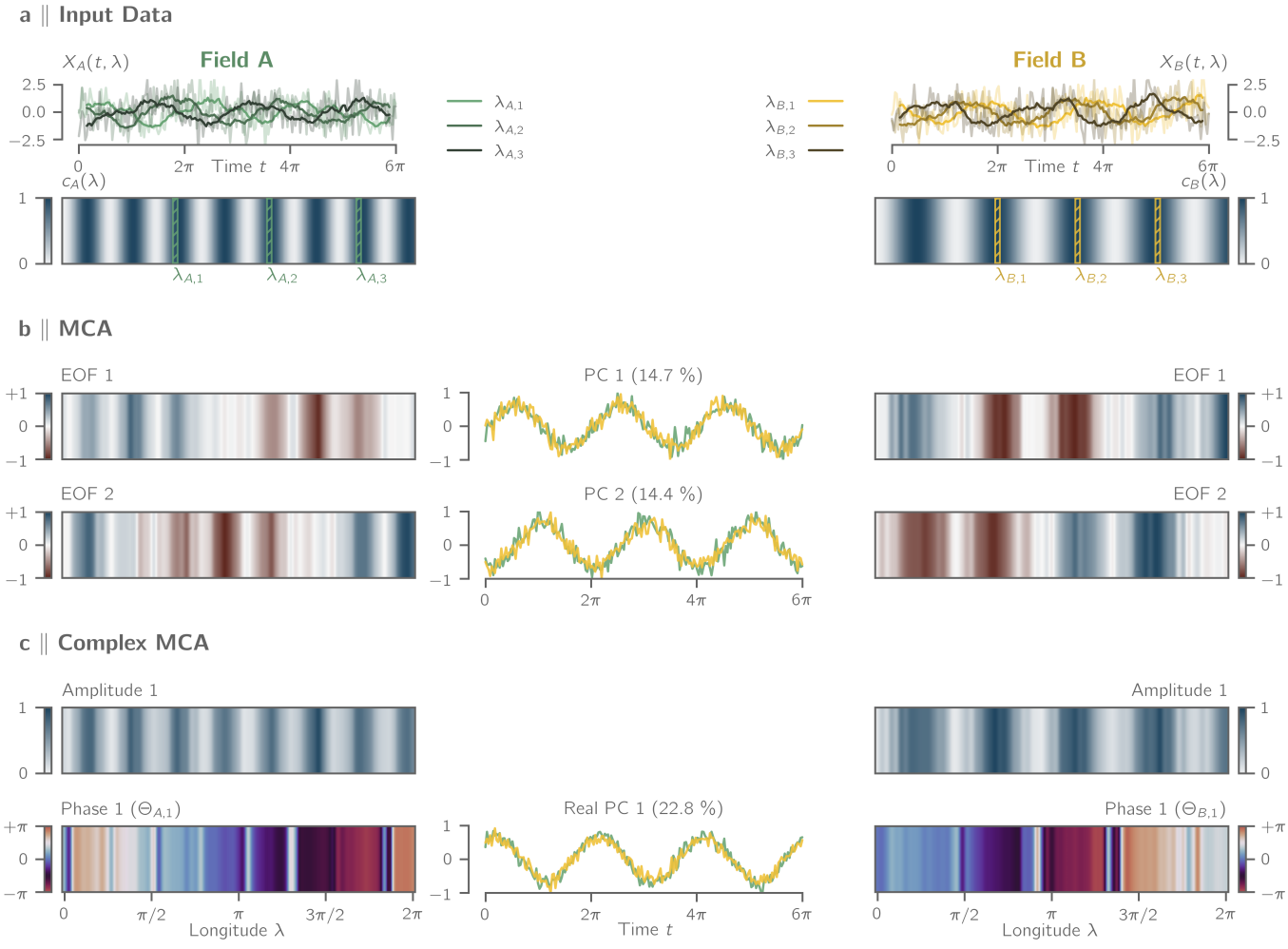


Figure 2: Results of Experiment I. **(a)** Illustration of synthetic data fields A (left) and B (right) showing the temporal evolution $X_i(t, \lambda)$ for some selected longitudes λ (grey) and the smoothed 10-step moving average (greenish, yellowish), respectively, (top) and the associated damping patterns $c_i(\lambda)$ (bottom) for both fields, respectively. **(b)** Mode 1 and 2 of the standard MCA solution showing the EOFs of field A (left) and field B (right) as well as their respective PCs (center). **(c)** Mode 1 of the complex MCA solution showing the spatial amplitude (top) and phase functions (bottom) for field A (left) and B (right) as well as the real part of the corresponding PCs (center). The phase function indicates the phase shift that has to be applied to the time series of a specific grid point in order to obtain the non-shifted PC (shown in center bottom). Percentages in parenthesis represent the covariance fraction γ_k. All EOFs and spatial amplitude functions are max-normalised for the sake of readability.

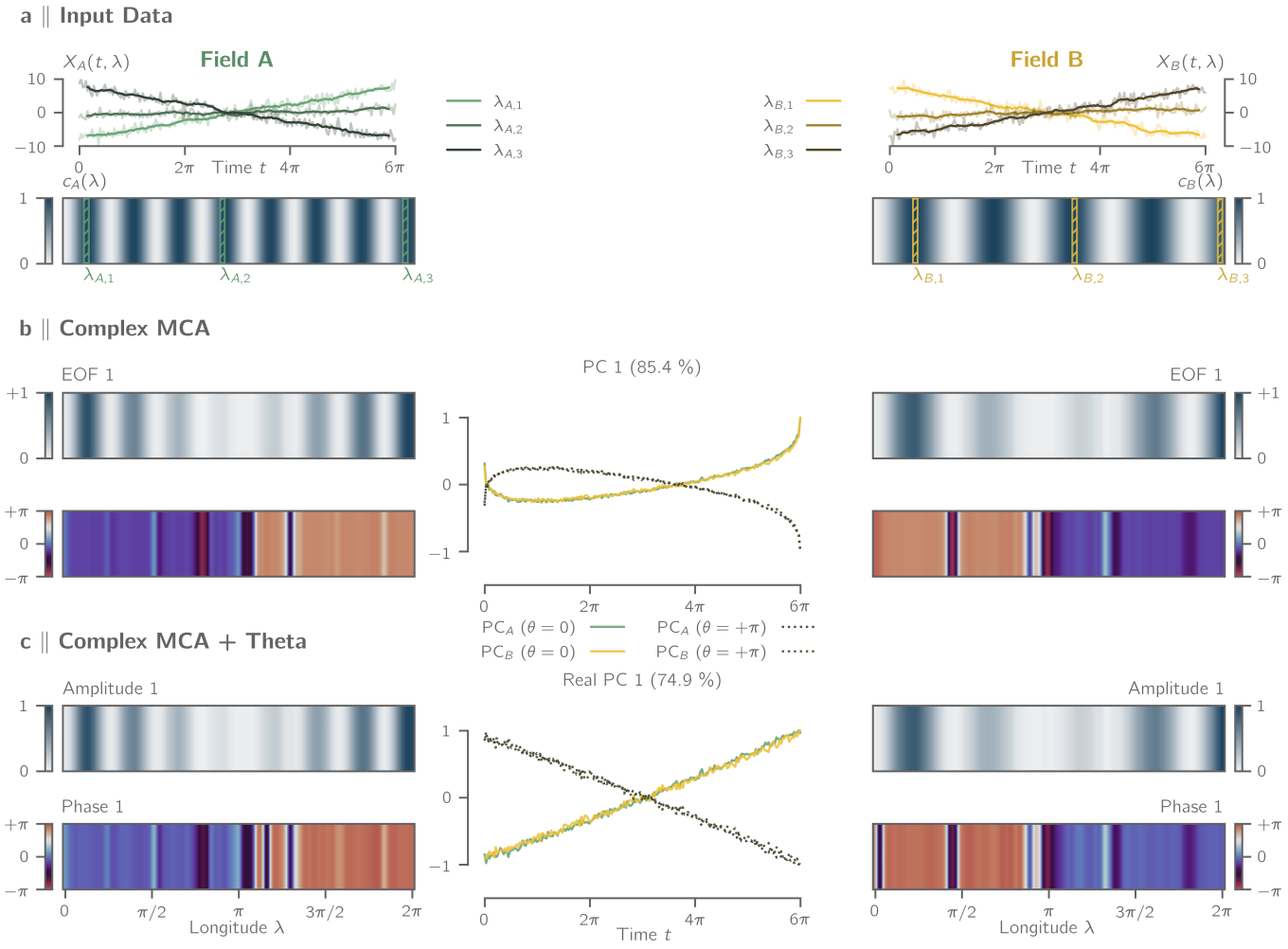


Figure 3: Results of Experiment II. **(a)** Illustration of synthetic data fields A (left) and B (right) showing the temporal evolution $X_s(t, \lambda)$ for some selected longitudes λ (grey) and the smoothed 10-step moving average (greenish, yellowish), respectively, (top) and the associated damping patterns $c_s(\lambda)$ (bottom) for both fields, respectively. **(b)** Mode 1 of the complex MCA solution showing the spatial amplitude (top) and phase functions (bottom) for field A (left) and B (right) as well as the real part of the corresponding PCs (center). **(c)** As in **(b)** but for Theta extended complex MCA. Percentages in parenthesis represent the covariance fraction γ_k. All spatial amplitude functions are max-normalised for the sake of readability. Phase functions show regions correlating with PC_s $1(\theta = 0)$ (blueish) and PC_s 1 $(\theta = \pi)$ (redish).

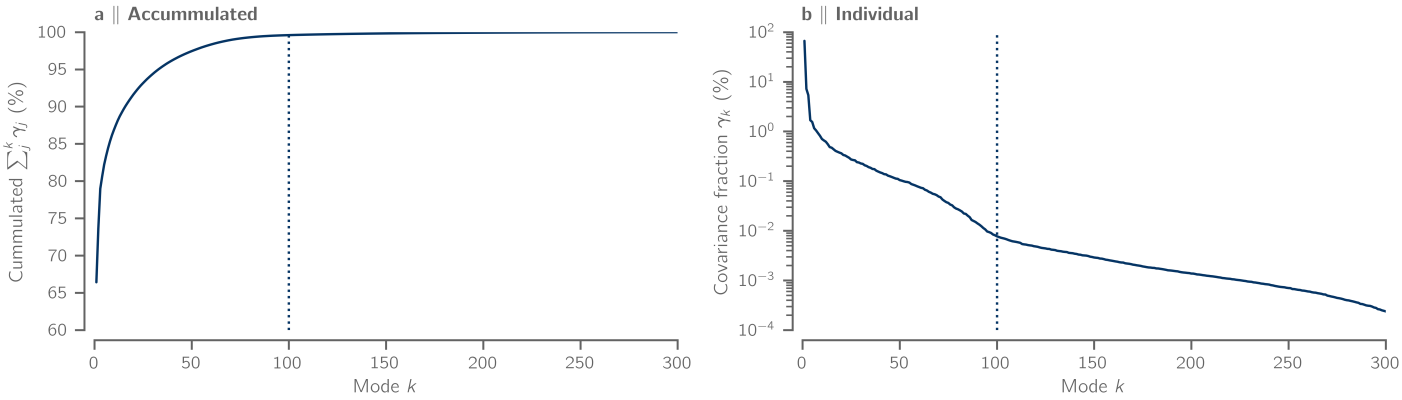


Figure 4: Described covariance of the first 300 modes of complex MCA applied to SST and continental precipitation showing the (a) individual and (b) accumulated share. Dashed line estimates the boundary of suspected "noisy" modes.

as much information as possible without including the noise which dominates the higher modes. Since we observe a slight drop in singular values at about the mode number 100 followed by an exponential decrease in singular values (Fig. 4b), we guess that there is a negligible information content in higher modes. Moreover, the quality of the reconstructed signal is rather independent of the exact number of rotated modes, with 150 ± 50 yielding basically identical results for the first 6 modes for both Varimax orthogonal and Promax $p = 2$ oblique rotation. Therefore, we will restrict our discussion in the following to the first 6 modes. In order to address the question of the rotation method to be chosen, we note that Promax oblique rotation performs better when simple structures are present and correlations among PCs are high [21]. We expect our first mode to be dominated by the shared dynamics of the seasonal cycle of both SST and continental precipitation, which is indeed what we find for both Promax (not shown here) and Varimax solution (Fig. 5a). This mode, however, is fairly global and as such does not represent a simple structure. Furthermore, we observe that the correlations among the first six Promax-rotated PCs range from -0.15 to 0.13 only, underlining that the Promax oblique solution is close to orthogonal and differences to the Varimax solutions only marginal, at least for the first six modes. Since the Promax oblique solution seemingly does not provide a better results, we opt for the somewhat simpler Varimax orthogonal rotation. In the following, we will discuss the Varimax-rotated modes by investigating the real part of the PCs, the spatial amplitude and phase functions for both fields, SST and continental precipitation, respectively.

DERIVED COMPLEX PATTERNS In our representations of the modes, we remove non-significant, "noisy" phase values (see Sec. 4.1) by masking out regions in the spatial amplitude and phase function exhibiting a max-normalised amplitude of < 0.25.

Mode 1 describes 62.7 % of the covariance between SST and continental precipitation clearly showing the annual cycle (Figs. 5a and 9a). As expected, the annual cycle shows itself in both variables on a global scale, with the exception of the equatorial ocean, where the seasonal SST variations are only weak. The phase function correctly identifies the anti-correlation between the northern and southern oceans. It also suggests that the eastern equatorial Pacific and the equatorial Indian Ocean nevertheless show a weak seasonal signal, which is, however, positively phase shifted relative to the rest of the southern ocean. On the continents, precipitation dominates mainly in monsoon areas and the tropics. The phase function illustrates the division into boreal and austral summer rainfall dominated regions and identifies corresponding transition zones, as e.g. over the South American rain forest and central North America. It also highlights some interesting dynamical regions which stand out of their respective environment, namely, the West and East coast of North America, the Mediterranean region, as well as the East Asian monsoon region in China.

The obtained mode provides an instructive example to highlight the benefits of the complexified approach due to the mode's corresponding narrow-band frequency spectrum (Fig. 9a). Using the dominant periodicity of

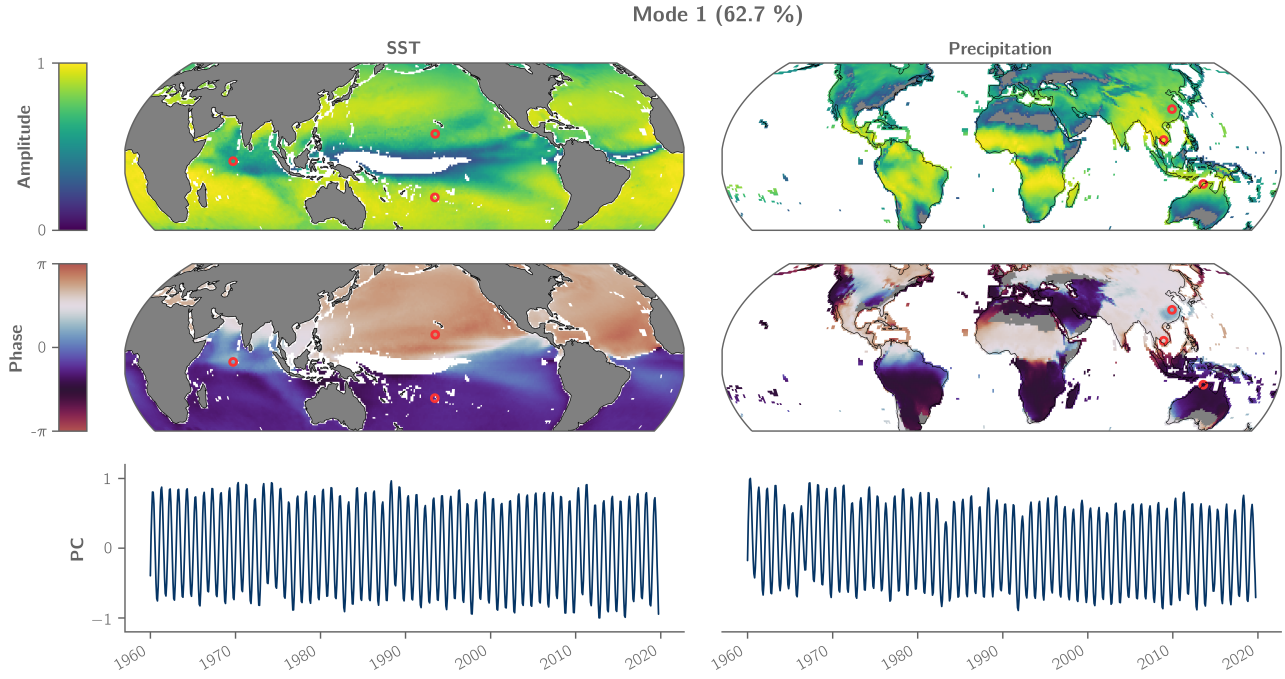
mode 1, $T_1 = 12\,\text{mon}$, the interpretation of a phase shift θ, given by the spatial phase function, as time lag τ is straightforward and can be computed via $\tau = \frac{T_1}{2\pi}\theta$. Doing this for some exemplary locations (denoted by ○ in Fig. 5a), we observe that the seasonal SST maximum of the North Pacific follows the one in the South Pacific by 196 days, that is, being almost perfectly anti-correlated (Fig. 6). The relative time shift of 81 days of the SST maximum in the Indian Ocean and 86 days at Poyang lake, China, is close to 3 months and as such translates to a phase shift of about $-\pi/2$, something that can not be picked up as feature within one mode when using standard MCA. It should also be noted that although the sampling frequency of SST and precipitation is monthly, the phase function is continuous and thus allows to infer time lags on shorter time scales. This is, for instance, the case for the seasonal precipitation maximum in Darwin, Australia, which precedes the seasonal SST maximum of the South Pacific by 27 days.

Mode 2 (7.0 %) can be clearly associated to the ocean-atmosphere phenomenon of El Niño - Southern Oscillation (ENSO) (Fig. 5b and 9b). During *El Niño*, higher SST in the central and eastern Pacific and lower SST in the western Pacific positively correlate with heavier rainfall within a narrow band along the west coast and the southeastern coast of South America [28], the east and west coasts of North America [29], the East Asian monsoon region [30] and the Horn of Africa [31]. At the same time, precipitation decreases in northern South America [28], Oceania [32], South Africa [33] and in the Indian monsoon region [34]. During *La Niña* (the counterphase of *El Niño*), these correlations are reversed.

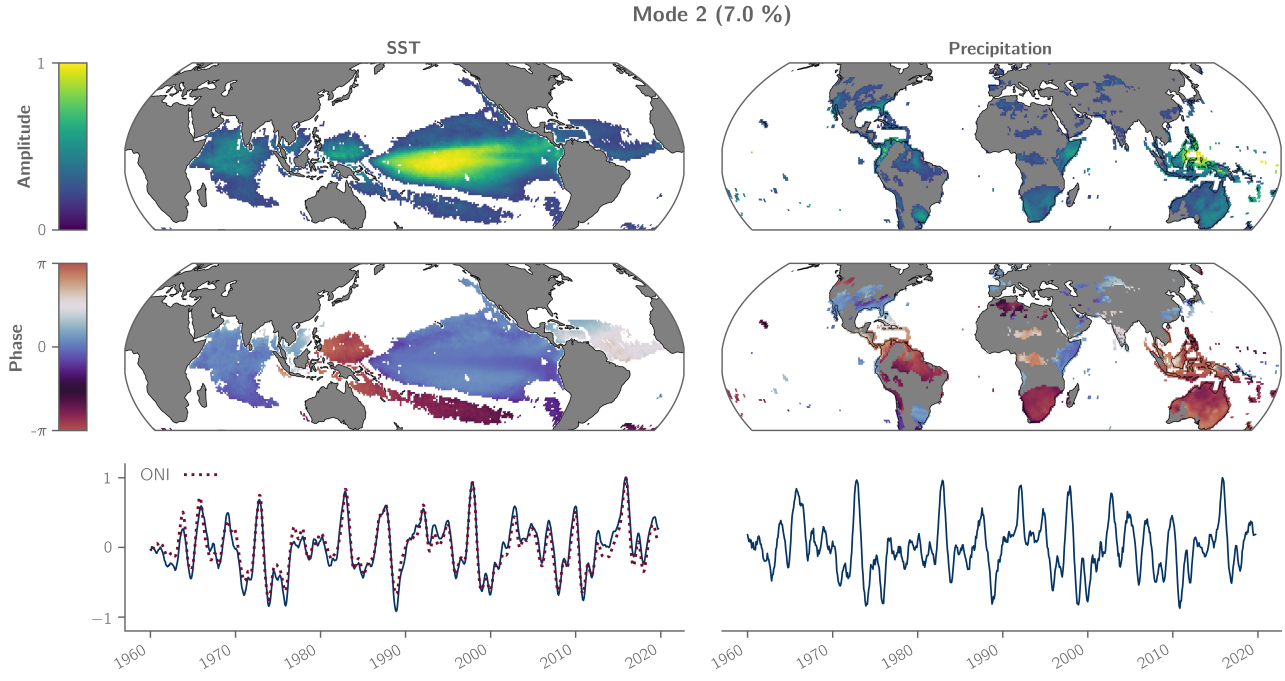
Our result also shows the interrelationship of Pacific ENSO with the Indian Ocean [35], the South China Sea [36] and the Tropical North Atlantic [37, 38, 39, 40] in accordance with previous studies. Interestingly, it was found that the ENSO related SST teleconnections in the remote oceans often occur with some delays, with the Indian Ocean typically peaking ~ 3 months and the South China Sea and tropical North Atlantic ~ 5 months after the ENSO SST [37, 36, 38]. The mechanism behind these lagged responses, known as the atmospheric bridge, is based on the characteristic atmospheric circulation during El Niño which causes changes in cloud cover and evaporation over the remote oceans, leading to increased net heat flux and SSTs [41, 36]. However, due to the broadband frequency spectrum of the SST PC (Fig. 9b), with most of the energy contained at four different peaks around 2.5, 3.5, 5 and 11 years, the phase cannot simply be translated into a time shift. Nevertheless, the tropical North Atlantic clearly exhibits more positive phase values compared to the Pacific El Niño region, therefore indicating to be positively time shifted relative to the Pacific.

Mode 3 (6.9 %) represents global warming and the associated changes in precipitation patterns (Fig. 7a). The warming SST patterns clearly emerge in all major ocean basins, although more pronounced in the northern hemisphere due to the asymmetric response of the northern and southern trade winds to global warming [42]. We also note a pronounced warming of the western part of both, the Pacific and the Atlantic basins, both regions of enhanced ocean heat transport [42]. Similar to these oceanic trends, we also observe global trends in the precipitation patterns, with decreasing rainfall over the Mediterranean, South Africa, Australia, South America and parts of western North America. There seems to be also a decrease of rainfall over the west Asian monsoon region. In contrast to that, the results suggest increased precipitation over the Indian monsoon region as well as some localised regions in Europe, South and North America. These results are largely in agreement with studies based on observational data [43] and, more recently, on CMIP5 climate simulations [44]. Finally, it should be stressed, that both PCs, SST and precipitation, do only provide a meaningful interpretation for phases $\theta \approx 0; \pm\pi$ (correlation, anti-correlation), due to their broadband frequency spectra (Fig. 9c). For phase shifts different from that, no clear conclusions can be drawn, as it is the case e.g. for equatorial Africa where the phase shift is approximately $+\pi/2$.

Mode 4 (1.6 %) shows a slow-oscillating pattern of SST in the northeastern Pacific correlating with localized precipitation patterns distributed over all continents (Fig. 7b). The typical spatial SST pattern is known as the Pacific Decadal Oscillation (PDO) [46] and a well-established climate index. A combination of various processes originating in the tropics and extra-tropics has been proposed as the physical source of the PDO [47], with ENSO and PDO likely responding to the same forcing function [48]. Our analysis, however, provides a mean to disentangles ENSO and PDO related precipitation patterns, which are often similar, though reveal important differences e.g. for Australia, the Indian subcontinent or the African Sahel region. Yet, care must be



(a) Mode 1 representing the seasonal cycle. Red circles (○) mark North Pacific (170° E, 15° N), Indian Ocean (70° W, 0° N) and South Pacific (160° E, 20° S) for SST and Poyang lake, China (116.3° W, 29.1° N), Phnom Penh, Cambodia (104.9° W, 11.6° N) and Darwin, Australia (130.8° W, 12.5° S) for precipitation. The phase shifted PCs of these locations are examined in Fig. 6.



(b) Mode 2 compared to the Oceanic Niño Index (ONI; ···) provided by NOAA Climate Prediction Center as described in Section 3.2.

Figure 5: Complex Varimax-rotated MCA of SST (left) and continental precipitation (right) showing **(a)** Mode 1 and **(b)** Mode 2 with their relative importance indicated by the covariance fraction, respectively. The amplitude functions (top) show the regions predominantly contributing to the mode of the respective variable. The phase functions (centre) depict the relative phase shifts with respect to the corresponding PC (bottom), where 0 (blue) means correlation and $\pm\pi$ (red) shows anti-correlation. For each grid point, the corresponding PC can be computed from the given phase value by applying the negative phase shift to the PC of phase 0 (bottom). The amplitude functions, PCs and indices shown are all max-normalised for the sake of comparability. In both amplitude and phase function regions with a max-normalised amplitude below 0.25 are masked out.

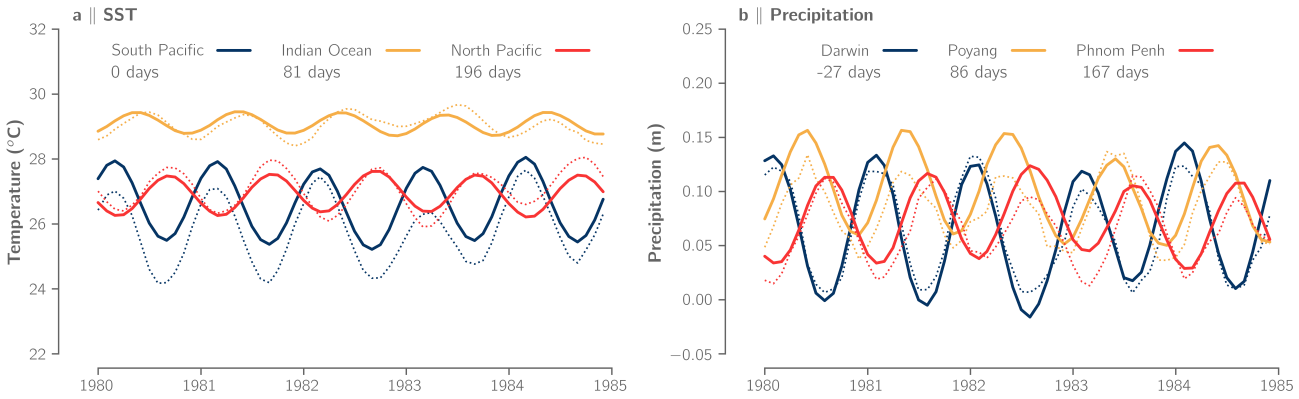


Figure 6: Comparison between 6-month moving average of ERA5 SST and continental precipitation (dotted lines) and the reconstructed time series based on mode 1 only (thick lines) for the exemplary locations defined in Fig. 5a covering the years 1980 to 1985. Days in legend refer to the time shift τ of the individual locations with respect to the SST of the South Pacific derived from the spatial phase functions (Fig. 5a.)

taken when interpreting regions which have a phase shift different from $0, \pm\pi$. Although the PDO exhibits a strong spectral energy at about 35 yr, the mode contains also important features at about 1 yr to 4 yr (Fig. 9d) making the phase interpretation physically less clear.

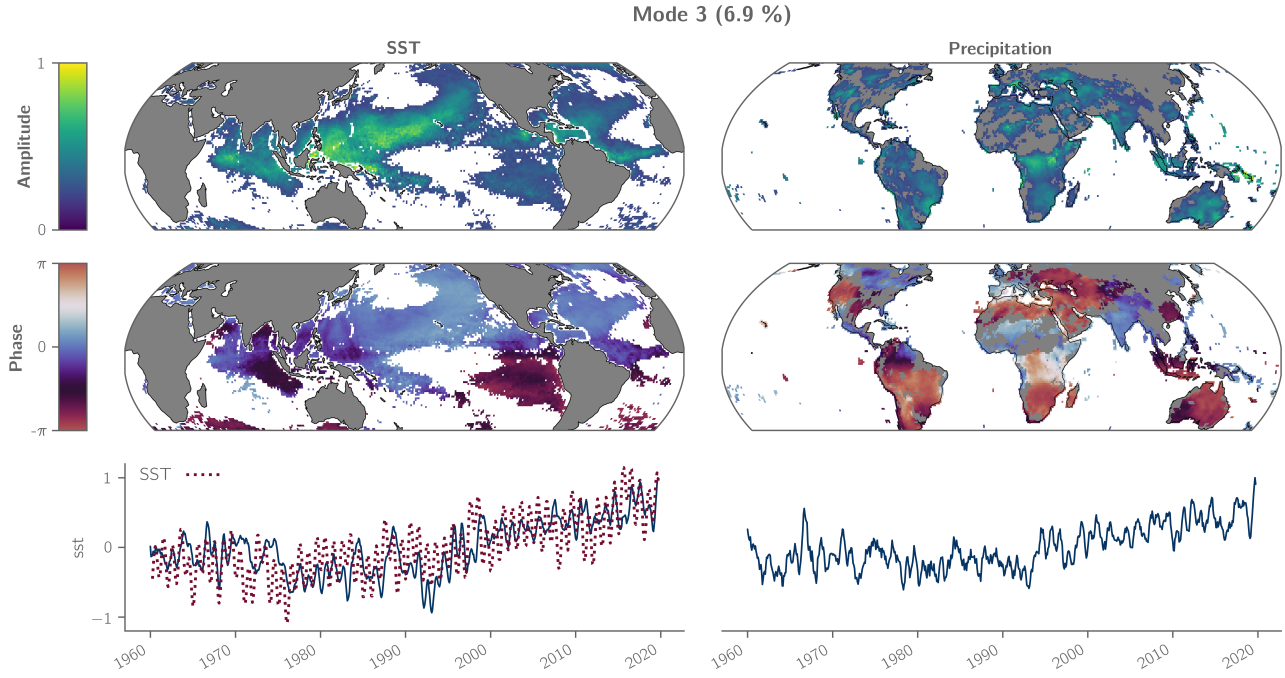
Mode 5 (1.6 %) describes an oscillating SST anomaly mainly limited to the central Pacific (Fig. 8a and 9e) describing El Niño Modoki [49] and represented by the El Niño Modoki Index (EMI). Higher SST in the central Pacific and lower SST in the east Pacific correlate with reduced precipitation in the East Asian monsoon region [50], Australia [51], parts of South America [28] and South Africa [52] and vice versa. Our result suggests that continental rainfall in certain regions of Africa, Arabia and the Americas are time-delayed El Niño Modoki expressions, albeit with weak amplitude.

Mode 6 (1.1 %) is characterised by a SST pattern concentrated to the tropical and subarctic North Atlantic (Fig. 8b and 9f). The SST pattern, representing the Atlantic Meridional Mode (AMM), is the dominant coupled ocean-atmospheric phenomenon in the tropical Atlantic [54] and its impact on precipitation of the African Sahel zone, Central America and the northern South American continent are well known [55, 56]. Only very recently, [57] found a link between the AMM and Indian summer monsoon rainfall, which they used to improve precipitation forecast. In addition, our result suggests more tele-connections between the AMM and precipitation variability over the Mediterranean, East Africa, the Congo basin, North America and the East Asian monsoon region, providing much potential of advancing local rainfall predictions in those areas. Due to the missing intrinsic time scale with no clear periodicity (Fig. 9f), only correlating and anti-correlating patterns should be interpreted. However, most of the regions identified by this mode satisfy being correlated or anti-correlated.

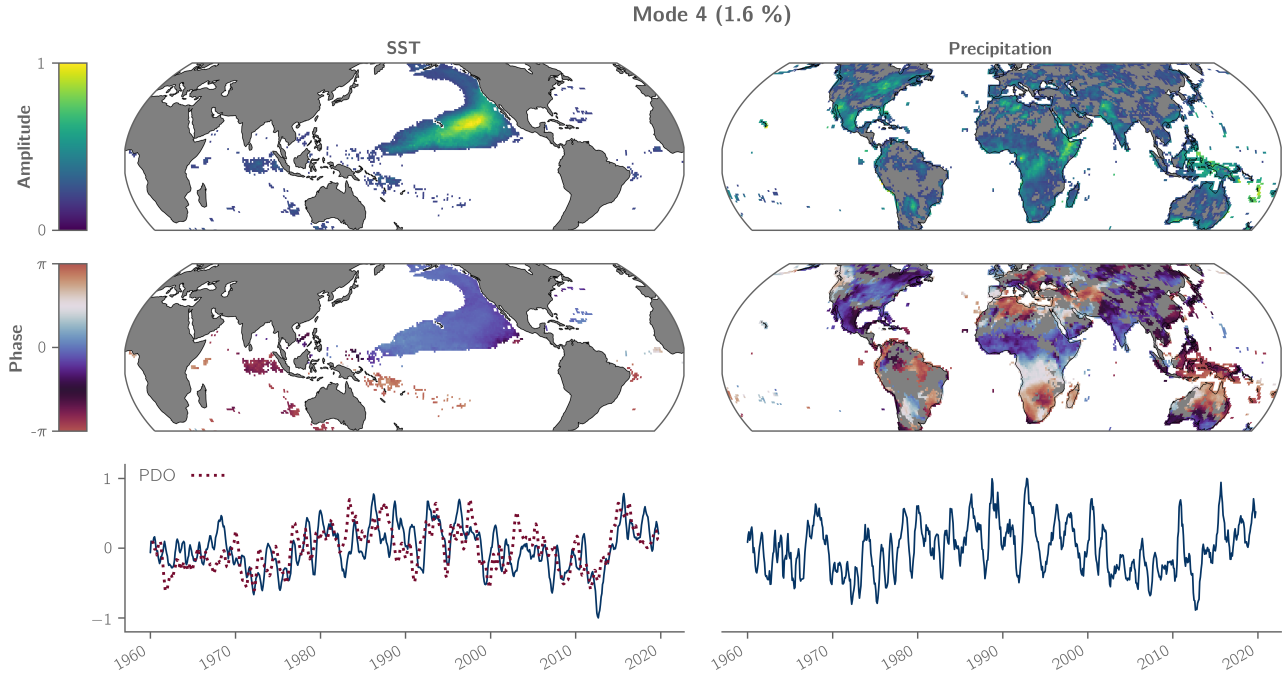
5 CONCLUSION

Understanding the entangled and dynamical climate system is a challenging task, often requiring advanced statistical methods. Finding correlations among a set of different climate variables is complicated by the frequently present lagged responses of different variables to the same forcing. Assuming the processes of interest possess a cyclical nature, we show that complex rotated MCA provides a practical tool to single out such modes from a high-dimensional data space.

By taking into account the spatial amplitude and phase function of the obtained complex modes, we obtain a simple approach to examine otherwise complicated spatial and temporal structures. Our synthetic experiments highlight that, in the case of phase-shifted signals, complex MCA can capture a more comprehensive and complete picture of the correlations present. However, they also show the sensitivity of the Hilbert transform to the boundary conditions of the given time series, e.g. when the time series clearly consists of a non-integer

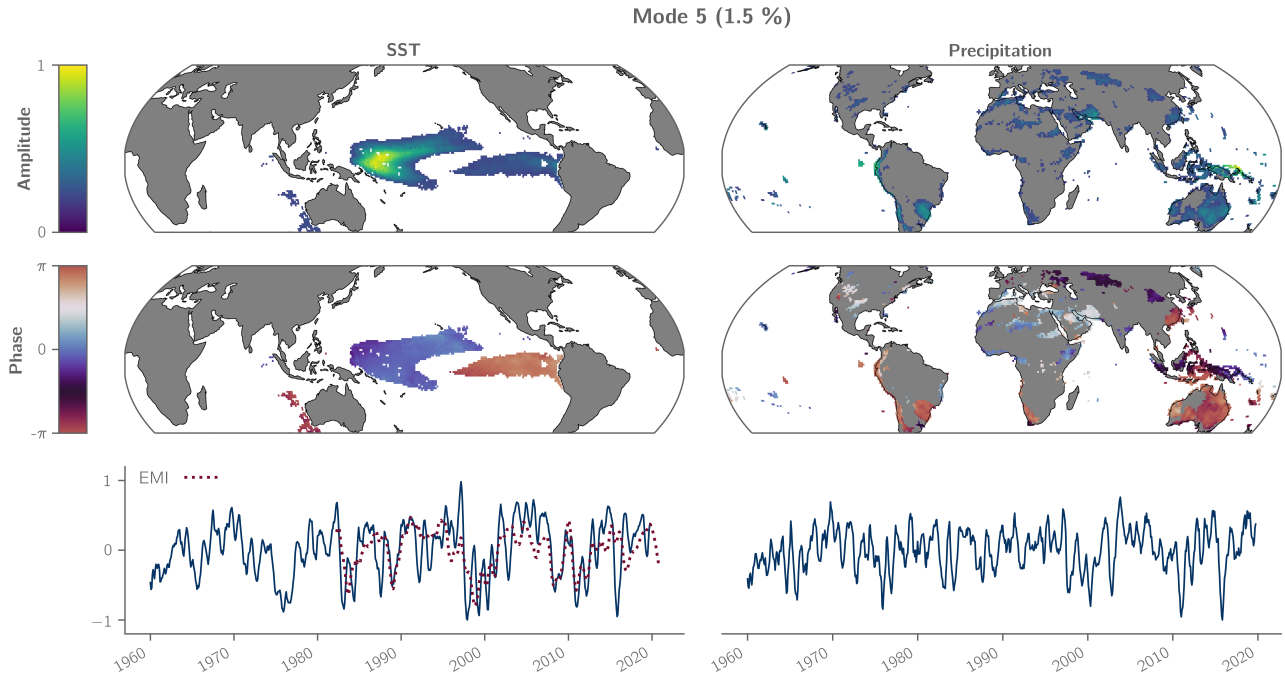


(a) Mode 3 compared to the mean SST field (SST; · · ·) based on latitude weighted SST smoothed with a 6-month moving average as described in Section 3.2.

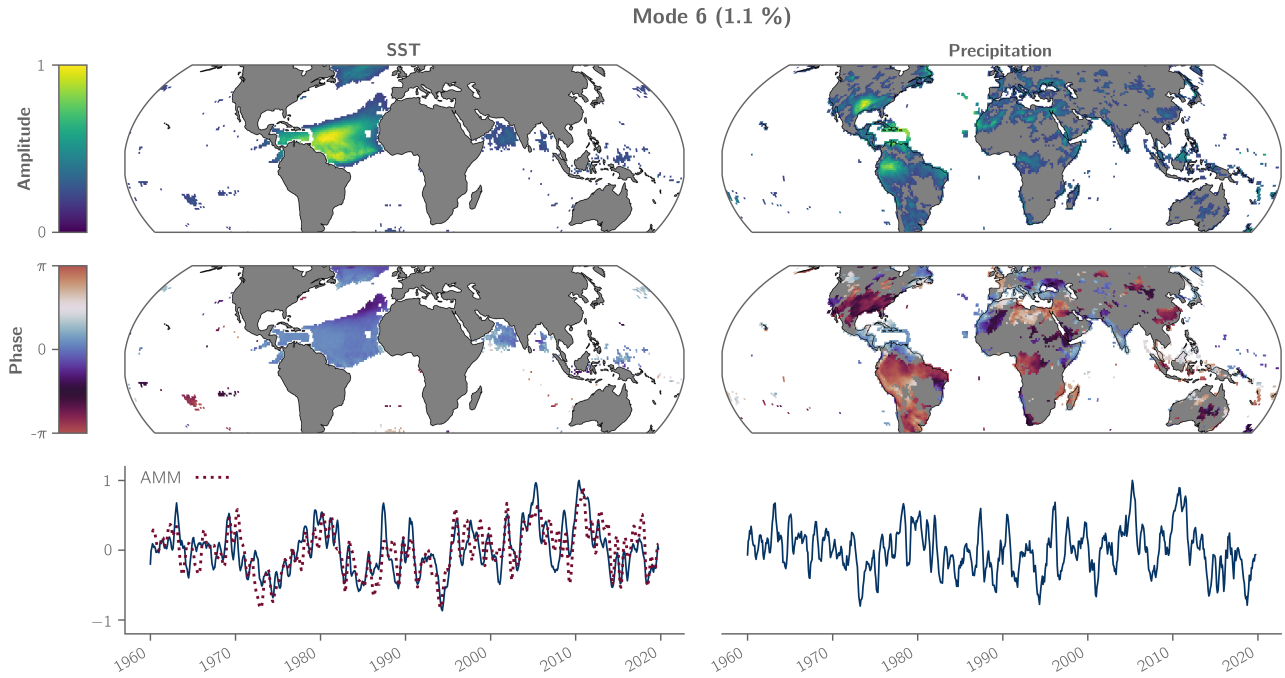


(b) Mode 4 compared to the Pacific Decadal Oscillation (PDO; · · ·) [45] as described in Section 3.2.

Figure 7: As in Fig. 5 but showing **(a)** Mode 3 and **(b)** Mode 4.



(a) Mode 5 compared to the ENSO Modoki Index (EMI; · · ·) [49] as described in Section 3.2.



(b) Mode 6 compared to the Atlantic Meridional Modulation (AMM; · · ·) [53] as described in Section 3.2.

Figure 8: As in Fig. 5 but showing **(a)** Mode 5 and **(b)** Mode 6.

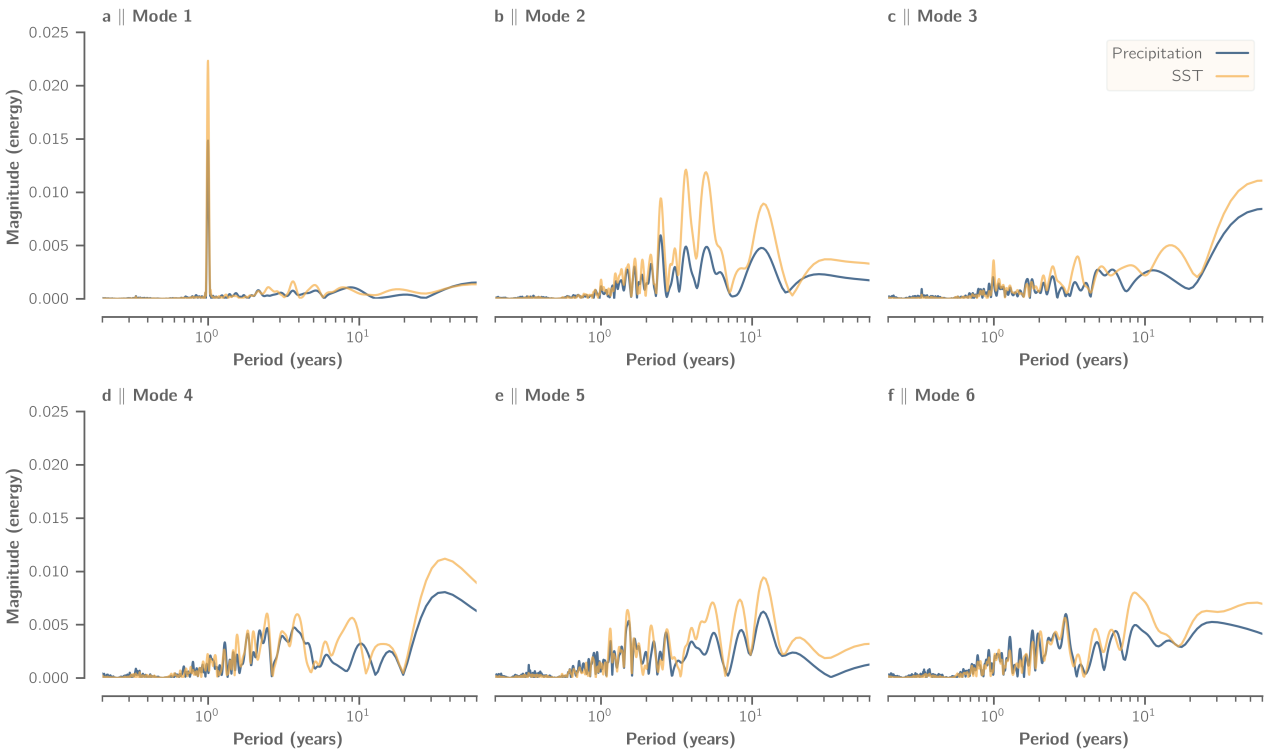


Figure 9: Magnitude spectrum of the first 6 PCs considering only the real part and using a Hanning window.

number of cycles and/or in the presence of trends. Since the stationary assumption does often not hold anymore for many processes because of climate change, time series should generally be pre-processed when applying complex MCA.

Extending the time series via the optimised Theta model mitigates the effect of spectral leakage and produces physically reasonable PCs. This procedure allows us to resolve trends and non-cyclic signals, although for trends the obtained spatial phase function have a simple physical interpretation only for correlating and anti-correlating patterns. Nevertheless, this approach provides a mean of applying complex MCA without the need to detrend the time series of interest. Moreover, excluding the first mode, the original fields can be reconstructed without the seasonal cycle, providing an advanced tool to preprocess time series containing non-stationary seasonal features.

A general caveat in complex MCA is the fact, that the phase function loses its interpretation for PCs with a broadband frequency spectrum. But although the spatial phase function has not always a simple physical interpretation for most of the modes due to their broadband frequency spectrum, complex MCA nevertheless can always be interpreted for the correlating and anti-correlating patterns.

Applying complex rotated MCA to SST and continental precipitation, we clearly identify the main shared dynamics in both variables, namely (i) the seasonal cycle, (ii) the canonical ENSO, (iii) the trends associated to global warming (iv) the PDO, (v) ENSO Modoki and (vi) the AMM. We also retrieve phase shifted signals between the two climate variables. While for the seasonal cycle these phase shifts can directly be translated into a time shift, the remaining modes generally do not lend itself to such a simple interpretation due to their broadband frequency spectra. However, even without a precise equivalent as time lag, the phase function provides a mean to identify regions of lagged correlations, for instance, between the SST of the Pacific and the tropical North Atlantic during El Niño events. Many of the obtained correlation patterns in SST and continental precipitation had already been evidenced by a multiplicity of different, partly regional studies. The great advantage of complex rotated MCA is that it allows to obtain all those patterns by a single analysis of the correlation of two geophysical variables, in a more compact and easy-to-interpret way. Besides, our results

also point out to new ocean-atmospheric teleconnections, that, to our knowledge, have not been reported, most notably for the PDO and the AMM.

Regarding future applications of complex rotated MCA, this method has the potential for shedding light in the investigation of seasonal and sub-seasonal phenomena, as well as for spatially propagating patterns. As future work, we plan to analyse the Madden-Julian oscillation. Besides, complex rotated MCA could be used to evidence other connections between less studied variables, as for instance sea surface salinity, sea surface height, soil moisture, winds, etc, what has the potential of evidence new phenomena and novel aspects of existing or new teleconnections.

ACKNOWLEDGMENTS

This work is part of the Climate Advanced Forecasting of sub-seasonal Extremes (CAFE) project and has been prepared in the framework of the doctorate in Physics of the Autonomous University of Barcelona. The authors gratefully acknowledge funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813844, and also from the Spanish government through the 'Severo Ochoa Centre of Excellence' accreditation (CEX2019-000928-S). This work is a contribution to CSIC Thematic Interdiciplinary Platform TELEDETECT. The study relied on publicly available data only, which has been provided by different institutions and services, namely the ECMWF MARS sever, the Copernicus Climate Change Service, the Physical Sciences Laboratory of NOAA and JAMSTEC.

AUTHOR CONTRIBUTIONS

NR developed the method, coded the Python package, run the analysis and wrote the initial manuscript. All co-authors equally contributed to the work by discussing the methods and results, developing the design of the study and substantially revising the manuscript.

CODE AVAILABILITY

The method is freely available as a compiled Python package, with install instructions and manual provided on GitHub (`https://github.com/nicrie/pycca`).

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.

[2] John E. Kutzbach. Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America. *Journal of Applied Meteorology and Climatology*, 6(5):791–802, 1967.

[3] C Bretherton, C Smith, and JM Wallace. An intercomparison of methods for finding coupled patterns in climate data. *Journal of climate*, 5(6):541–560, 1992.

[4] Harold Hotelling. Relations Between Two Sets of Variates. *Biometrika*, 28(3/4):321–377, 1936. Publisher: [Oxford University Press, Biometrika Trust].

[5] Hrishikesh D. Vinod. Canonical ridge and econometrics of joint production. *Journal of econometrics*, 4(2):147–166, 1976. Publisher: Elsevier.

[6] Raul Cruz-Cano and Mei-Ling Ting Lee. Fast regularized canonical correlation analysis. *Computational Statistics & Data Analysis*, 70:88–100, February 2014.

[7] L Li, R Schmitt, CC Ummenhofer, and KB Karnauskas. North Atlantic salinity as a predictor of Sahel rainfall. *Science Advances*, 2(5):e1501588, May 2016.

[8] JD Horel. Complex Principal Component Analysis: Theory and Examples. *Journal of Climate and Applied Meteorology*, 23(12):1660–1673, December 1984.

[9] Peter Bloomfield and Jerry M. Davis. Orthogonal rotation of complex principal components. *International Journal of Climatology*, 14(7):759–775, 1994. _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.3370140706.

[10] Michael B. Richman. Rotation of principal components. *Journal of climatology*, 6(3):293–335, 1986. Publisher: Wiley Online Library.

[11] Xinhua Cheng and Timothy J. Dunkerton. Orthogonal Rotation of Spatial Patterns Derived from Singular Value Decomposition Analysis. *Journal of Climate*, 8(11):2631–2643, November 1995. Publisher: American Meteorological Society.

[12] John M. Wallace and Robert E. Dickinson. Empirical Orthogonal Representation of Time Series in the Frequency Domain. Part I: Theoretical Considerations. *Journal of Applied Meteorology*, 11(6):887–892, September 1972. Publisher: American Meteorological Society.

[13] Eugene M. Rasmusson, Phillip A. Arkin, Wen-Yuan Chen, and John B. Jalickee. Biennial variations in surface temperature over the United States as revealed by singular decomposition. *Monthly Weather Review*, 109(3):587–598, 1981.

[14] B. Boashash. Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals. *Proceedings of the IEEE*, 80(4):520–538, April 1992. Conference Name: Proceedings of the IEEE.

[15] V. Assimakopoulos and K. Nikolopoulos. The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, 16(4):521–530, October 2000.

[16] Jose A. Fiorucci, Tiago R. Pellegrini, Francisco Louzada, Fotios Petropoulos, and Anne B. Koehler. Models for optimising the theta method and their relationship to state space models. *International Journal of Forecasting*, 32(4):1151–1161, October 2016.

[17] Rob J. Hyndman and Baki Billah. Unmasking the Theta method. *International Journal of Forecasting*, 19(2):287–290, April 2003.

[18] Henry F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200, September 1958.

[19] Alan E. Hendrickson and Paul Owen White. Promax: A Quick Method for Rotation to Oblique Simple Structure. *British Journal of Statistical Psychology*, 17(1):65–70, 1964. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2044-8317.1964.tb00244.x.

[20] L. Thurstone. The simple structure concept. *Multiple Factor Analysis: A Development and Expansion of The Vectors of Mind*, pages 319–346, 1947. Publisher: The University of Chicago Press Chicago.

[21] Holmes Finch. Comparison of the Performance of Varimax and Promax Rotations: Factor Structure Recovery for Dichotomous Items. *Journal of Educational Measurement*, 43(1):39–52, 2006. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-3984.2006.00003.x.

[22] Tao Lian and Dake Chen. An Evaluation of Rotated EOF Analysis and Its Application to Tropical Pacific SST Variability. *Journal of Climate*, 25(15):5361–5373, August 2012. Publisher: American Meteorological Society.

[23] Xinhua Cheng, Gregor Nitsche, and John M. Wallace. Robustness of Low-Frequency Circulation Patterns Derived from EOF and Rotated EOF Analyses. *Journal of Climate*, 8(6):1709–1713, June 1995. Publisher: American Meteorological Society Section: Journal of Climate.

[24] H. Hersbach, B. Bell, P Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and J-N. Thépaut. ERA5 monthly averaged data on single levels from 1979 to present. 2019.

[25] DP Dee, SM Uppala, AJ Simmons, P Berrisford, P Poli, S Kobayashi, U Andrae, MA Balmaseda, G Balsamo, P Bauer, P Bechtold, ACM Beljaars, L van de Berg, J Bidlot, N Bormann, C Delsol, R Dragani, M Fuentes, AJ Geer, L Haimberger, SB Healy, H Hersbach, EV Hólm, L Isaksen, P Kållberg, M Köhler, M Matricardi, AP McNally, BM Monge-Sanz, J-J Morcrette, B-K Park, C Peubey, P de Rosnay, C Tavolato, J-N Thépaut, and F Vitart. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597, 2011.

[26] Adrian Simmons, Hans Hersbach, Joaquin Munoz-Sabater, Julien Nicolas, Freja Vamborg, Paul Berrisford, Patricia de Rosnay, Kate Willett, and Jack Woollen. Low frequency variability and trends in surface air temperature and humidity from ERA5 and other datasets. ECMWF Technical Memoranda 881, ECMWF, February 2021.

[27] G North, T L. Bell, R Cahalan, and F J. Moeng. Sampling Errors in the Estimation of Empirical Orthogonal Functions. *Monthly Weather Review*, 110, 1982.

[28] Renata G. Tedeschi, Iracema FA Cavalcanti, and Alice M. Grimm. Influences of two types of ENSO on South American precipitation. *International Journal of Climatology*, 33(6):1382–1400, 2013. Publisher: Wiley Online Library.

[29] C. F. Ropelewski and M. S. Halpert. North American Precipitation and Temperature Patterns Associated with the El Niño/Southern Oscillation (ENSO). *Monthly Weather Review*, 114(12):2352–2362, December 1986. Publisher: American Meteorological Society Section: Monthly Weather Review.

[30] Na Wen, Zhengyu Liu, and Laurent Li. Direct ENSO impact on East Asian summer precipitation in the developing summer. *Climate Dynamics*, 52(11):6799–6815, June 2019.

[31] Matayo Indeje, Fredrick HM Semazzi, and Laban J. Ogallo. ENSO signals in East African rainfall seasons. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 20(1):19–46, 2000. Publisher: Wiley Online Library.

[32] Aiguo Dai and T. M. L. Wigley. Global patterns of ENSO-induced precipitation. *Geophysical Research Letters*, 27(9):1283–1286, 2000. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/1999GL011140.

[33] Andrea E. Gaughan, Caroline G. Staub, Andrew Hoell, Ariel Weaver, and Peter R. Waylen. Inter- and Intra-annual precipitation variability and associated relationships to ENSO and the IOD in southern Africa. *International Journal of Climatology*, 36(4):1643–1656, 2016. _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.4448.

[34] Annalisa Cherchi and Antonio Navarra. Influence of ENSO and of the Indian Ocean Dipole on the Indian summer monsoon variability. *Climate dynamics*, 41(1):81–103, 2013. Publisher: Springer.

[35] V. Krishnamurthy and Ben P. Kirtman. Variability of the Indian Ocean: Relation to monsoon and ENSO. *Quarterly Journal of the Royal Meteorological Society*, 129(590):1623–1646, 2003. _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1256/qj.01.166.

[36] Stephen A. Klein, Brian J. Soden, and Ngar-Cheung Lau. Remote Sea Surface Temperature Variations during ENSO: Evidence for a Tropical Atmospheric Bridge. *Journal of Climate*, 12(4):917–932, April 1999. Publisher: American Meteorological Society Section: Journal of Climate.

[37] David B. Enfield and Dennis A. Mayer. Tropical Atlantic sea surface temperature variability and its relation to El Niño-Southern Oscillation. *Journal of Geophysical Research: Oceans*, 102(C1):929–945, 1997. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/96JC03296.

[38] R. Saravanan and Ping Chang. Interaction between Tropical Atlantic Variability and El Niño–Southern Oscillation. *Journal of Climate*, 13(13):2177–2194, July 2000. Publisher: American Meteorological Society Section: Journal of Climate.

[39] Michael Alexander and James Scott. The influence of ENSO on air-sea interaction in the Atlantic. *Geophysical Research Letters*, 29(14):46–1–46–4, 2002. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2001GL014347.

[40] John C. H. Chiang and Adam H. Sobel. Tropical Tropospheric Temperature Variations Caused by ENSO and Their Influence on the Remote Tropical Climate. *Journal of Climate*, 15(18):2616–2631, September 2002. Publisher: American Meteorological Society Section: Journal of Climate.

[41] Ngar-Cheung Lau and Mary Jo Nath. The role of the "atmospheric bridge" in linking tropical Pacific ENSO events to extratropical SST anomalies. *Journal of Climate*, 9(9):2036–2057, 1996.

[42] Shang-Ping Xie, Clara Deser, Gabriel A. Vecchi, Jian Ma, Haiyan Teng, and Andrew T. Wittenberg. Global Warming Pattern Formation: Sea Surface Temperature and Rainfall. *Journal of Climate*, 23(4):966–986, February 2010. Publisher: American Meteorological Society Section: Journal of Climate.

[43] Guojun Gu and Robert F. Adler. Spatial Patterns of Global Precipitation Change and Variability during 1901–2010. *Journal of Climate*, 28(11):4431–4453, June 2015. Publisher: American Meteorological Society Section: Journal of Climate.

[44] Filippo Giorgi, Francesca Raffaele, and Erika Coppola. The response of precipitation characteristics to global warming from climate projections. *Earth System Dynamics*, 10(1):73–89, February 2019. Publisher: Copernicus GmbH.

[45] Nathan J. Mantua and Steven R. Hare. The Pacific decadal oscillation. *Journal of oceanography*, 58(1):35–44, 2002. Publisher: Springer.

[46] Nathan J. Mantua, Steven R. Hare, Yuan Zhang, John M. Wallace, and Robert C. Francis. A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production*. *Bulletin of the American Meteorological Society*, 78(6):1069–1080, June 1997. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

[47] Matthew Newman, Michael A. Alexander, Toby R. Ault, Kim M. Cobb, Clara Deser, Emanuele Di Lorenzo, Nathan J. Mantua, Arthur J. Miller, Shoshiro Minobe, Hisashi Nakamura, Niklas Schneider, Daniel J. Vimont, Adam S. Phillips, James D. Scott, and Catherine A. Smith. The Pacific Decadal Oscillation, Revisited.

Journal of Climate, 29(12):4399–4427, June 2016. Publisher: American Meteorological Society Section: Journal of Climate.

[48] David W. Pierce. The Role of Sea Surface Temperatures in Interactions between ENSO and the North Pacific Oscillation. *Journal of Climate*, 15(11):1295–1308, June 2002. Publisher: American Meteorological Society Section: Journal of Climate.

[49] Karumuri Ashok, Swadhin K. Behera, Suryachandra A. Rao, Hengyi Weng, and Toshio Yamagata. El Niño Modoki and its possible teleconnection. *Journal of Geophysical Research: Oceans*, 112(C11), 2007. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2006JC003798.

[50] Juan Feng, Wen Chen, C.-Y. Tam, and Wen Zhou. Different impacts of El Niño and El Niño Modoki on China rainfall in the decaying phases. *International Journal of Climatology*, 31(14):2091–2101, November 2011.

[51] Andréa S. Taschetto and Matthew H. England. El Niño Modoki impacts on Australian rainfall. *Journal of Climate*, 22(11):3167–3174, 2009.

[52] J. V. Ratnam, S. K. Behera, Y. Masumoto, and T. Yamagata. Remote effects of El Niño and Modoki events on the austral summer precipitation of southern Africa. *Journal of Climate*, 27(10):3802–3815, 2014.

[53] John C. H. Chiang and Daniel J. Vimont. Analogous Pacific and Atlantic Meridional Modes of Tropical Atmosphere–Ocean Variability. *Journal of Climate*, 17(21):4143–4158, November 2004. Publisher: American Meteorological Society Section: Journal of Climate.

[54] Jacques Servain, Ilana Wainer, Julian P. McCreary, and Alain Dessier. Relationship between the equatorial and meridional modes of climatic variability in the tropical Atlantic. *Geophysical Research Letters*, 26(4):485–488, 1999. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/1999GL900014.

[55] Peter J. Lamb, Randy A. Peppler, and Stefan Hastenrath. Interannual variability in the tropical Atlantic. *Nature*, 322(6076):238–240, July 1986. Number: 6076 Publisher: Nature Publishing Group.

[56] Elinor R. Martin, Chris Thorncroft, and Ben B. B. Booth. The Multidecadal Atlantic SST—Sahel Rainfall Teleconnection in CMIP5 Simulations. *Journal of Climate*, 27(2):784–806, January 2014. Publisher: American Meteorological Society Section: Journal of Climate.

[57] H. Vittal, Gabriele Villarini, and Wei Zhang. Early prediction of the Indian summer monsoon rainfall by the Atlantic Meridional Mode. *Climate Dynamics*, 54(3):2337–2346, February 2020.

[58] John R. Hurley and Raymond B. Cattell. The Procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behavioral science*, 7(2):258, 1962. Publisher: University of Michigan, Mental Health Research Institute.

Appendix

6 DEFINITION OF arctan2

Every non-zero complex number in Cartesian coordinates, $z = x + iy$, can be transformed into Polar complex coordinates, $z = ae^{i\theta}$, where $a = (x^2 + y^2)^{1/2}$ is the amplitude, and θ the phase of z. For each $z \neq 0$, the phase is only defined up to an integer multiple of 2π, resulting in an infinite number of possible values. In order to construct a well-defined function $\theta(x, y)$, one typically limits the phase θ to $(-\pi, \pi]$. Then, the two-argument arctangent function $\arctan2(y, x)$ converts the slope y/x to the Polar phase via:

$$\theta = \arctan2(y, x) = \begin{cases} \arctan\left(\frac{y}{x}\right) & \text{if } x > 0 \\ \arctan\left(\frac{y}{x}\right) + \pi & \text{if } x < 0 \text{ and } y \geqslant 0 \\ \arctan\left(\frac{y}{x}\right) - \pi & \text{if } x < 0 \text{ and } y < 0 \\ +\frac{\pi}{2} & \text{if } x = 0 \text{ and } y > 0 \\ -\frac{\pi}{2} & \text{if } x = 0 \text{ and } y < 0 \\ \text{undefined} & \text{if } x = 0 \text{ and } y = 0 \end{cases}$$

7 FINDING THE ROTATION MATRIX R

Let us assume a complex loading matrix $L_r \in \mathbb{C}^{n \times r}$ containing only the first r columns (modes) which are to be rotated and n denoting the number of grid points. The number of grid points may be the sum of the grid points of two different fields $n = n_A + n_B$ as it is the case for MCA and described by Equation (7) or simply the total number of grid points if only one field is considered as it is for PCA. In the following, we will drop the subscript r in order to keep the notation simple. For this section, * refers to the conjugate transpose of a matrix and $|\Delta|$ denotes the absolute value of a complex number.

7.1 Orthogonal Varimax rotation

The goal of Varimax rotation is to approximate simple structures [20] of the EOFs which is achieved by simplifying the columns of L via an orthogonal rotation R. For this purpose, [18] defines the simplicity S_k,

$$S_k = \frac{1}{n} \sum_{j=1}^{n} \left(|\widetilde{l}_{jk}|^2\right)^2 - \frac{1}{n^2} \left(\sum_{j=1}^{n} |\widetilde{l}_{jk}|^2 \right)^2, \quad k = 1, \ldots, r \tag{9}$$

which measures the variance of the squared amplitude of the rotated loadings $\widetilde{l}_{jk}$. With increasing variance, the squared amplitudes $|\widetilde{l}_{jk}|^2$ either become low or large, thus increasing simplicity. The normalised Varimax criterion S then reads

$$S = \sum_{k=1}^{r} \left[\frac{1}{n} \sum_{j=1}^{n} \left(\frac{|\widetilde{l}_{jk}|^2}{h_j^2} \right)^2 - \frac{1}{n^2} \left(\sum_{j=1}^{n} \frac{|\widetilde{l}_{jk}|^2}{h_j^2} \right)^2 \right], \tag{10}$$

where $h_j = \left(\sum_{k=1}^{r} |l_{jk}|^2\right)^{1/2}$ represents the *communality* of grid point j, which is the amount of variance of the jth variable accounted for by the r retained modes. Subsequently, the normalised Varimax-rotated EOFs $\widetilde{L}_n$ are the solution to

$$\widetilde{L}_n = H^{-1} L R \quad \text{s.t.} \quad \underset{l}{\operatorname{argmax}}(S), \tag{11}$$

with the communality matrix $H \in \mathbb{R}^{n\times n}$ whose elements are given by $\mathrm{diag}(h_1, \ldots, h_n)$. Equation (11) can be solved by an iterative process, in which the EOFs are rotated in pairs in order to maximise S. Finally, the de-normalised Varimax-rotated EOFs can be computed via $\widetilde{L} = H\widetilde{L}_n$.

7.2 Oblique Promax rotation

Achieving simple structures with Promax is done via an oblique Procrustes transformation [58]. Any target matrix of rotated EOFs T, can always be approximated from a base matrix B via a linear transformation R,

$$T = BR + E, \tag{12}$$

where E is an error matrix. Minimising $\mathrm{trace}(E^*E)$ yields the complex Procrustes equation,

$$R = (B^*B)^{-1} B^*T. \tag{13}$$

The basic assumption of Promax is that an Varimax orthogonal rotation is a reasonable approximation to an optimal oblique solution. Therefore, the base matrix is chosen to be $B = \widetilde{H}^{-1}\widetilde{L}$ whose entries are normalised by the Varimax communalities $\widetilde{h}_j = \sum_{k=1}^{r} |\widetilde{l}_{jk}|^2$. Then, the Promax equation defines the elements of the target matrix T,

$$t_{jk} = |b^+_{jk}|^{p+1}/b^+_{jk}, \tag{14}$$

where $^+$ denotes the max-normalised entries given by $b^+_{jk} = b_{jk}/\max_j |b_{jk}|$. The power parameter p thus defines the strength of the Promax operation, while the sign remains unchanged. Using Equation (13), the normalised Promax-rotated EOFs are given by

$$\begin{aligned}\widetilde{L}_P =& \widetilde{H}^{-1}\widetilde{L}RD \\ =& \widetilde{H}^{-1}\widetilde{L}\left[\left(\widetilde{L}^*\widetilde{H}^{-2}\widetilde{L}\right)^{-1}\widetilde{L}^*\widetilde{H}^{-1}T\right]D,\end{aligned} \tag{15}$$

where the normalisation matrix is given by $D^2 = \mathrm{diag}\left(R^TR\right)^{-1}$.