Conference Proceedings Paper

New insights on land surface-atmosphere feedbacks over tropical South America at interannual timescales

Juan Mauricio Bedoya-Soto [1],* and Germán Poveda [1]

[1] Universidad Nacional de Colombia, Sede Medellín, Departamento de Geociencias y Ambiente, Facultad de Minas, Medellín, Colombia.

* Correspondence: jmbedoya@unal.edu.co, Tel.: +01-057-4-5988669

Abstract: Using monthly data for the period 1979-2010, we study the dynamics and strength of land surface-atmosphere feedbacks (LAFs) among variables involved in the heat and moisture fluxes, at interannual timescales in Tropical South America (TropSA). The variables include precipitation, surface air temperature, specific humidity at 925 hPa, evaporation, and estimates of volumetric soil water content. We use a Maximum Covariance Analysis (*MCA*) to reduce the dimensionality and to rank the relative contributions to LAFs and group the time series into Maximum Covariance States (*MCS*) with common mechanisms among variables. We estimate linear (Pearson correlations) and non-linear association (information transfer or causality) metrics among pairs of variables to configure the structure of linkages. The main *MCS* associated with LAFs over TropSA are strongly influenced by ENSO, and the meridional and equatorial SSTs modes over the Atlantic and Indian Oceans. ENSO favors a unimodal behavior, with center of action in the Amazon River basin, while the SSTs mode over the Tropical North Atlantic (TNA) results in a dipole between northern and southern TropSA. Our results show that soil moisture plays a leading role in regulating heat and water anomalies, and provides the memory of the LAFs and their subsequent influence. Thus, volumetric soil water becomes a fundamental variable leading up to 9 month-lags. ENSO enhances the interannual connectivity and memory inside LAF mechanisms with respect to other modes. Within the identified multivariate structure, evaporation and soil moisture enhance the interannual connectivity of the whole set of variables since both variables exhibit more frequent two-way feedbacks with the remaining variables.

Keywords: Soil Moisture, Feedbacks, Tropical South America, Interannual Variability

1. Introduction

The humid tropics concentrate a large amount of net radiation and water vapor, and consequently, intense heat and humidity fluxes dominate the interactions between the soil and the lower atmosphere [1]. The excess of net radiation is balanced through latent heat and sensible heat fluxes [2–4]. By controlling the partition of these fluxes, soil moisture modulates diverse land-atmosphere feedbacks (LAFs); similarly, the rates of change between dry and wet conditions in the soil necessarily impact surface temperatures [5,6].

Traditionally, the study of LAFs have been approached from physical modelling and numerical experiments [7-13]; analysis of observations and models with statistical tools [3, 14-17]; traces of moisture trajectories [18,19], among others. An important body of literature has focused on the role

of vegetation and land uses in the dynamics of LAFs [20-24], and the conditions under which LAFs determine the stability of the lower atmosphere [25-27].

Tropical South America (TropSA) and, in particular the Amazon River basin, is important for study of LAFs on a broad range of spatiotemporal scales, given the importance of recycled precipitation over the region [28-35]. [31] describe a recycling pattern in South America that regionally connects evapotranspiration in the Amazon with precipitation in La Plata River basin [36-39]. These authors also discuss the recycling east of the Andes in TropSA owing to the blocking orographic effect on moisture transport at low levels [40,41]. In addition, several studies have identified the ways in which LAFs are associated with the South American Monsoon (SAM) [36,42-43], and the important role of LAFs over TropSA in connecting the Tropical Pacific with the Tropical North Atlantic SSTs anomalies at interannual timescales [44-46].

The study of [47] provide a synthesis of the most recurrent links involved in LAFs (Figure 1), and differentiate between *state* variables and *process* variables. In this context, the state variables are temperature and humidity content of both the lower atmosphere and soil, while process variables are evaporation, precipitation, latent and sensible heat fluxes, albedo, and cloudiness in conjunction with net radiation. Within this framework, evapotranspiration is the most heterogeneous process involved in LAFs since it connects soil moisture with the other state variables (Figure 1). Overall, soil moisture regulates the entire LAF system by its direct linkage to evapotranspiration and precipitation [30, 48,49]. Soil moisture also influences soil temperature through albedo and latent heat flux [50], and at the same time non-linear vegetation-induced turbulence, determined by vegetation stature and cover fraction, significantly influences ET-soil moisture relationships [51]. Another important feedback represented by [47] is the influence of soil temperature, in connection with air temperature, sensible heat flux and evaporation.

The scheme of relationships between state and process variables shown in Figure 1 is a hierarchical adaptation using graph theory of the original figure presented by [47]. Graph theory is a specific branch of mathematics that studies the interconnections between objects. [42] propose an interpretation of couplings, feedbacks, and interactions via graph theory to further understand LAFs. According to this interpretation, the concept of feedback with graphs coincides with the type of bidirectional link established in Figure 1 between evaporation and soil moisture; on the contrary, the precipitation-moisture link in the soil is defined as a coupling, or a hierarchical interaction whereby one variable dominates over the other. The structure of links shown in Figure 1, derived from graph theory, provides a clearer understanding of the processes and mechanisms involved in LAFs [52-54].

Mathematically, the concept of feedback also represents cycles of interactions between two or more variables (Figure 1). In graph theory and other sciences, cycles are closely connected with Markov Chains [54,55]. One of the main applications of Markov Chains is Information Theory [56]. To take advantage of what is known in advance about a transmitted message, information theory suggests that all source data can be considered as a Markovian process. Precisely, [56] defined the concept of Entropy referring to the amount of noise contained in a message. Entropy in Thermodynamics and Information Theory are analogous terms but different concepts, although there are efforts to link them [57,58].

The aim of this paper is to explore new ideas and tools to advance our understanding of land surface-atmospheric feedbacks in Tropical South America (TropSA) at interannual timescales, in particularly by studying the connectivity between state and process variables involved in dynamical LAFs over TropSA. To achieve this goal, we propose to integrate classical analyses of climatic fields with more recent methods of information theory and graph theory. Our approach will include linear and non-linear analysis, including metrics such as causality between variables (non-linear couplings). An explicit goal of this paper is to study the essential role of such feedbacks to understand the mechanisms involved in water and heat anomalies in TropSA at interannual timescales. In addition, we aim to assess how LAFs in TropSA are associated with ENSO, and SSTs anomalies over the TNA, and other macro-climatic phenomena [44, 59].
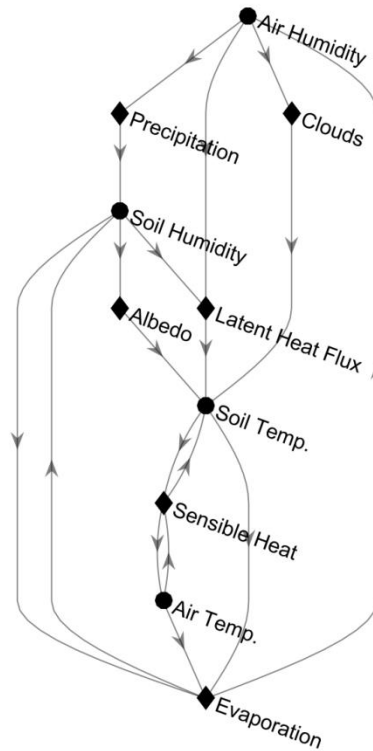
Figure 1. Conceptual scheme of interconnections among the main variables involved in the studied land surface-atmosphere feedbacks (LAF), adapted as a graph from [7], and hierarchized by number of connections of each variable. The state (process) variables are denoted as circles (diamonds) nodes.

The study is organized as follows. We present data and methods in Section 2. In section 3, we first assess data in order to standardized and remove noise (subsection 3.1). Then, we define spatiotemporal patterns of maximum covariance between pairwise selected variables over TropSA during the period 1979-2010, and, in order to estimate the interannual linear and non-linear connectivity of the set of variables, we use correlation and causalities among the representative time series of the estimated maximum covariance patterns (subsections 3.2 and 3.3). Using the same time series, we categorize the identified patterns according to its ENSO influence (subsections 3.4 and 3.5). Finally, with these categorized patterns, we evaluate the structure of relations between the proposed state and process variables through the linear and non-linear couplings using elements of graph theory (section 3.6 and 3.7).

2. Experiments

2.1. Data

We focus our study of LAFs at interannual timescales on tropical South America (TropSA) (15°N-20°S and 82°W-40°W, Figure 2) during the period from 1979 to 2010, through a detailed analysis of the interactions existing among five variables involved in LAFs dynamics, namely surface air temperature (T_{2m}), specific humidity at 925 hPa (SH^{925}), and soil moisture (VSW) (state variables), precipitation (PRC) and evaporation (EVP) (process variables) [7].

To that end, we use monthly climatic fields at 1° x 1° spatial resolution, as follows: Precipitation from the Global Precipitation (GPCC) Full Data Product version 7 [60]. The GPCC monthly precipitation fields are suitable to study hydrological and atmospheric linkages at regional scales on land. From the ERA-Interim Reanalysis [61], monthly fields of T_{2m}, SH^{925} and EVP, which is represented by the Instantaneous Moisture Flux. Volumetric Soil Water content (VSW) is obtained

from the ERA-Interim/Land Reanalysis V.2 [62, 63]. This latter product is focused on those land processes not well represented in the ERA-Interim Reanalysis [64].

We also use the Niño 3.4 index to represent the dynamics of ENSO over the tropical Pacific Ocean at interannual timescales, and monthly fields of SST from the Hadley Center (HADISST) to map the statistically significant correlations between hydro-climatic fields over TropSA.
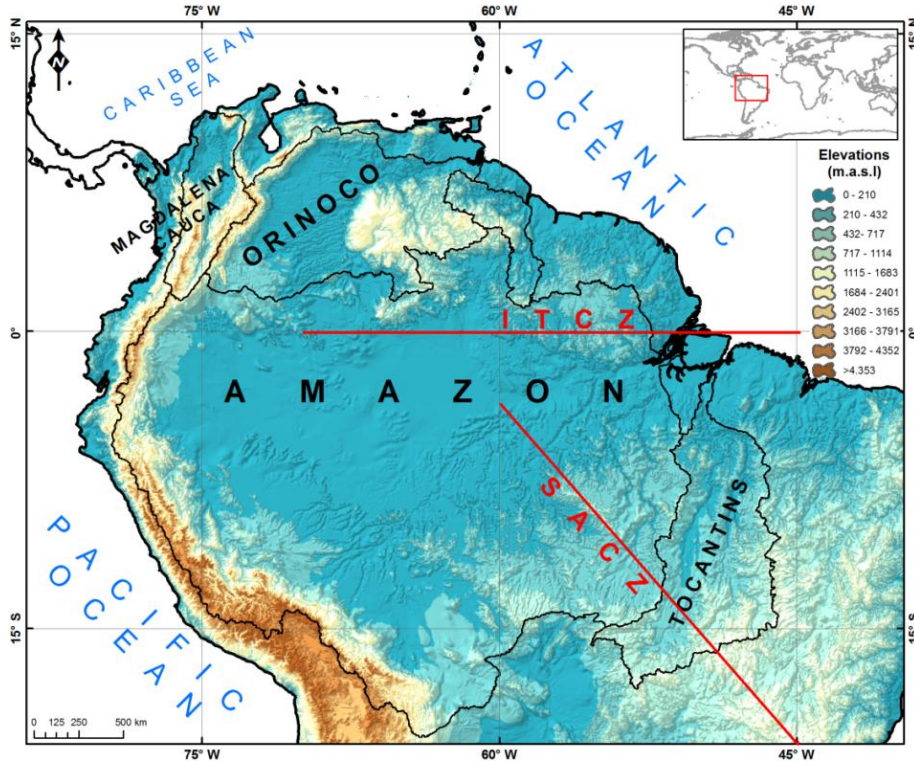

Figure 2. Study region with major river basins included in the analysis: Amazon, Tocantins, Orinoco, and Magdalena.

2.2. Standardization and Noise Removal Procedures

Monthly time series of all variables are available between 1979 and 2010 for grid cells within the TropSA domain. To study the interannual variability we removed the annual cycle by a simple standardization procedure (subtracting the monthly mean and scaling by the monthly standard deviation) of all data sets. Data from the ERA-Interim reanalysis require a split climatology associated with systematic discontinuities in the water balance due to changes in the sensor channel of the Microwave Special Sensor (SSM/I) [65]. Thus, we estimated separated means and standard deviations for the periods 1979-1991, 1992-2002, and 2003-2010 to estimate monthly anomalies.

With the aim of removing noise from data, we build a matrix $X \in \Re^{mxn}$ $m > n$, with monthly anomalies for each variable, with m columns containing the 1° x 1° cells within the TropSA domain, and n rows corresponding to the months of the 1979- 2010 period (n=384). To extract the spectral components required to remove noise from each data set, we apply a Singular Value Decomposition (SVD) to each data matrix, X.

The SVD procedure establishes that any rectangular matrix $X \in \Re^{mxn}$ can be factorized in matrices, $U \in \Re^{mxm}$, $\Sigma \in \Re^{mxn}$ and $V \in \Re^{nxn}$, such that:

$$X_{mxn} = U_{[mxm]} \Sigma_{[mxn]} V^T_{[nxn]} = \sum_{i=1}^{\min(m,n)} u_i \sigma_i v_i^T = u_1 \sigma_1 v_1^T + u_2 \sigma_2 v_2^T + \ldots + u_m \sigma_m v_n^T \qquad (1)$$

The matrix U is dense, and the Left Singular Vectors (u_i) are linear combination of the columns of the matrix X; while the matrix Σ is sparse, the singular values (σ_i) are scalars such that $\sigma_1 \geq \sigma_2 \geq \dots \geq 0$ located on the diagonal; and the matrix V is dense, with the Right Singular Vectors (v_i) being linear combination of the rows of the matrix X.

The singular values, σ_i, quantify the dimensions of the variance of the matrix X. The u_i and v_i are unitary orthonormal vectors that conform the row and column spaces of X. In this sense the rank (r) of X, with $r < \min(m,n)$, is the maximum number of linearly independent columns/rows, and corresponds to the amount of singular values, $\sigma_i > 0$. At the same time, u_i and v_i are eigen-vectors of the matrices XX^T and X^TX, and the singular values are the square root of the eigen-values.

The Eckart-Young-Mirsky theorem describes the dimensional reduction of each matrix X of rank r as an approximation through a matrix $\hat{X} \in \Re^{mxn}$ of rank k, such that $k<r$ [66]. In summary, upon disregarding the smaller singular values of the SVD of X, we derive the truncated (lower-rank) matrix $\hat{X}$. The removed σ_i are the least representative dimensions in the row space ($U_{i=1:r}$) and column space ($V_{i=1:r}$) of each matrix X. If we use the first k singular factors in eqn. (1), the form of the truncated matrix $\hat{X}$ is:

$$X \approx \hat{X}_{mxn} = U_{[mxk]}\Sigma_{[kxk]}V_{[nxk]}^T = \sum_{i=1}^{k} u_i \sigma_i v_i^T = u_1\sigma_1 v_1^T + u_2\sigma_2 v_2^T + \dots + u_k\sigma_k v_k^T \text{ with } k < r \qquad (2)$$

With this low-rank approximation, each matrix X can be explained as the linear combination of k underlying patterns; which are the first k singular values and vectors of X. The selection of k is a critical problem since $\hat{X}$ must maintain enough information of the matrix X, while avoiding the largest amount of noise . In order to determine k, we use the singular values, σ_i, to estimate the Frobenius matrix norm, as

$$\|X\|_F^2 = \sum_{i=1}^{min(m,n)} \sigma_i^2 \qquad (3)$$

The dimensional reduction of matrix X tries to minimize its difference with its truncated version ($\hat{X}$). Such difference can be represented as the sum of the $\sigma_i < \sigma_{k+1}$, such that

$$\|X - \hat{X}_k\|_F^2 = \sum_{i=k+1}^{min(m,n)} \sigma_i^2 \qquad (4)$$

And the relative contribution associated with each singular value, σ_i, to the total variance of the matrix X is estimated as,

$$f_k = \sigma_k^2 \Big/ \sum_{i=1}^{min(m,n)} \sigma_i^2 \qquad (5)$$

The sum of the k relative fractions is denoted by F_k. The most common criteria to determine the reduction is established in such a way that the k factors of the truncated matrix contain between 70 - 90% of the total variance, or $X \approx \hat{X}_k$ when $F_k \approx 0.7 - 0.9$.

Another criterion to select k is based upon the mathematical concept of entropy used within the theory of information framework [56]. Here we associate the random degree of a matrix X to the entropy, H_x, as a measure of uncertainty that varies between 0 and 1. An entropy value close to 1 indicates that X is highly random, whereas close to 0 indicates that X takes a limited number of configurations. Thus, the most important k singular values to reduce the matrix X are related with

the fraction of the variance associated with each one (f_i, eqn. 6). The relative entropy of X is defined as the relation between the maximum rank that it can take according to the minimum between m and n, and therefore

$$H_X = -\frac{1}{\log_2(min\{n,m\})} \sum_{i=1}^{min(n,m)} f_i \log_2 f_i \qquad (6)$$

To find the number of singular values k that best represent the variance structure of X using the entropy criteria, it is necessary to search for the minimum number of singular values, k_H, such that:

$$k_H = arg\ min(k) \sum_{i=1}^{k} f_i \geq H_X \qquad (7)$$

For our purposes, we removed the noise from each variable by extracting the k_H singular values using eqn. (7). With such number of singular factors, we used eqn. (2) to construct the truncated matrix, $\hat{X}$.

2.3 Maximum Covariance Associated with Interannual Variability in TropSA

After removing the noise from the three state variables (SH[925], T2m, VSW) and the two process variables (EVP and PRC), we analyze the spatiotemporal covariance between all possible pairs of variables (10 combinations). Owing to the standardization procedure, the covariance represents the interaction between each pair of variables x and y at interannual timescales.

Mathematically, we use the matrices $X \in \Re^{m \times n}$ (m, 1° x 1° cells) and $Y \in \Re^{q \times n}$ (q, 1° x 1° cells), during the n=384 time periods in common to derive their truncated versions $\hat{X} \in \Re^{m \times n}$ and $\hat{Y} \in \Re^{q \times n}$. Then, we estimate the covariance matrix, C_{XY}, as:

$$C_{XY} = \frac{1}{n-1} \hat{X} \hat{Y}^T \qquad (8)$$

We factorize the C_{XY} matrix employing SVD (eqn. 1) to detect those patterns explaining the maximum fraction of the covariance between $\hat{X}$ and $\hat{Y}$ at the interannual timescale. This method is also known as Maximum Covariance Analysis (MCA) [53, 67, 68]. Through MCA, the C_{XY} matrix can be decomposed as:

$$C_{XY[m \times q]} = U_{[m \times m]} \Sigma_{[m \times q]} V_{[q \times q]}^T = \sum_{i=1}^{min(m,q)} u_i \sigma_i v_i^T = u_1 \sigma_1 v_1^T + u_2 \sigma_2 v_2^T + \ldots + u_m \sigma_m v_q^T \qquad (9)$$

In this case, the first singular factor, $u_1 \sigma_1 v_1^T$, is the first Maximum Covariance State (MCS_1) between $\hat{X}$ and $\hat{Y}$. The magnitude of the first singular value, σ_1, indicates the amount of covariance between both variables, which is captured by the MCS_1. The second singular factor, $u_2 \sigma_2 v_2^T$, is orthogonal to the first one, and maximizes the remaining portion of the covariance, and so on until explaining all the covariance contained in C_{XY}. A description of the way both variables maximize their temporal covariance within each MCS_k, is provided through the estimation of the series x_k and y_k, by projecting the matrices u_k and v_k (containing the singular vectors of C_{XY}) on the original matrices $\hat{X}$ and $\hat{Y}$ (eqn. 10), as:

$$x_k = u_k^T \hat{X} \quad , \quad y_k = v_k^T \hat{Y} . \qquad (10)$$

Each MCS_k contains a fraction of the square covariance, f_k^{XY}, defined as:

$$f_k^{XY} = \sigma_k^2 \Big/ \sum_i \sigma_i^2 \, ,$$

(11)

Such fraction of the relative covariance, f_k^{XY}, represents the magnitude of the covariance that is explained by each MCS_k between each pair of variables. The sum of the k relative fractions is denoted by F_k^{XY}.

Finally, we estimate the spatial patterns associated with $\hat{X}$ and $\hat{Y}$ for each MCS_k. To that end, we use correlation maps (ρ) between the series x_k and y_k and the column vectors of the truncated matrices $\hat{X}$ and $\hat{Y}$ (Table 1).

Table 1. Types of maps employed to represent the spatial patterns resulting from Maximum Covariance Analyses.

Symbol	Description of map
$\rho\left(x_k, \hat{Y}\right)$	Correlation between vector x_k with each column vector of matrix $\hat{Y}$
$\rho\left(y_k, \hat{X}\right)$	Correlation between vector y_k with each column vector of matrix $\hat{X}$
$\rho\left(x_k, \hat{X}\right)$	Correlation between vector x_k with each column vector of matrix $\hat{X}$
$\rho\left(y_k, \hat{Y}\right)$	Correlation between vector y_k with each column vector of matrix $\hat{Y}$

2.4 A Graph Model of LAF in Tropical South America at Interannual Timescales

[53] review the combined usage of MCA and Graph Theory. [52] and [55] propose using graph models to evaluate climatic interactions in a context of causality. For our purposes, we first classify the interaction of the expansion coefficients, x_k, and y_k (eqn. 10), according to the interannual mode that determines their maximum covariance. The criterion to group the singular vectors of each matrix is defined by the degree of correlation with the Niño 3.4 index.

Thus, we identify three interannual modes that dominate the maximum covariance between the interactions of the variables involving the time series which represent each MCA. Then, we construct a complete graph to study the connectivity at interannual time scale for each Maximum Covariance State (*MCS*) between variables, for lags from 1 to 9 months. We used the concept of links between two variables x_k and y_k [48] to construct the LAF graph model, Γk, in TropSA, as:

$$\Gamma_k = f(V_k, E_k),$$

(12)

where the nodes V_k are the variables with maximum covariance under the k-th interannual mode. The weight of the edges, E_k, are the results of the correlation and causality metrics estimated between the x_k, and y_k from each MCS (Apendix A).

When a graph contains all the possible interactions among nodes is called a Complete Graph. In this type of graphs, the number of edges $\left|E_k\right|$ connecting nodes $\left|V_k\right|$ is estimated as:

$$\left|E_k\right| = \frac{\left|V_k\right|\left(\left|V_k\right| + 1\right)}{2}$$

(13)

In our case, we use five variables ($|V_k| = 5$) to construct the complete graph, and thus there are $|E_k| = 10$ possible links between all variables for each graph.

The graph depicting the linear associations between the *MCS* is constructed by assigning the weight in the edges as correlations (ρ). A graph of non-linear associations among the same *MCS* is constructed with causalities (τ) (Appendix A).

Aditionally, we introduce a metric to quantify the ratio of bi-directional causalities between two series *y* and *x* (eqn. 14), as:

$$\Phi_{y \leftrightarrow x} = \frac{\tau_{y \to x}}{\tau_{x \to y}} \tag{14}$$

A value of $\Phi_{x \leftrightarrow y} = 1$ indicates a bidirectional feedback, while $\Phi_{x \leftrightarrow y} \neq 1$ quantifies the degree of coupling between both variables. For our purposes, we established a threshold $\tau_{y \to x} > |1\%|$ to define the relative causalities that indicate the interaction strength.

Estimation of simultaneous correlation does not provide information about directionality, and thus the resulting linking graph is called Undirected Graph. Otherwise, by applying bidirectional metrics, such as Causality described in Appendix A [69], it is possible to establish a Directed Graph, denoting directions with arrows (Figure 3).

However, lagged correlations and causalities also quantify bidirectional characteristics. Figure 3 shows a basic relationship between two time series x_k, and y_k as a schematic view in two time steps: *X(t-1)* influences *Y(t)* and *Y(t-1)* influences *X(t)*. The red arrow in Figure 3 connects series when x_k leads y_k, and the blue arrow when y_k leads x_k. Hence, in both cases, we systematically estimate correlations, $\rho(x_k, y_k)$ and $\rho(x_k, y_k)$, and causalities, $\tau_{x \to y}$ and $\tau_{y \to x}$ (Appendix A and B).
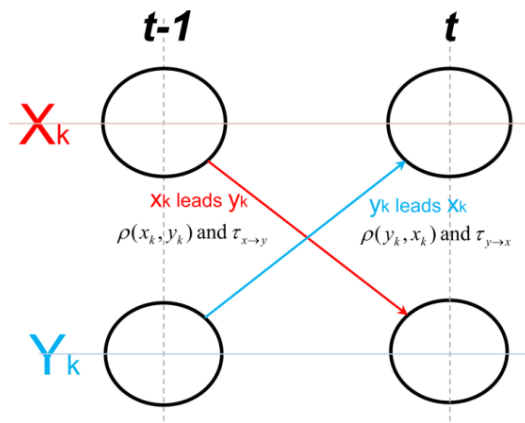
Figure 3. Basic scheme of the interaction between two singular vectors resulting from a Maximum Covariance Analysis (MCA) between two variables X and Y for each mode *k* assuming 1 lag-month.

3. Results

3.1 SVD, Noise Reduction, and Entropy

We constructed the matrix of standardized anomalies, $X_{[mxn]}$, for each data set for the period spanning from 1979 to 2010. Although all fields have the same cell size (1 ° x 1 °), they have different particular grid arrangements, thus implying small differences in the number of cells covering the TropSA region among the three data sources (m, Table 2). In all cases, *n* corresponds to the 384

months of the study period. Moreover, the rank defines the number of the linearly independent columns and rows of X. From this perspective, if we assume that few modes control most of the hydroclimatic variance at interannual time scale in TropSA [44], a high rank of X indicates the presence of high amounts of noise in the data sets (Table 2).

To remove the noise, we factorize each matrix $X_{[mxn]}$ using SVD (eqn. 1) and then we estimate the truncated matrix $\hat{X}_{[mxn]}$ having a lower rank k [66]. As a result, each matrix $\hat{X}_{[mxn]}$ contains the variance of X explained by the most relevant k singular values, σ_k.

Table 2. Characteristics of the matrices containing standardized spatiotemporal anomalies of the studied variables for the TropSA region during the period from 1979 to 2010. Values of f_k for k =1, 2 denote the relative variance estimated through SVD. For each matrix X also are shown the estimated values of entropy (Hx, eqn. 6), the number factors selected to construct the denoised matrix $\hat{X}$ (k_H, eqn. 7), and the cumulated variance contained in $\hat{X}$ (F_{k_H}).

Data Source	X_{mxn}	m (# of cells 1°)	n (# of months)	r	F_1	F_2	Hx	k_H	F_{k_H}
GPCC	Precipitation (PRC)	1101		372	0.12	0.20	0.76	64	0.76
ERA-Interim	Evaporation (EVP)	1047		372	0.12	0.22	0.71	31	0.71
ERA-Interim Land	Volumetric Soil Water (VSW)	1039	384	348	0.16	0.31	0.61	9	0.63
ERA-Interim	Air Temperature (T2m)	1047		348	0.30	0.43	0.53	4	0.55
ERA-Interim	Specific Humidity at 925 hPa (SH[925])	1047		348	0.31	0.45	0.50	3	0.53

In order to select the value of k, we obtain the complete spectrum of singular values, σ_X, of each matrix X (Figure 4A). The form of this spectrum describes the cumulative variance of each data set over TropSA. The slope of such spectra is high for SH[925] and T2m because both state variables of the lower atmosphere capture most of the variance in their first $\hat{X}_{[mxn]}$ (Table 2). On the other hand, process variables (PRC and EVP) exhibit a more distributed variance across their set of σ_i values, and thus the smaller slopes of their spectra (Figure 4A). While the first singular values of SH[925] and T2m take up around 30% of the total square variance, it amounts to 12% for PRC and EVP (f_1, Table 2). Finally, the spectrum of VSW lies between both of them (Figure 4A).

The form of the spectrum of σ_X is also related to the relative entropy, Hx, estimated through eqn. (6) (Figure 4B). A matrix X with low Hx concentrates most of its variance in few singular values (σ_X). This interpretation implies, in our case, that a matrix X with low relative entropy (Hx) represents a variable controlled by few modes over TropSA at interannual timescales. In terms of Hx, and according to the hierarchy of the singular values spectra shown in Figure 4A, the state atmospheric variables SH[925] and T2m are the less complex of our data sets (Hx= 0.5 , Figure 4B).

On the other hand, the process variables (PRC and EVP) are the most complex in the region H_{PREC}= 0.76 and H_{EVP}= 0.71. This result confirms the diversity of factors that make part of the hydro-climatic variance at interannual timescale, thus reflecting the implicit complexity of the linkages between these two processes over TropSA [44, 46, 59,70-72]. The high entropy of EVP can be also explained in terms of the multivariate influence of water and energy fluxes in TropSA [22, 47, 73]. Finally, entropy of soil moisture exhibits an intermediate value (H_{VSW}= 0.61), although it exhibits the highest value among the state variables. Such relative complexity suggests that VSW is essential for the LAFs over TropSA, given its relation with EVP and PRC [3, 48].

The direct relationship between entropy, Hx, of matrix X and k implies having truncated matrices $\hat{X}_{[mxn]}$ with few singular values and low entropy. With this criterion, we select the most

relevant k singular factors (σ_k) to estimate the truncated matrices $\hat{X}_{[mxn]}$ (eqn. 2). Among the methods of selection, the least amount of k singular factors is obtained using Entropy (k_H, eqn. 7). After the dimensional reduction, each matrix acquires a rank k_H and stores an amount of interannual variance given by f_{k_H} (Table 2). Therefore, to estimate the truncated matrices, $\hat{X}_{[mxn]}$, of PRC and EVP we require, respectively, $k_H = 64$ and $k_H = 31$ singular values (Table 2). On the other hand, SH[925] ($k_H = 3$) and T2m ($k_H = 4$) demand less singular values among the datasets. Specifically, SH[925] contains the state variable with less entropy, and therefore it can be represented with the fewest singular values (Table 2).
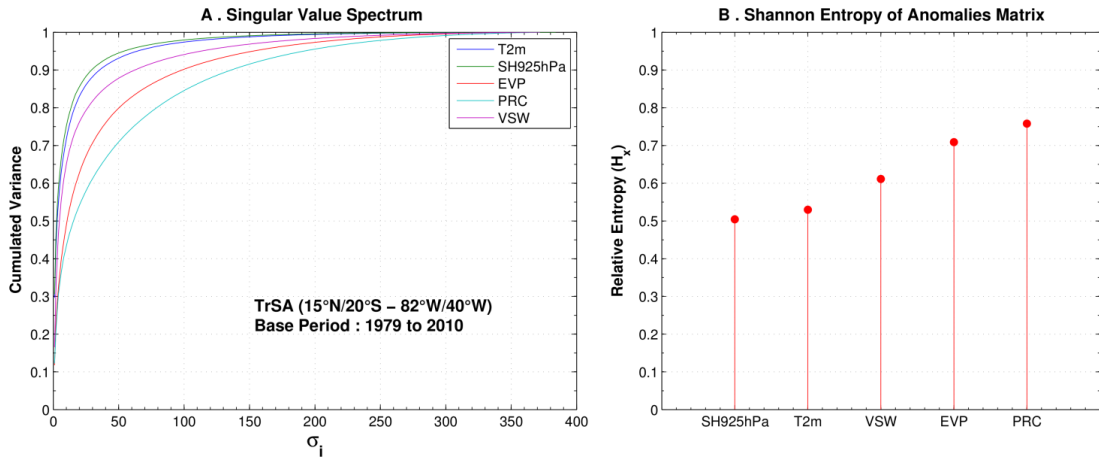

Figure 4. (A) Spectrum of singular values expressed as a function of the cumulative variance, corresponding to SH[925], T2m, VSW, EVP and PRC over TropSA during 1979-2010. **(B)** Relative entropy associated with each matrix of anomalies X (eqn. 8)

3.2. Maximum Covariance States of Land Surface-Atmosphere Feedbacks (LAF)

First, we use the set of reduced variables over TropSA (Table 1) to derive all possible pairs of truncated matrices $\hat{X}_{[mxn]}$ and $\hat{Y}_{[qxn]}$ for a crosslink analysis. With these pairs, we estimate a total of 10 covariance matrices, C_{XY} (Table 2). Then, we apply the Maximum Covariance Analysis (*MCA*, eqn. 9) to each C_{XY}, to find the k spatiotemporal patterns associated with each pair $\hat{X}$ and $\hat{Y}$. These k patterns between two variables define the Maximum Covariance States (*MCS_k*), which in turn are represented by the fraction of contained variance (f_k^{XY}, eqn. 10), and by the time series x_k and y_k representing the common pattern of $\hat{X}$ and $\hat{Y}$ .in time (eqn. 11). Therefore, the correlation coefficients $\rho(x_k, y_k)$ between these series (x_k and y_k) represent the temporal coupling of the two variables for each *MCS_k*.

The first $k=3$ *MCS_k* capture more than 85% of the cumulative covariance, F_3^{XY}, of each C_{XY} (Table 3); moreover, each *MCS_k* of C_{XY} including SH[925] describes the 100% of the covariance $F_3^{XY} = 1$. Correlation coefficients among pairs also describe the strength of the crosslink between concurrent time series (Table 3). Consistently, the first three *MCS_k* are quite high for the C_{XY} pertaining to the pairs T2m - SH[925] and PRC-VSW.

We hierarchically sorted the *MCA* results using the first covariance fraction (f_1^{XY}, Table 2). The matrix C_{XY} for the pair T2m - SH[925] (both state variables) has the highest f_1^{XY} corresponding to 78% of its total covariance. In the humid tropics, these two variables are strongly related by the

Clausius-Clapeyron phase-transition curve concerning temperature and saturated water vapor pressure [1].

In contrast, both matrices C_{XY} of PRC-EVP (42 %) and PRC-VSW (41 %) concentrate the lowest fraction in the first relative fraction f_1^{XY} (Table 2). Besides, the latter matrices show the second relative fraction f_2^{XY} nearest to their f_1^{XY}; exactly, the f_2^{XY} of PRC-EVP is 34% (8% less than f_1^{XY}), and the f_2^{XY} of PRC-VSW is 36% (5% less than f_1^{XY}). Both f_2^{XY} are higher respect to the f_2^{XY} of T_{2m} - SH^{925}, whose estimated value is 16% (62% less than f_1^{XY}). These results put forward that the essential mechanisms of the hydrological cycle (PRC – VSW – EVP) define two equilibrated maximum covariance modes over TropSA cumulating 75% of the covariance F_2^{XY} (last two rows of Table 2). In contrast, the covariance matrix between T_{2m} and SH^{925} highly concentrate most of the variance in the first singular value.

Table 3. Results of the MCS_k, for k=1,2, and 3, of all possible covariance matrices C_{XY}. Fraction of relative and cumulative square covariance (f_k^{XY} and F_k^{XY}), simultaneous correlation between time series, $\rho(x_k,y_k)$, and Shannon entropy, H_{XY}.

C_{XY}		MCS1			MCS2			MCS3			H_{XY}
$\hat{X}$	$\hat{Y}$	f_1^{XY}	$\left(F_1^{XY}\right)$	$\rho(x_1,y_1)$	f_2^{XY}	$\left(F_2^{XY}\right)$	$\rho(x_2,y_2)$	f_3^{XY}	$\left(F_3^{XY}\right)$	$\rho(x_3,y_3)$	
SH^{925}	T_{2m}	**0.78**	(0.78)	0.95	0.16	(0.94)	0.96	0.06	(1.00)	0.94	0.09
EVP	T_{2m}	0.64	(0.64)	0.54	0.23	(0.87)	0.63	0.08	(0.95)	0.53	0.14
PRC	SH^{925}	0.63	(0.63)	0.75	0.29	(0.92)	0.51	0.08	(1.00)	0.49	0.12
EVP	VSW	0.59	(0.59)	0.76	0.18	(0.77)	0.62	**0.12**	(0.89)	0.61	0.18
EVP	SH^{925}	0.59	(0.59)	0.52	0.31	(0.90)	0.60	0.10	(1.00)	0.51	0.13
PRC	T_{2m}	0.59	(0.59)	0.69	0.32	(0.91)	0.54	0.06	(0.97)	0.50	0.13
SH^{925}	VSW	0.58	(0.58)	0.67	**0.36**	(0.94)	0.50	0.06	(1.00)	0.38	0.12
T_{2m}	VSW	0.57	(0.57)	0.62	0.33	(0.90)	0.55	0.06	(0.96)	0.55	0.14
EVP	PRC	0.42	(0.42)	0.70	0.34	(0.76)	0.72	0.11	(0.87)	0.70	**0.23**
PRC	VSW	0.41	(0.41)	0.77	0.36	(0.75)	0.77	0.08	(0.88)	0.78	0.20

Precisely, given that the matrix C_{XY} corresponding to the pair of process variables, PRC and EVP, exhibit the largest entropy H_{XY} (Table 3), the fractions of covariance are proportionally distributed across the spectra of singular values σ_{XY} (Table 2). Such complexity in the interaction between process variables is also pointed out by [47]; regarding the multiple interactions that connect PRC within the LAFs so as to EVP returns humidity to the lower atmosphere (Figure 1).

Regarding the interactions between the hydrological cycle variables (PRC, VSW and EVP), the squared covariance fraction of the C_{XY} between EVP and VSW captures $f_{XY1} = 59\%$ (Table 2); which means that EVP has 17% more covariance with VSW than PRC in the first MCS_k. Such a strong covariability between soil moisture (VSW) and evaporation (EVP) in TropSA reinforces the idea that droughts are also associated with LAFs at local scales, and with teleconnections at larger scales [23, 73]. At this point, we hypothesize that the high entropy (H_X) from hydrological cycle variables crosslinking PRC–VSW-EVP arises as a consequence of the recycled precipitation that physically enhances non-linear interactions among those variables at TropSA [2, 13, 34, 44].

In the context of LAFs over TropSA, the strong coupling between VSW and EVP is also confirmed through cross-correlation maps [$\rho(y_k,\hat{X})$ and $\rho(x_k,\hat{Y})$] between the obtained time series and their respective truncated $\hat{X}$ and $\hat{Y}$ matrices described by the MCS_1 (Figure 5, with $\hat{X} = EVP$

and $\hat{Y} = VSW$). We infer that the core of such interaction is located in the Amazon River basin, in connection with the Magdalena-Cauca River basin over north-western TropSA; whereas, some regions of the Orinoco and Tocantins River basins do not show statistically significant values (P < 1%). Spatially, the MCS_1 exhibits more cells with higher significant correlations over TropSA between the series x_k (EVP) and $\hat{Y}$ (VSW), than the inverse case (Figures 6B and 6C). Both maps exhibit positive correlations, thus suggesting that negative (positive) anomalies of soil moisture go along with negative (positive) anomalies in evaporation. For one-month lag, the values of the correlation coefficients suggest that the effect of VSW on EVP is higher than in the reverse sense (Figure 5D). This effect is coherent but less evident at 2 months-lag. Furthermore, the cross-correlation between both series is high up to 6 months-lag.
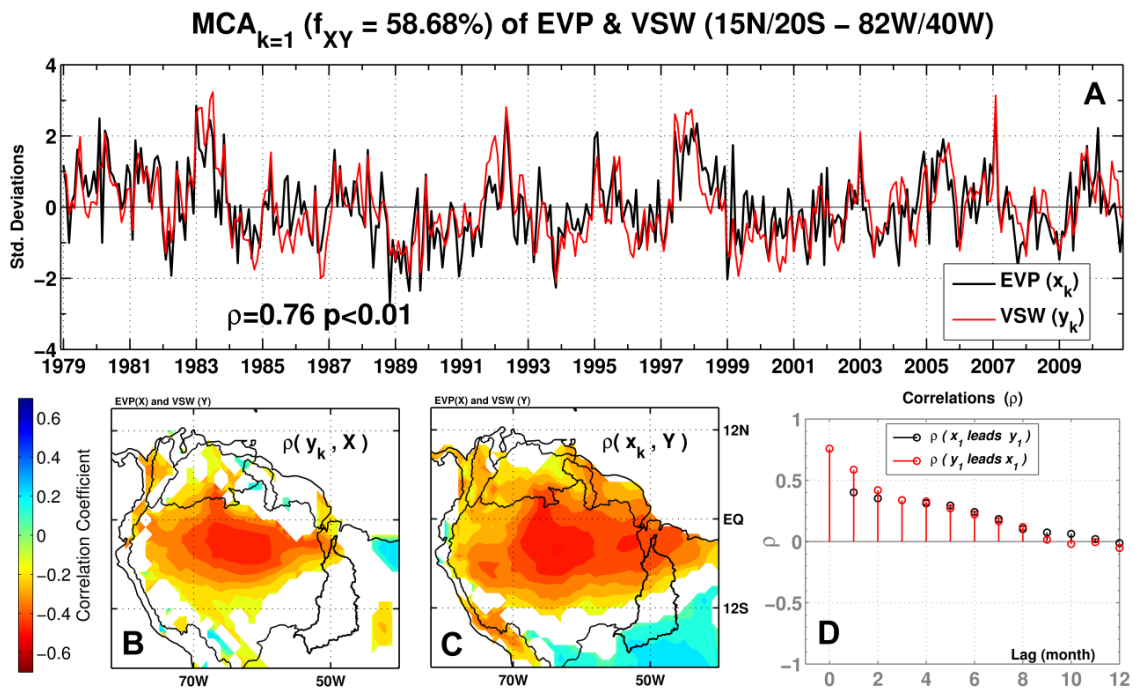

Figure 5. First Maximum Covariance State (MCS_1) between evaporation (EVP) and soil moisture (VSW) over TROSA. The correlation -pattern shows a core over the Amazon River basin: **(A)** Time series x_1 and y_1 associated with MCS_1 during 1979-2010; **(B and C).** Correlation maps using series x_1 and y_1 of anomalies (elaborated according to Table 1). **(D).** Lagged cross-correlations between series x_1 and y_1.

Similarly, we found that specific humidity in the lower atmosphere (SH^{925}) is the state variable most well-combined with precipitation (PRC) over the region on the MCS_1 ($f_{XY1} = 63\%$, Table 3). Comparing the time series x_1 and y_1 from MCS_1, we found that the interannual frequency of SH^{925}-PRC is higher than the frequency of EVP-VSW previously shown (Figure 6A). The spatial correlation pattern associated with this pair of variables (MCS_1 of SH^{925}-PRC) exhibits a dipole with one center in the northwest (direct correlations), and other in the southeast (inverse) (Figure 6B-C). In general, most cross-correlations are positive, $\rho = 0.5$, $P < 1\%$, over the Orinoco, Esequibo (Guyanas), and the northern Amazon River basin (Figure 6B-C); and correlations are negative over the southern Amazon and Tocantins River basins. This is a well-defined east-west dipole with a region of non-significant correlations in the meridional Amazon. The influence of atmospheric humidity on precipitation variance is significant over the Magdalena-Cauca River basin (Figure 6B), with more significant simultaneous cross-correlations (Figure 6C).

Finally, the main Maximum Covariance State (MCS_1) for the pair of process variables PRC-EVP concentrates a relative covariance fraction of $f_{XY1} = 42\%$. The MCS_1 of this pair shows a high

simultaneous correlation ($\rho = 0.7$) between x_1 and y_1 (Figure 7A). In addition, the spatial pattern of correlations that corresponds to the MCS_1 exhibits a northwest to southeast dipole (Figure 7B-C). This correlation pattern is similar to that shown in Figure 6B-C. In contrast, correlations over the northern pole of the MCS_1 of PRC-EVP are significant over the piedmont of Amazon-Orinoco River basins. Cross-correlations between series x_1 and y_1 of PRC-EVP (Figure 7D) are relevant just in the simultaneous case and almost null for lagged time steps correlations.
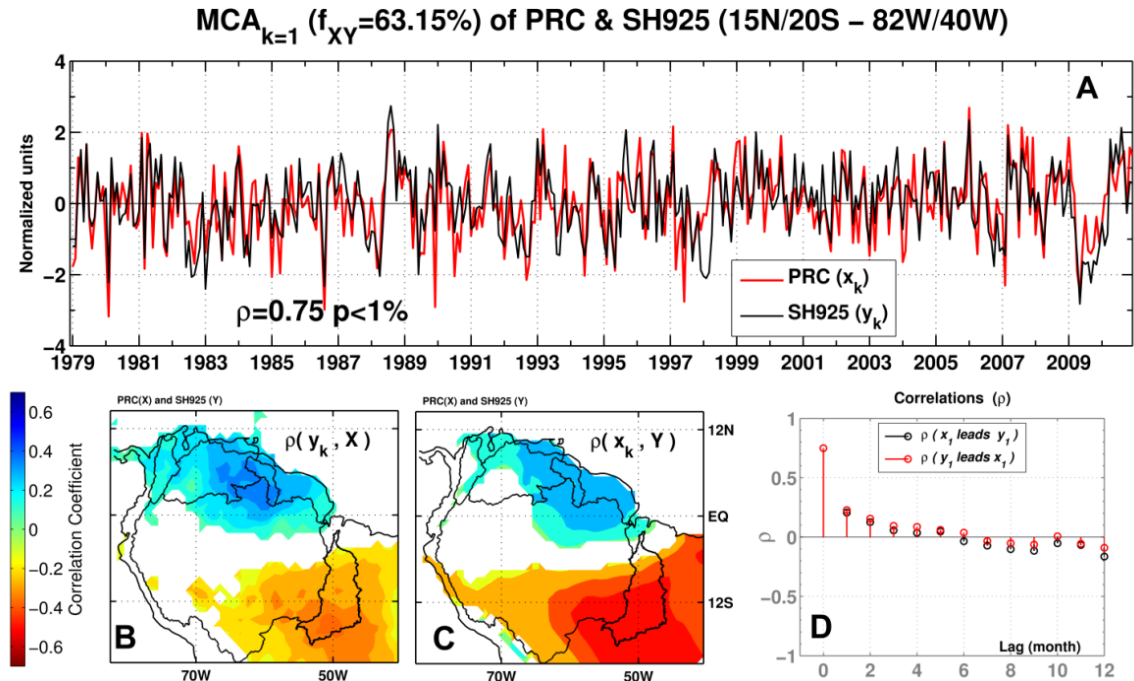

Figure 6. As in Figure 5 for precipitation (PRC) and specific humidity at 925 hPa (SH^{925}).
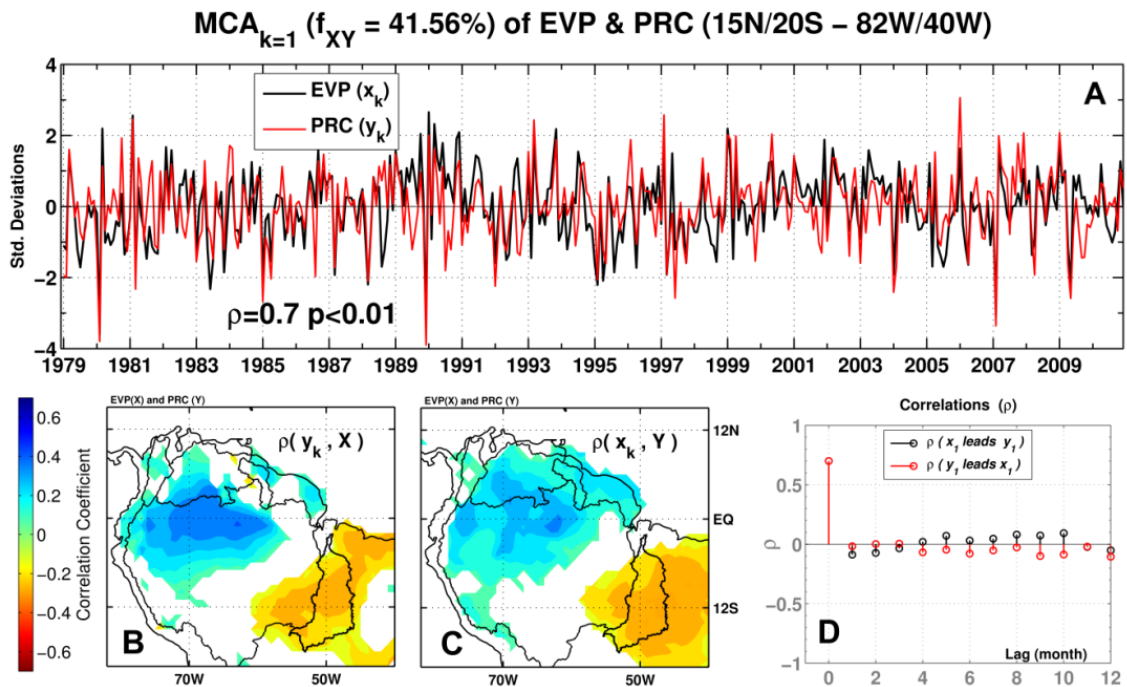

Figure 7. As in Figure 5 for precipitation (PRC) and evaporation (EVP).

3.3 Non-linear analysis of Maximum Covariance States

With the aim of quantifying the non-linear connections from the estimated MCS_k of each covariance matrix , we estimate the relative causalities $\tau_{y \to x}$ using each pair of time series x_k and y_k, for $k = 1,2,3$ (eqn. A12). For these main three MCS_k of each $\mathbf{C_{XY}}$, the simultaneous (non-lagged) causalities $\tau_{y_k \to x_k}$ and $\tau_{x_k \to y_k}$ are shown in Table 4. Similar to the results regarding the first fraction of the squared covariance (f_{XY1}) and correlations (ρ_{XY}), we find that the $\mathbf{C_{XY}}$ of SH^{925} -T$_{2m}$ in TropSA shows the highest relative causalities among the set of estimated values (Table 4). In this interaction, the T$_{2m}$ reaches up to -27 % of relative causality on SH^{925} for the MCS_1. As previously discussed in the methodology section, the negative sign of τ does not affect the amount of causality but the interpretation of the measure. According to [69], a negative value of τ indicates that surface temperature diminishes the entropy of specific humidity at 925 hPa on this particular Maximum Covariance State. However, this is only true for the first mode, but not so for second and third $MCSs$. After SH^{925} - T$_{2m}$, the PRC-VSW interaction shows the highest causality values, τ . Through the three main MCS_k, VSW consistently subordinates PRC (Table 4, last row). Particularly for MCS_1, VSW exerts the maximum causality on PRC among the MCS_k of this pair of variables: $\tau_{VSW \to PRC}$ is 18.1 % while $\tau_{PRC \to VSW}$ is 1.4%, thus implying a coupling $\mathbf{\Phi}_{VSW \leftrightarrow PRC} = 12.9$. Physically, these non-linear connections suggest that VSW is fundamental in modulating the interannual variability of PRC over TropSA.

In the Table 4, we highlight in red those values of Φ (eqn 14) within the interval $0.8 \leq \Phi \leq 1.2$ that define a bidirectional feedback. Having in mind this interval, and with respect to the second and third $MCSs$, bidirectional feedbacks are more frequent in the first MCS simultaneous causalities. Surface temperature (T$_{2m}$) is the most frequent variable in those pairs of feedbacks (SH^{925} - T$_{2m}$, EVP - T$_{2m}$, T$_{2m}$ - VSW). Additionally, the MCS_1 of the pair T$_{2m}$ -VSW shows the nearest value to a bidirectional feedback ($\Phi_{y_1 \leftrightarrow x_1} = 1.1$ Table 4). The main conclusion of these results is that the T$_{2m}$ is essential in capturing the non-linear interactions at interannual timescale within the LAFs of TropSA.

Table 4. Estimated values of simultaneous relative causality, τ, and feedback $\phi_{y_k \leftrightarrow x_k}$ for the first three $MCSk$, (estimated through eqn. A12) for all pairs of variables over TropSA (1979 - 2010)..

C_{XY}		MCS_1			MCS_2			MCS_3		
$\hat{X}$	$\hat{Y}$	$\tau_{(x_1 \to y_1)}$	$\tau_{(y_1 \to x_1)}$	$\Phi_{y_1 \leftrightarrow x_1}$	$\tau_{(x_2 \to y_2)}$	$\tau_{(y_2 \to x_2)}$	$\Phi_{y_2 \leftrightarrow x_2}$	$\tau_{(x_3 \to y_3)}$	$\tau_{(y_3 \to x_3)}$	$\Phi_{y_3 \leftrightarrow x_3}$
SH^{925}	T$_{2m}$	**22.4**	**-26.9**	**-1.2**	**-22.8**	15.1	-0.7	**-26**	15.2	-0.6
EVP	T$_{2m}$	12.2	10	**0.8**	0.1	**-5.2**	-52.0	6.4	4.9	**0.8**
PRC	SH^{925}	8.3	**-9.7**	**-1.2**	5.1	**-1.1**	-0.2	7.7	**-18.1**	-2.4
EVP	VSW	11.4	9.2	**0.8**	5.1	1.9	0.4	10.7	8.3	**0.8**
EVP	SH^{925}	10.4	7.4	0.7	**-2.8**	**-5.5**	2.0	6.0	-2.2	-0.4
PRC	T$_{2m}$	5.4	**-3**	-0.6	8.7	**-2.4**	-0.3	5.2	**-9.2**	-1.8
SH^{925}	VSW	7.6	13.2	1.7	3.8	7.1	1.9	**-1.4**	8.4	-6.0
T$_{2m}$	VSW	9.8	10.8	**1.1**	-2.6	11.5	-4.4	3.8	10.5	2.8
EVP	PRC	**-16.2**	**-7.6**	0.5	**-6.6**	5.5	**-0.8**	5.2	14.3	2.8
PRC	VSW	1.4	18.1	12.9	14.7	**21.6**	1.5	2.6	10.5	4.0

3.4 Grouping the Maximum Covariance States through ENSO

First, we assess the influence of ENSO on the main Maximum Covariance States (MCS_k) during 1979-2010. To that end, we correlate the series x_k and y_k with the Niño 3.4 index (Table 5). Then, according to the magnitude of these correlations, we hierarchically categorize the ENSO effect on the MCS_k of each $\boldsymbol{C}_{XY}$ in three groups d = {1, 2, 3}: high (blue), medium (green) and low (red). This categorization does not necessarily correspond to the k-sorting of the MCS originally coming from the covariance fractions (f_k^{XY}). For example, those MCS_1 of $\boldsymbol{C}_{XY}$ that include precipitation (PRC) are not the best correlated with N3.4 (Table 5). This means that the maximum fraction of covariance of the different pairs is not homogeneously organized according to the influence of different oceanic modes, and depends on the interaction of the variables.

Also, these results allow us to identify which specific $\boldsymbol{C}_{XY}$ exhibits the highest association with ENSO (blue entries in Table 5). From those highly ENSO-related MCS (d=1), the largest one corresponds to the covariance matrix of VSW ($\hat{X}$) - T2m ($\hat{Y}$) of the MCS_1 [$\rho(y_1, N3.4) = -0.55$, Table 5]. Again, the T2m - VSW interaction is quite relevant within the LAFs as the most ENSO-influenced among the paired variables. This result indicates that soil moisture (VSW) influences the atmospheric response to surface warming (T2m), and, in combination with atmospheric moisture (SH925), VSW contributes to determine the spatiotemporal pattern of PRC [74]. Regularly, the pairs that involve T2m exhibit the highest correlation values with the Niño 3.4 index (Table 5); surface temperature is essential to propagate the ENSO-effect on variables involved on LAF over TropSA.

Table 5. Correlation between series x_k and y_k with the Niño 3.4 index, $\rho(x_k, N3.4)$, and $\rho(y_k, N3.4)$ for the MCS_k, k=1,2, and 3. High correlation values are denoted in blue, medium values in green and low values in red.

$\boldsymbol{C}_{XY}$		MCS_1		MCS_2		MCS_3	
$\hat{X}$	$\hat{Y}$	$\rho(x_1, N3.4)$	$\rho(y_1, N3.4)$	$\rho(x_2, N3.4)$	$\rho(y_2, N3.4)$	$\rho(x_3, N3.4)$	$\rho(y_3, N3.4)$
SH925	T2m	-0.44	-0.50	0.15	0.13	-0.20	-0.17
EVP	T2m	-0.48	-0.52	0.11	0.12	0.14	0.14
PRC	SH925	-0.25	-0.31	0.40	0.45	-0.20	0.10
EVP	VSW	0.4	0.44	0.32	0.36	0.11	0.18
EVP	SH925	-0.44	-0.44	0.20	0.21	0.22	0.05
PRC	T2m	0.32	0.43	-0.32	-0.45	-0.17	0.03
SH925	VSW	-0.43	-0.47	0.36	0.35	-0.10	-0.03
T2m	VSW	-0.52	**-0.55**	-0.30	-0.21	-0.05	-0.09
EVP	PRC	-0.08	-0.19	-0.48	-0.4	-0.22	-0.31
PRC	VSW	0.3	0.37	-0.35	-0.4	0.21	0.21

Besides, each $\boldsymbol{C}_{XY}$ involving PRC (PRC- SH925, PRC- T2m, EVP-PRC, PRC-VSW) consistently shows the highest correlations with N3.4 in the MCS_2 instead of MCS_1 (Table 5, blue entries column MCS_2). This means that the MCS_1 of each $\boldsymbol{C}_{XY}$ involving PRC, and in contrast with each $\boldsymbol{C}_{XY}$ involving T2m, are not fully related to ENSO as shown with the other $\boldsymbol{C}_{XY}$ (Table 5). The correlation coefficients of the PRC- T2m and PRC-VSW pairs with N3.4 are almost the same for MCS_1 and MCS_2 (Table 5), which implies that the ENSO effect is very similar in both Maximum Covariance States. This is especially significant for the PRC-VSW connection because in Section 3.2 we established that the squared covariance fractions f_1^{XY} and f_2^{XY} are also very similar. The PRC- SH925 pair presents

a clear difference of correlation values among the three MSC_k with N3.4; in that case, the MSC_2 of PRC- SH[925] is best correlated with N3.4 having a f_2^{XY} of 30%. The pair EVP and PRC exhibits the lower correlation with N3.4 among the set of MCS_1 (Table 5, red entries in column MCS_1), which implies that the interactions between those process variables at interannual time scales are not completely due to ENSO as we show for the state variables SH[925] and T[2m].

3.5 ENSO-intensities Categorizing Interannual Oceanic Modes

Once we grouped each Maximum Covariance State (MCS_k) through correlations with ENSO, we correlate the monthly fields of global SSTs with each pair series x_k and y_k during the study period 1979-2010. To that end, we used the HADISST product from the Hadley Center. We show one representative SST map per each one of these groups: high ENSO-related using the MCS_1 EVP-VSW (Figure 8A), medium ENSO-related using the MCS_1 PRC- SH[925] (Figure 8B), and low ENSO-related using the MCS_1 EVP-PRC (Figure 8C). Simultaneously, using the same time series, such patterns correspond to the TropSA correlations maps shown in Figures 5 to 7. Each group of MCS_k series (high, medium, and low ENSO-related) correlate at global scale with a particular SSTs pattern (Figure 8A-C).

We conclude that Maximum Covariance States (MCS_k) capture the LAFs determined by the conjoint (non-isolated) action of few SST-modes (Figure 8A). The conjoint effect is triggered because the Amazon River basin occupies the largest portion of TropSA, and, at interannual time scales, several SST-modes determine the characteristics and magnitude of the LAFs therein [59; 75-77].

First, we analyzed the time series group which best correlates N3.4 (High ENSO-related, blue entries in Table 5). Those series exhibit highly statistically significant correlations with the SSTs over the different Niño regions (Figure 8A). Additionally, the Niño-SSTs act in phase (same sign of coefficients) with SSTs over the Tropical North Atlantic (TNA), and with the Indian Ocean SSTs (IO) (Figure 8A). This global SST pattern also exhibits the influence of the Pacific Decadal Oscillation (PDO) acting in phase-shift with Niño regions correlations (reverse sign of coefficients).

The second group of MCS_k (medium ENSO-related) captures in a dipole the ENSO effects and an equatorial oceanic mode over the Tropical Atlantic (Figure 8B). Consistently, this SST-pattern represents the covariance among variables pairs involving precipitation (Table 5). The described SST-pattern connects with the correlation dipole over the TropSA continental zone shown in Figure 6 for the MCS_k PRC-SH925.

The third group (low ENSO-related) presents the weaker correlations with global SSTs respect to the other groups (Figure 8C). The SST-pattern of correlations is dominated by the interannual variability over the Eastern Tropical Pacific (Niño 1+2 regions) and over the TNA region (Figure 8C). Both regions act in phase as per the same sign of correlations (Figure 8C). Likewise, this SST-pattern connects the interactions inside the MCS_1 of LAFs process variables PRC-EVP, whose $f_1^{XY} = 0.41$, with the northeastern to southeastern dipole at TROSA shown in Figure 7C.

3.6 Visualizing Dependences Among Variables with Graph Theory

To synthesize and further understand the LAFs structure including *state* and *process* variables at interannual time scale over TropSA, we constructed graphs Γ_d for the SST-mode depicting the highest correlation with the Niño 3.4 index along the MCS_k time series (Table 5). As we stated in section 3.4, this group Γ_1 captures the MCS_k best correlated with the Niño 3.4 index from each C_{XY} matrix (Table 3).

We define a threshold for the magnitude of the edges to construct correlation and causality graphs Γ_d. This allow us to study the strongest connections among the studied variables. In all cases,

we selected the percentile of 50% of the empirical probability distribution of causalities and correlations values for each time lag (Figure 9, Table 6).
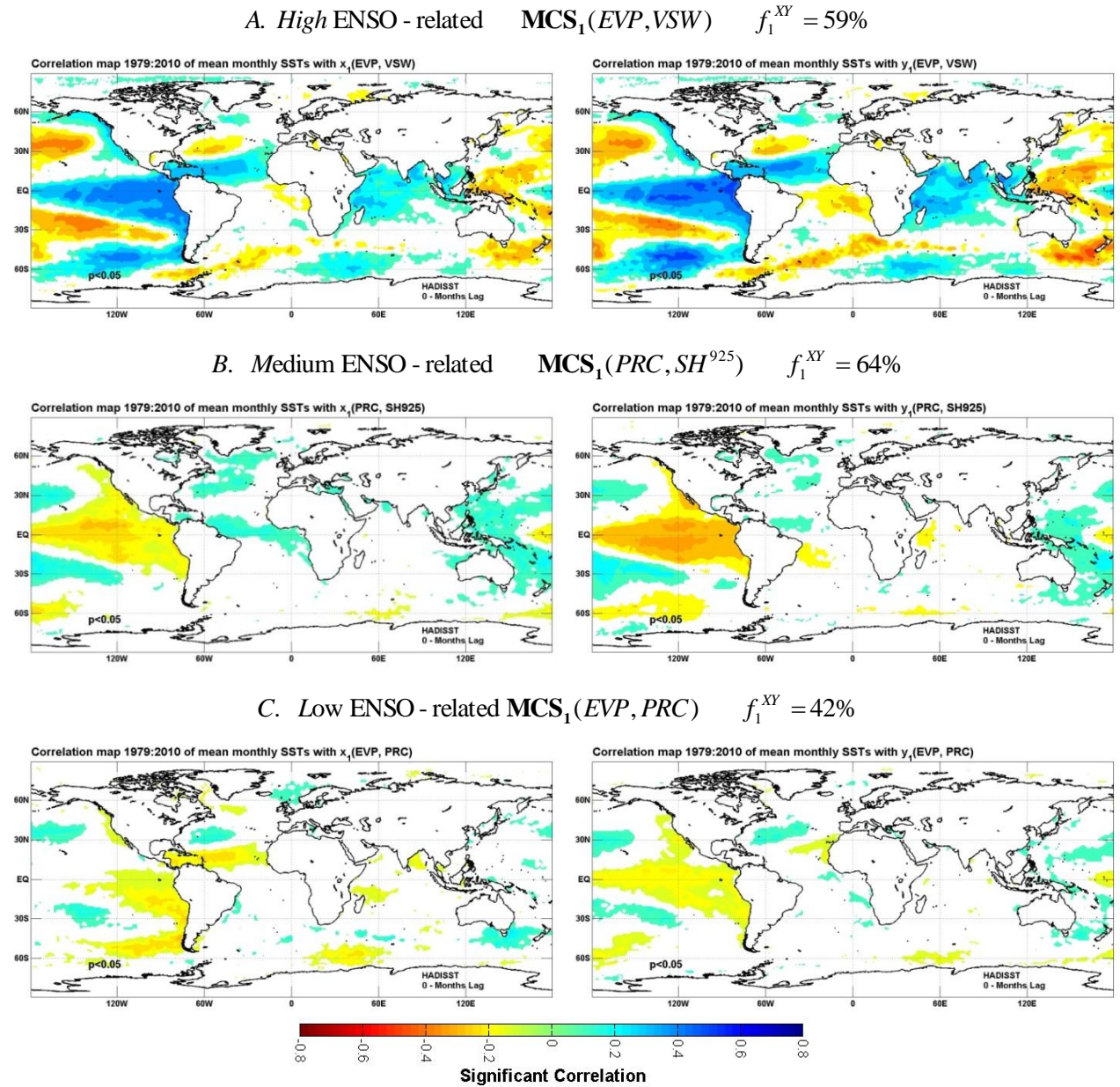
A. *High* ENSO - related $\quad$ $\mathbf{MCS_1}(EVP, VSW)$ $\quad$ $f_1^{XY} = 59\%$

B. *M*edium ENSO - related $\quad$ $\mathbf{MCS_1}(PRC, SH^{925})$ $\quad$ $f_1^{XY} = 64\%$

C. Low ENSO - related $\mathbf{MCS_1}(EVP, PRC)$ $\quad$ $f_1^{XY} = 42\%$

Figure 8. Maps of simultaneous correlations between SST anomalies and the time series associated with the first Maximum Covariance State (MCS_1): A. EVP - VSW (series in Figure 5A); B. PRC-SH925 (series in Figure 6A); C. EVP - PRC (series in Figure 7A).

Each node graph from Γ_d represents the variables inside the MCS_k in a common SST-mode crosslinking state and process variables. Using series x_k and y_k from each MCS_k, we represented the edges of the graph Γ_d as correlations (linear coupling; see Figure 10), and as information flow or causality (non-linear coupling; see Figure 10). Without lagging series x_k and y_k, the correlation coefficient is a single value. Conversely, causality takes two senses: $x_k \rightarrow y_k$ and $y_k \rightarrow x_k$, and thus, both do not necessarily have the same value. Consequently, we coupled graphs and feedbacks lagging the series from 1 to 9 months (edges as correlation coefficients, Appendix B).
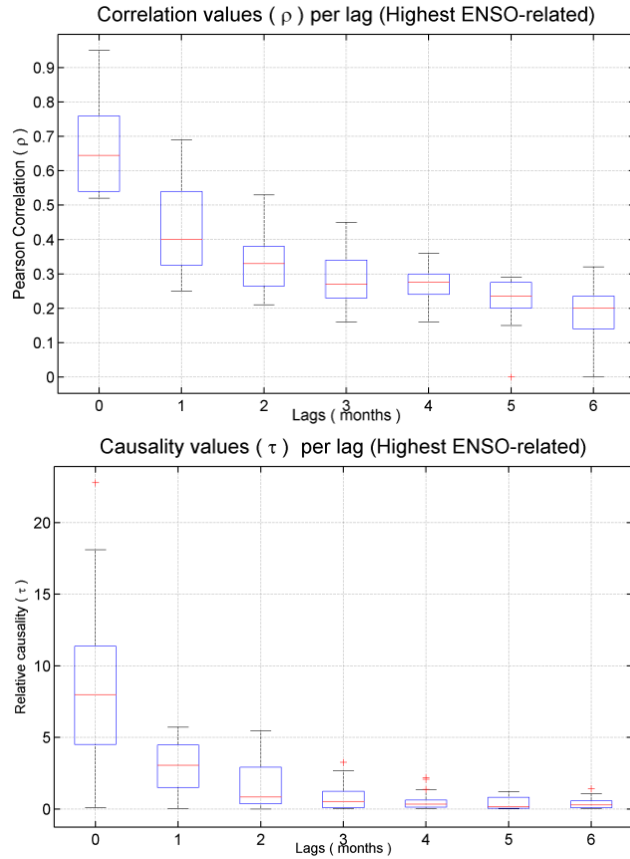
Figure 9. Thresholds to define weight of graphs edges as correlations (up) and causalities (bottom) for the highest ENSO-related group

Table 6. Threshold of edges magnitude to construct correlation and causality graphs (Percentile of 50% of the empirical probability distribution of causalities and correlations per lag)

Metric type	Group	Lag						
		0	**1**	**2**	**3**	**4**	**5**	**6**
Correlation	*Highest*	0.64	0.4	0.33	0.27	0.275	0.235	0.2
Relative Causality (%)	*ENSO-related*	9.93	4.8	3.44	3.94	2.69	1.75	1.08
Correlation	*Medium*	0.65	0.31	0.225	0.18	0.155	0.135	-
Relative Causality	*ENSO-related*	5.0	2.75	1.52	1.14	0.84	0.18	0.13
Correlation	*Lowest*	0.58	0.23	0.23	0.15	-	-	-
Relative Causality	*ENSO-related*	1.37	3.05	0.83	0.53	0.17	-	-

Using this methodology, we constructed a graph for each lag from 1 to 9 months, and then we extracted the adjacency matrix $W_{[ixj]}$ (eqn. A14) of each $\varGamma_d$. To study the links structure among all variables connected by each graph, we factorized each adjacency matrix $W_{[ixj]}$ applying a SVD as shown in eqn. A14 . Then, following the HITS algorithm proposed by [88], we systematically extracted the highest entries of the first singular vectors pair (u_1 and v_1) from each matrix factorization. These are essential values because, in this context, the adjacency matrix $W_{[ixj]}$ directly connects the whole set of graph nodes crosslinking their columns (*u*/emitters) and rows (*v*/receivers). As the matrix $W_{[ixj]}$ defines all the possible associations among nodes, the variable (node) corresponding to the maximum vector u_1 value transferring (emitter) the highest amount of

information to the whole set of variables, while the maximum vector v_1 value is the best information receiver. Moreover, we present the singular value spectrum containing the cumulative variance from the adjacency matrix $W_{[ixj]}$ (Figures 10A, 10B, and 10C). Red nodes in each graph indicate the highest value inside the singular u_1 and v_1 estimated from the SVD adjacency matrix. We respectively show those red graph nodes as red horizontal bars in Figure 10A-C-right. Likewise, red paired-edges in each graph represent correlations and causality ratios (eqn. 14) ranged between 0.9 and 1.1. With these conventions, we encapsulate the concept of feedback or equal-bidirectional effect in each graph. We used both metrics (correlation and causality) comparing systematically linear and non-linear couplings among *MSCs* ensembles.

3.6.1 Linear coupling Graphs

To illustrate the proposed graph analysis, Figure 10A shows the 1 month-lag correlations graph Γ_1 with significant *p*-values < 0.01 (group Γ_1 high ENSO-related). This graph Γ_1 associates the SST-pattern established in Figure 8A with the spatial pattern depicted in Figure 5 over TropSA as interactions among nodes (variables). Through this graph, whose edges are 1-month lagged correlations higher than the median (ρ=0.4, Table 6), we identify a structural interaction between VSW- SH^{925} (soil moisture-atmospheric humidity relation) which consistently are the highest entries of vectors u_1 and v_1 (Figure 10A-right).
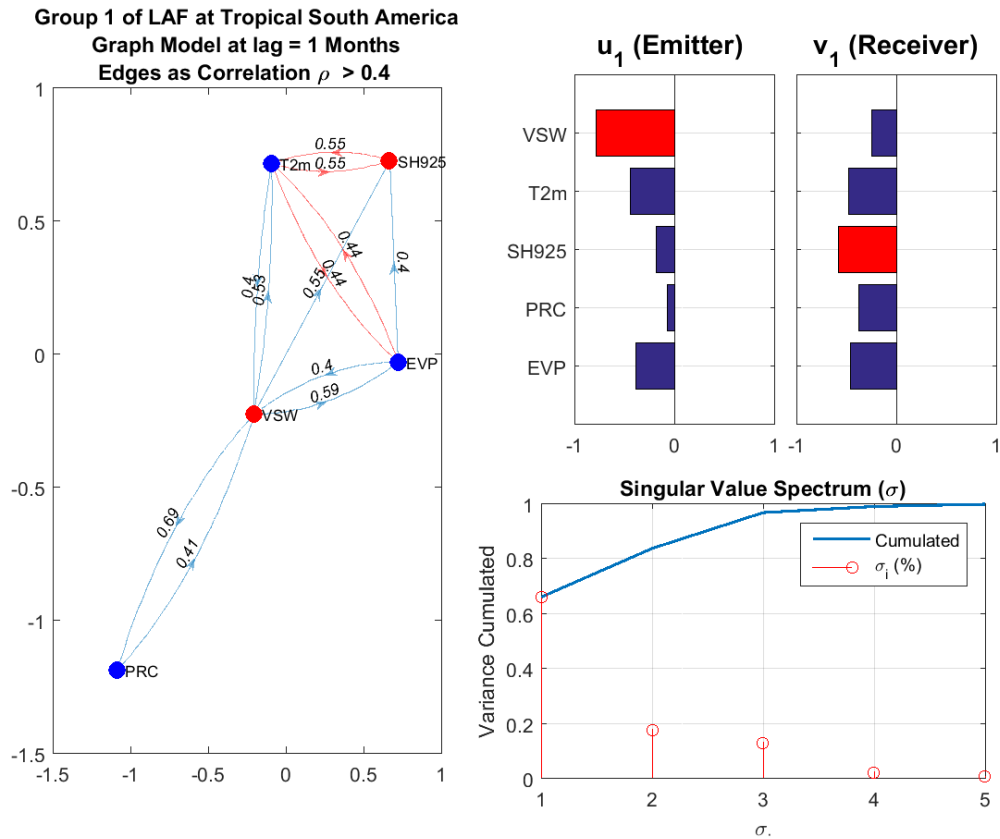

Figure 10A. (Left) Correlation graph, Γ_1, constructed with the highest ENSO-related *MSC*$_k$ time series from each pair of variables (corresponding to the SST-pattern shown in Figure 8A). The weights of the edges between nodes are estimated as the 1 month-lag correlation between the series x_k and y_k from each *MSC*$_k$. Red edges represent feedbacks between variables. **(Right-Up)** Vectors u_1 and v_1 estimated from the SVD of the adjacency matrix associated with the graph Γ_1, red bars represent the highest and lowest values of these vectors; **(Right-Bottom)** Spectrum of singular values estimated from the SVD on the adjacency matrix associated with the graph Γ_1

We catalogue this interaction as structural because the first singular value of the adjacency matrix of graph Γ_1 (high ENSO-related, 1-month lagged correlations) captures a high fraction of its variance (80%). This high percent characterizes the relations assembly expressed by vectors u_1 and v_1 leading the most of the linear coupling expressed by the graph Γ_1. Our results point out that T2m, SH925, PRC and EVP contain an important chain of feedbacks favoring the regulation of VSW within the structure of relations (Figure 10A). Firstly, VSW depicts the highest correlations among the variables at 1-month lag. Aditionally, PRC is just strongly connected with VSW. Led by the ENSO forcing, VSW acts to trigger the LAFs mechanisms. Surface temperature is regulated by VSW directly and through linear feedbacks with EVP and SH925 (Figure 10A). SH925 does not influence directly VSW but through changes in T2m. Instead, VSW linearly controls changes in SH925 directly and through T2m and EVP.

Figure 10B shows the graph Γ_2 constructed with 2 months-lag correlations associated with the highest ENSO-related group. Again, this graph reflects a relational structure dominated by VSW as the best emitter (maximum value of u_1); but, in this case, EVP is the best receiver (maximum value of v_1). For this Γ_2, we found one more bidirectional relations among variables with respect to the Γ_1. T2m establishes those linear feedbacks with SH925, EVP and VSW. The highest correlation of the entire set shows the predominance of VSW on PRC (ρ=0.53) regulating the interaction of PRC with T2m, SH925 and EVP.
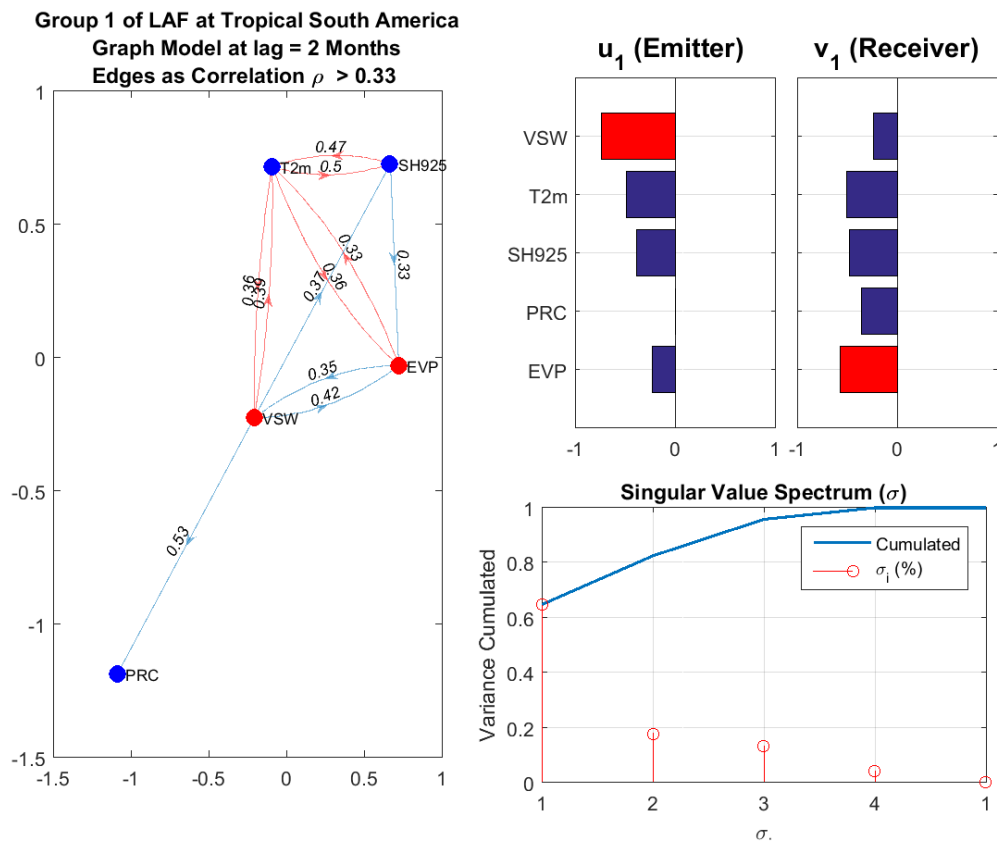

Figure 10B. As in figure 10A for 2 lag-months

Finally, Figure 10C shows the graph Γ_3 corresponding to 3 months-lag for the highest ENSO-related group. This analysis reveals a structural relation between surface temperature and evaporation (T2m- EVP). The estimated 1 month-lagged correlation between SHP925 and PRC is not strong in the graph. Again, represented in a hierarchical form, the structure of the graph Γ_3 evidences

that VSW constitutes a heterogeneous node connecting PRC with EVP, SH[925] and T[2m] (Figure 10C). The linear feedbacks are identified for the pairs EVP-VSW and T[2m]-SH[925] (Figure 10C).

In summary, using 1 to 3 month lagged correlation as a metric to link the graph nodes involving the state and process variables of LAFs over TropSA, we find that, structurally, soil moisture is the best regulating mechanism of interaction among the identified interannual modes.
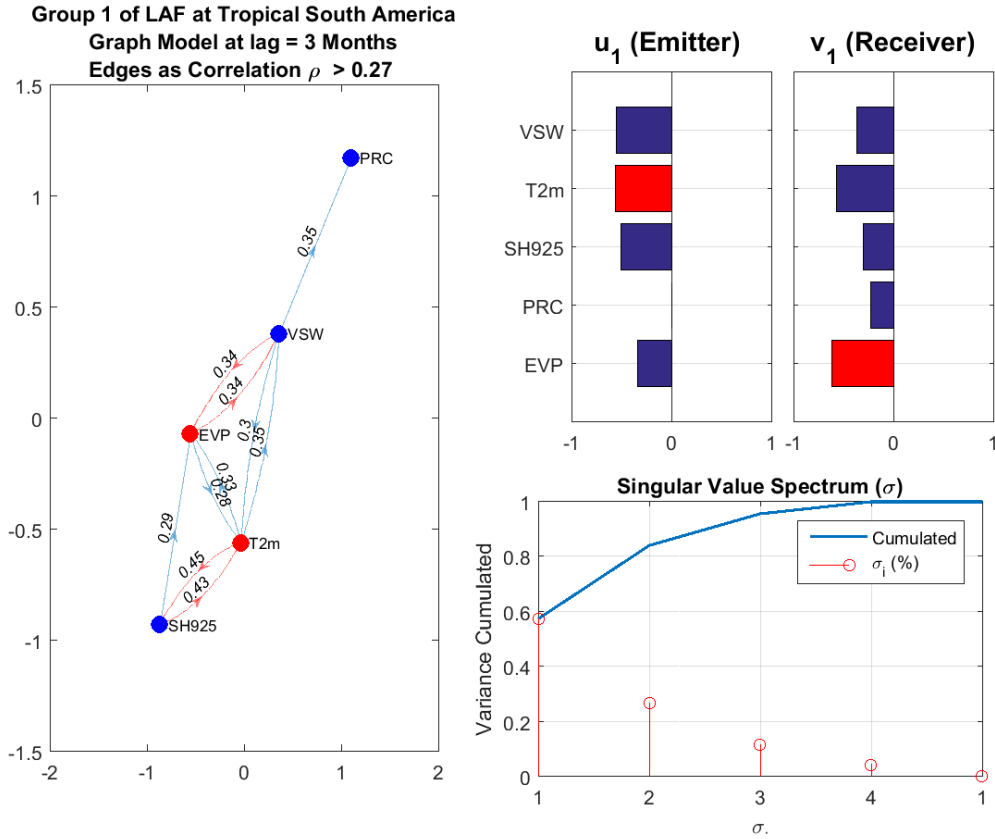

Figure 10C. As in figure 10A for 3 lag-months.

3.6.2 Non-Linear Coupling Graphs

We also explore a group of graphs whose edges denote causality relations ($\tau_{y_k \to x_k}$, non-linear associations). Figure 11 shows the simultaneous (non-lagged) graph for the high ENSO-related group defined previously. Such causalities are the same reported in Table 4. Contrary to correlations, relative causalities allow us to examine the relations structure without lagging. Because of non-linearities, relative causalities depend on direction between the two variables, given by $\tau_{y_k \to x_k}$ and $\tau_{x_k \to y_k}$ (Table 4).

The links structure is similar with those depicted by linear graphs of 1 month-lag (Figure 10A) and 2 months-lag (Figure 10B). VSW works as a transferring node variable connecting PRC with EVP, SH925 and T2m (Figure 11). The highest causality of this graph corresponds to the interaction T[2m] → SH[925] (26.88 %). Besides, the relational structure of this graph is nearly equally dominated by VSW as the best emitter (u_1), and the best receptor is SH[925] (Figure 11, red horizontal bars and red graph nodes), as we show for 1 month-lag linear graph (Figure 10A). At 0 month-lag (concurrent), VSW dominates non-linear relationships on PRC, SH[925] and T[2m] at interannual time scales. However, under this causalities structure, EVP influences VSW. VSW is essential over Tropical South America to structurally feedback on SH[925] under the group of highest ENSO-related Maximum Covariance

States at concurrent series (non-linear links). Therefore, we estimate the bi-directional influences and the symmetries between series without lags.
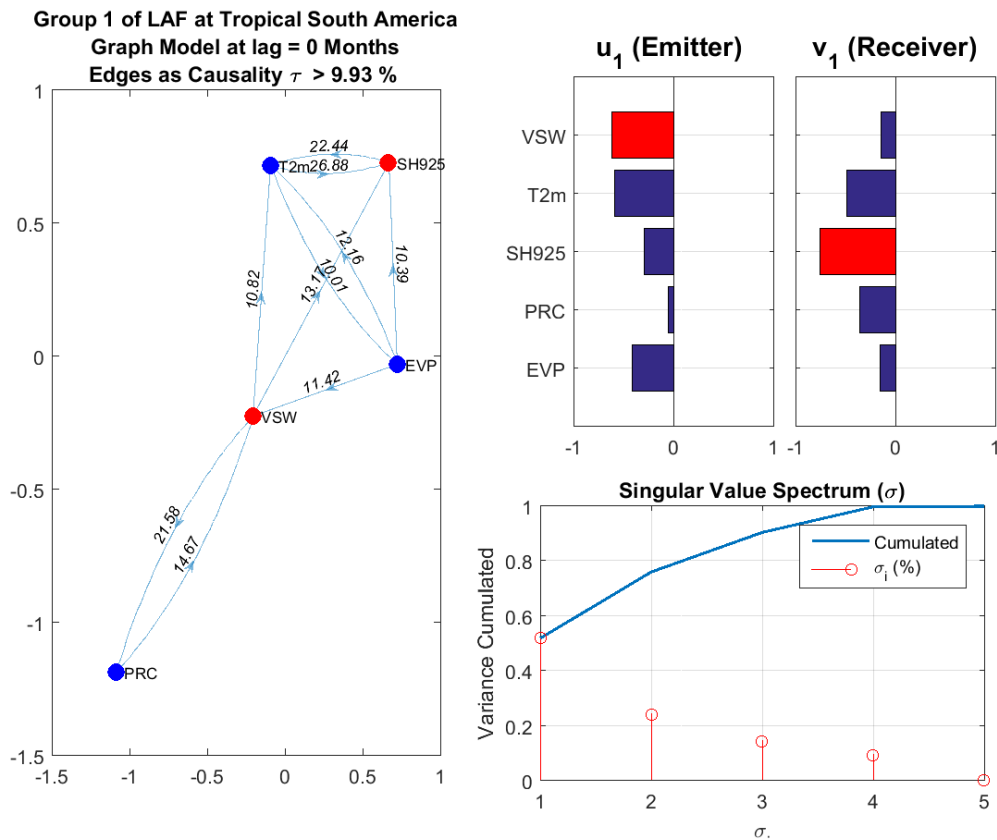

Figure 11. (Left) Causality Graph Γ_1 constructed with the highest ENSO-related MCS_k time series from each pairwise of variables (corresponding to the SST-pattern shown in Figure 8A). The weights of the edges between the nodes are estimated as the simultaneous causality ($\left|\tau_{x\to y}\right|$ and $\left|\tau_{y\to x}\right|$; eqn. A12) between the series

x_k and y_k from each. Red edges represent feedbacks between variables. **(Right-Up)** Vectors u_1 and v_1 estimated from the the adjacency matrix associated with the graph causality graph Γ_1; **(Right-Bottom)** Spectrum of singular values estimated from the SVD on the adjacency matrix associated with the graph Γ_1

3.7 Multivariate connectivity of LAFs at interannual timescale

Figure 12 shows a synthesis of the graphs constructed with the MCS for the three interannual modes (Group 1: High ENSO-related, Group 3: Low ENSO-related), correspondently with Figure 8, for each lag from 0 to 6 months, and for both metrics: correlation (Figure 12, left column) and causality (Figure 12, right column). Consistently, the rows in each box of Figure 12 constitute all possible interactions among the five variables; the columns represent a complete graph for each lag from 0 to 6 months. Figure 12 also represents with red filled circles the edges between two nodes feeding back on each other (feedbacks between variables). Therefore, this concept also corresponds to the edges feeding back in Figures 9 and 10.

For each graph lag, the blue square corresponds to the pair of variables with maximum values in the main vectors u_1 and v_1, that is, the pair of variables which control the relational structure of each graph. For example, a blue square is located in the pair SH^{925}-VSW for the 1 month-lag graph in Group 1 (edges as correlations). This square represents the maximum values of u_1 and v_1, in agreement with the graph shown in Figure 9A.

From this synthetic analysis and according to the graphs denoting correlations (Figure 12, left column), we found a direct relation between the number of feedbacks (defined as the fraction of

significant coefficients between 0.9 and 1.1), and the degree of correlation with N3.4 used to categorized graphs in three groups (High, medium, low ENSO-related). In other words, a larger influence of ENSO on the MCS_k favors the existence of feedbacks among variables. In the high ENSO-related graph group, EVP is always structurally involved in the connectivity making part of the pair with maximum u_1 and v_1 for several lags. On the other hand, the interaction VSW–EVP dominates the relations structure inside Group 2. Recurrently, this pair is associated with the maximum values of u_1 and v_1, for both correlation and causality in this group (Figure 12). Finally, the relations structure which defines a consistent pattern of LAFs appears in Group 3. For those graphs, and both metrics, the interaction of the pairs SH[925]-VSW dominates the vectors u_1 and v_1 for several lag-graphs. (Figure 12, Group 3).
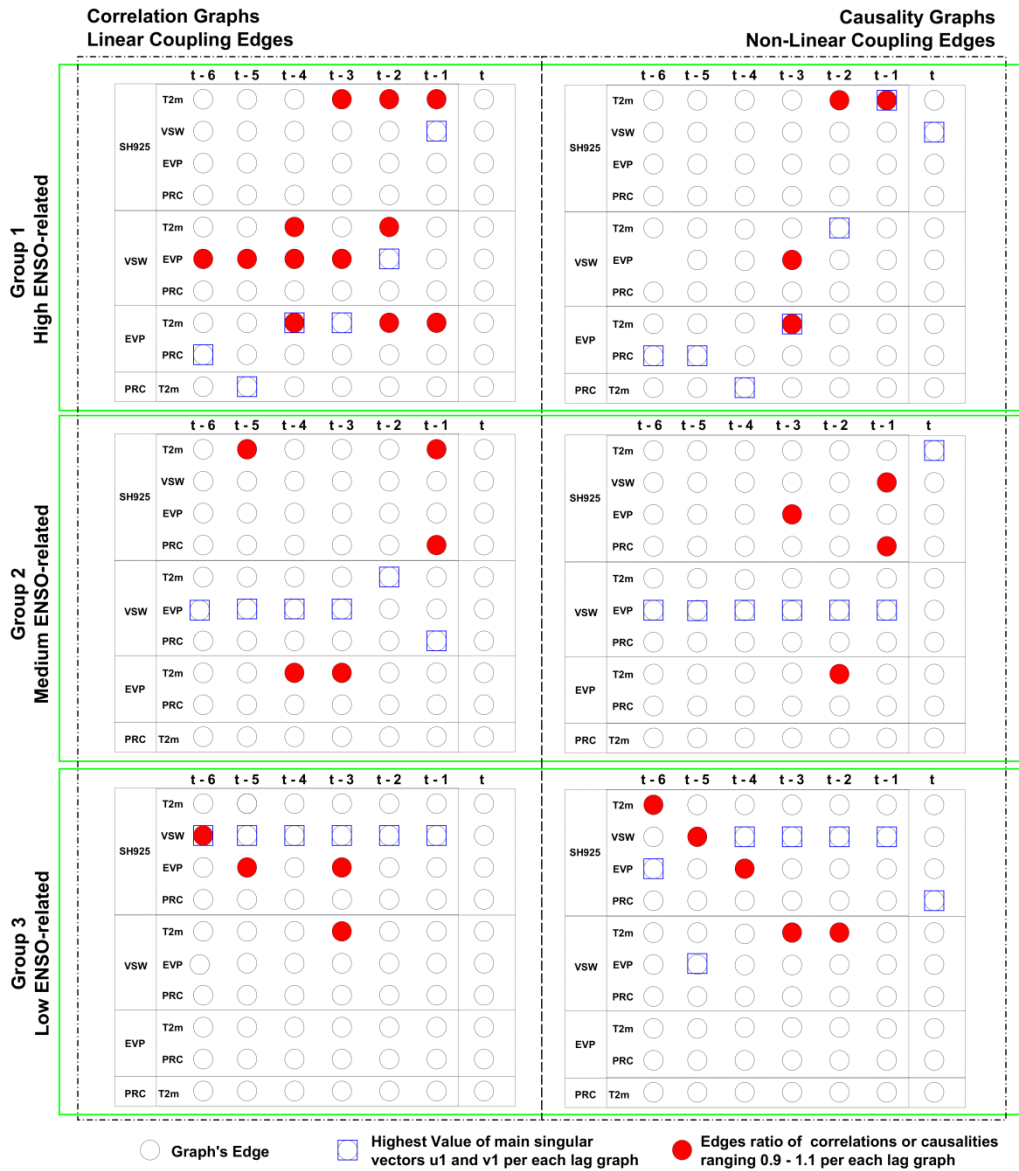

Figure 12. Synthesis of the structural analysis of graphs associated to each of the three groups established according to their correlation with the N3.4 index. We use linear (Correlation, left), and non-linear (Causality, right) metrics to establish the edges between graph nodes. Therefore, each column in each group represents the ten possible edges of the graph created for lags from 0 to 6 months.

As a final point, we estimate the algebraic connectivity, λ_2, (eqn. A18) for each graphs group previously defined by their correlation with the Niño 3.4 index (ENSO), and for the whole set of lags

from 0 to 9 months (Figure 13). The connectivity curves were constructed by summarizing the edges as correlations (Figure 13A) and as causalities (Figure 13B) per lag graph. Consistently for both metrics, the connectivity of the variables within the LAFs is higher if the MSC_k group is higher correlated with the Niño 3.4 index. In summary, the high ENSO-related graphs (Group 1, blue line in Figure 13A-13B) exceed the connectivity among variables in 25% with respect to the influence of Group 3 (Mode 3, green line). This effect augments the connectivity among variables up to a 7 month-lag. Likewise, the connectivity diminishes with the number of lags in the graphs (Figure 13).

The connectivity λ_2 is a measure of the structural resilience of a graph [78]. From this point of view, we demonstrate that ENSO has an important effect on the connectivity, and thus on the structural resilience of variables exchanges at interannual time scale on TropSA. For the high ENSO-related group, this resilience is basically the robust memory in time of the interactions among state and process variables favored by the SST-pattern presented in the figure 8A. It is worth recalling that, over the spatial domain of TropSA, the maximum connectivity of the variables of this high ENSO-related group is representative in regions of significant correlations shown in Figure 5 whose main action core is the Amazon rainforest and the north-west zone strongly influenced by the Choco Low-Level Jet [72]. At structural level, a large part of this connectivity resides in the soil moisture (VSW) as shown in Figure 12 for the linear and non-linear metrics. Non-linear metrics in the High ENSO-related group also reveal the essential role of surface temperature as a connectivity enabler between SH[925] and VSW for 0 to 3 month-lags. Besides, the structural power of process variables (PRC-EVP) in this group (High ENSO-related) is important from 3 to 6 months-lag, which means that the memory in the state variables are dominated by the previous conditions on process variables.

On the other hand, the second group of graphs connectivity (Medium ENSO-related) in time is dominated by the pair interaction VSW-EVP (see blue squares in Figure 12). The connectivity in the third group depends directly on the structural relation SH[925]-VSW (state variables). Therefore, VSW is an essential state variable favoring connectivity and resilience of the interactions on the complex mechanisms inside LAF over TropSA.
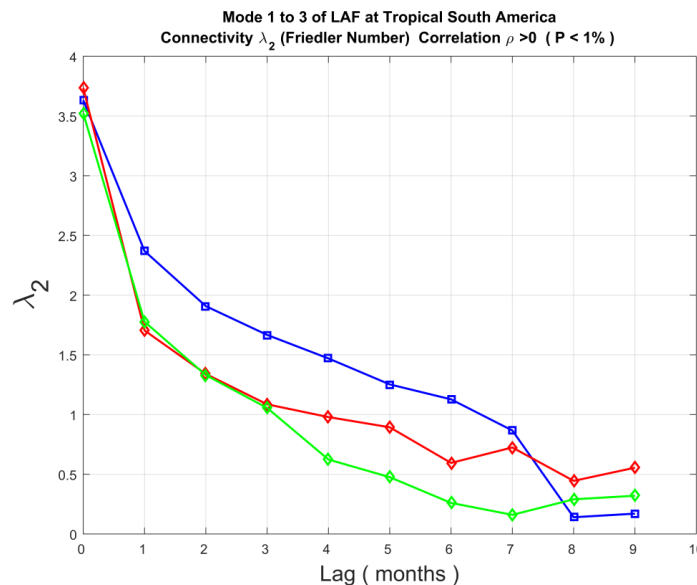

Figure 13A. Graph connectivity associated to each ENSO degree influence group over TROSA. Each lag represents a graph where the edges between the nodes are the correlation coefficients, $\rho(x_k, y_k)$. Blue: Highest correlations with N3.4; Red: Medium correlations with N3.4; Green: Lowest correlations with N3.4.
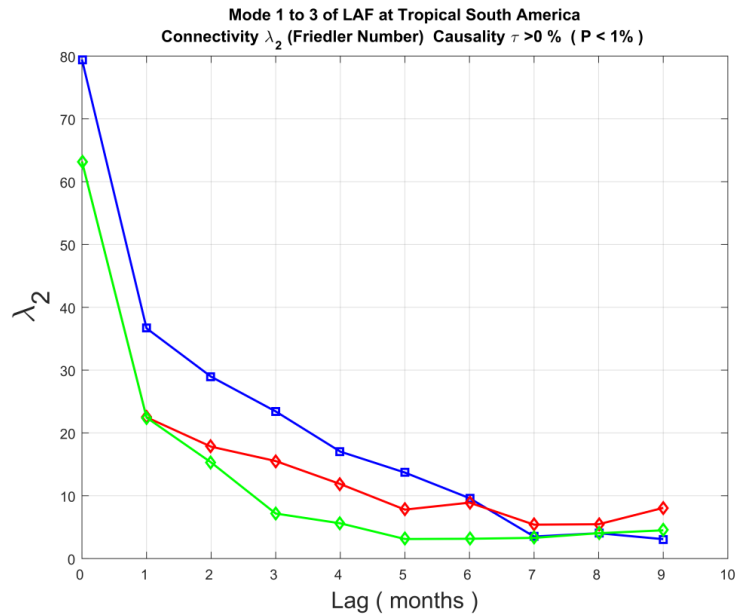
Figure 13B. As in Figure 13A for relative causality $\left|\tau_{y \to x}\right|$.

4. Discussion

Time series extracted from the MCA procedure were used to quantify linear (correlations) and nonlinear (causalities) metrics for different time lags among all pairs of variables. All sets of associations were summarized in the form of correlation and causality graphs in turn grouped according to their association degree with ENSO. This way, we analyzed how the LAFs over TropSA contribute to explain the simultaneous and lagged interannual anomalies over land and atmosphere. In particular, we found a strong statistical association among state and process variables with ENSO over the study region.

The identified structure of correlations among variables differs from that derived from causalities. The temporal dynamics of the identified structural relations was extracted from the SVD analysis of the adjacency matrix ($W_{[ix,j]}$) of the constructed graphs (eqn. A14, section 2.4.1). In particular, we based our analysis of the maximum values contained in the first singular vector, u_1 (best emissary) and its complement, v_1 (best receptor) extracted from $W_{[ix,j]}$. Given that both vectors come from the adjacency matrix, they describe the highest amount of the variance of all possible interactions among variables.

A synthesis of results from correlation graphs allowed us to conclude that:

(1) Regarding the group of graphs exhibiting the highest association with ENSO, a temporal sequence is evidenced whereby PRC (max. u_1) structurally leads EVP (max. v_1) at 5 to 6 month-lags. Also, T_{2m} leads EVP at 4 month-lag, and T_{2m} leads SH^{925} at 3 months-lag. VSW leads EVP at 2 months-lag; and, finally, VSW lead SH^{925} at 1 month-lag.

(2) Regarding the group of graphs exhibiting an intermediate association with ENSO, there is a consistent temporal lead of VSW at all lags (max. u_1). Such lead is evidenced on EVP for 3 to 6 months-lags, on T_{2m} at 2 months-lag, and leading PRC with one month lag. This latter conclusion is at the root of the important role of precipitation recycling in the Amazon River basin [35].

(3) Regarding the group of graphs exhibiting the lowest association with ENSO, VSW persistently leads the relational structure on SH^{925} from 1 to 6 months-lags.

On the other hand, a synthesis of results from the causality graphs allowed us to conclude the following temporal sequence of simultaneous and lagged relations among variables:

(1) Regarding the causality graphs exhibiting the highest association with ENSO, and similar to the correlation graphs, PRC dominates EVP at 5 to 6 month-lags, and T_{2m} at 4 months-lag. In turn, T_{2m} controls EVP at 3 months-lag, and also controls VSW at 1 to 2 months-lag. Besides, the simultaneous relations evidence that VSW dominates SH^{925}. Indeed, non-linear associations also reveals that T_{2m} is essential strengthening LAFs when variables interacts on highest ENSO frequency.

(2) Regarding the causality graphs exhibiting an intermediate association with ENSO, VSW leads EVP for lags between 1 to 6 months. Simultaneous relations among variables indicate that SH^{925} dominates on T_{2m}. From 6 to 3 months-lags of this group of graphs, the linear associations coincide with the nonlinear ones structuring the relations through VSW leading EVP.

(3) Regarding the causality graphs exhibiting the lowest association with ENSO, VSW dominates SH^{925} from 1 to 6 month-lags, similar to the correlation graphs. Finally, SH^{925} dominates over PRC on the simultaneous relations. Unlike the group of the MCS showing the highest association with ENSO, in this case results of linear and nonlinear metrics are quite consistent in the sense of the structural relation between VSW- SH^{925} (Figure 12, group 3).

On the other hand, using the ratio of linear (correlations) and nonlinear (causalities) metrics between two variables ($\boldsymbol{\Phi}$ eqn. 14), we defined a bi-directional relationship (feedback) whenever $0.9 \leq \boldsymbol{\Phi} \leq 1.1$. Using such definition, based on concepts defined by [48], we found that the causality graphs exhibiting the highest association with ENSO show much more frequent feedbacks among the studied variables for all lags between 1 to 6 months. Specifically, the linear and non-linear metrics coincide in identifying the following set of feedbacks:

(1) Regarding the graphs exhibiting the highest association with ENSO, both correlation and causality metrics reveal a consistent bi-directional relation between VSW- T_{2m} at 1-2 months-lags, and between EVP-PRC at 3 months-lag.

(2) Regarding the graphs exhibiting an intermediate association with ENSO, we found consistent linear and non-linear feedbacks between the SH^{925}-PRC pair at 1 month-lag.

(3) Regarding the graphs exhibiting the lowest association with ENSO, consistent results are obtained between linear and nonlinear feedbacks in the relation VSW- T_{2m} at 3 month-lags. For this graphs group, there is a much smaller set of feedbacks among variables in comparison with those graphs high ENSO-related.

Among the studied variables, T_{2m} and VSW exhibit the highest amount of linear feedbacks (correlation) including the one among them at 1 month-lag, and with process variables PRC and EVP. In turn, T_{2m} feedbacks on VSW, EVP and SH^{925} at 2 months-lag, and SH^{925} and VSW feedback on each other at 3 month-lags, and each one establishes an additional feedback (SH^{925}- T_{2m}, VSW-EVP). At 4 months-lag, EVP feedbacks with VSW, T_{2m} and SH^{925}, and for 5 month-lags, VSW feedbacks on EVP and PRC. Regarding the recycling of PRC in TropSA, the VSW-EVP pair (strongly linked to T_{2m}), exhibit the highest amount of linear feedbacks from 3 to 6 months-lags (Figure 12, Group 1, correlation graphs).

For both the simultaneous and lagged analysis, surface temperature (T_{2m}), as a state variable, activates non-linear associations with atmospheric moisture (SH^{925}) and soil moisture (VSW) among the high ENSO-related graphs (causality graphs). In summary, under the ENSO influence, T_{2m} is not only a key variable to diagnose the dynamics of interannual LAFs but also has a substantial role on the dynamics and thermodynamics of the lower troposphere and soil interfaces over TropSA [2, 3,44,73,80]. Besides, interannual anomalies in PRC are associated with perturbations of the surface water balance at regional scale by controlling T_{2m} inside the dissected interactions of VSW and EVP [10,17,30,31,45,48,81-83]. On the contrary, in the group of least ENSO-related graphs that reveals a direct relation between VSW- SH^{925}, T_{2m} is not fundamental in the relational structure among variables.

5. Conclusions

A multivariate analysis allowed us to further understand the role of land surface-atmosphere feedbacks (LAF) over Tropical South America (TropSA) at interannual timescales during the period spanning from 1979 to 2010. Methodologically, we used ideas and tools from Linear Algebra, Information Theory and Graph Theory to infer structural relations among the most relevant processes involved in LAFs over TropSA. In so doing, we defined the dominant LAF spatiotemporal patterns by a pair-wise categorization of variables through Maximum Covariance Analysis (MCA)/Singular-Value Decomposition (SVD), with the aim of reducing the dimensionality of the problem, as well as to identify and quantify the most salient factors associated with ENSO over TropSA. With such patterns, we evaluated the relational structure between state (T_{2m}, SH^{925}, and VSW) and process (PRC and EVP) variables using Graph Theory (Appendix B) to estimate the linear and nonlinear connectivity between the set of variables at interannual timescales. Regarding hydrology in the continental zone of South America, we used the lines of biggest watersheds that also indicate the highest heights of this domain. Among the five macro-basins analyzed, the Amazon basin has the most extensive drainage surface, followed respectively in size by the Tocantins, Orinoco, Magdalena-Cauca, and Essequibo-Guyanas basins (Figure 2). These limits are important in the context of local recycling of moisture in the Andean-Orinoco and Andean-Amazonian piedmont [41,79], and also to understand the resulting MCA spatial patterns.

The identified LAFs over TropSA are enhanced (in-phase) by SSTs over the tropical Pacific and Atlantic, but also over the Indian Ocean. In turn, the algebraic connectivity of graphs [78] representing the LAFs over TropSA is stronger owing to the action of ENSO in comparison with other interannual modes. The regulatory action of ENSO sets in a stronger memory and resilience of interactions among state and process variables. The spectral analysis of the composite graphs identified to study the linkages between variables allowed us to conclude that soil moisture is a key variable in defining the spatiotemporal patterns of PRC and EVP in TropSA at interannual timescales, and as such plays a major role in regulating LAFs at interannual timescales, mainly controlled by ENSO.

Finally, within the context of the humid tropics, the identified LAFs play a major role in controlling the connection between the soil and atmosphere subsystems. Our analyses and results using the concepts of process and state variables and their linear and nonlinear connectivities over TropSA shed new light on the fundamental role of soil moisture within the LAFs and the water and energy budgets over the region at interannual timescales.

Author Contributions: M. Bedoya and G. Poveda conceived and designed the experiments; M Bedoya performed the experiments and analyzed the data. M Bedoya and G Poveda wrote the paper; G Poveda revised the paper.

Conflicts of Interest: Authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

LAFs: Land Atmosphere Feedbacks
TropSA: Tropical South America
SAM: South American Monsoon
TNA: Tropical North America
SST: Sea Surface Temperature
T_{2m}: Surface Temperature
SH^{925}: Specific Humidity at 925 hPa
VSW: Volumetric Soil Water
PRC: Precipitation
EVP: Evaporation
GPCC: Global Precipitation Centre
SVD: Singular Value Decomposition
MCS: Maximum Covariance State
C_{XY}: Covariance Matrix of Variables X and Y

Appendix A Correlation and Causality

By applying the MCA on the truncated matrices $\hat{X}$ and $\hat{Y}$ we obtain the series x_k and y_k (eqn. A1) which capture the covariance between two variables within a determined MCS_k (Figure 3). These series derived from each MCA are used to estimate the correlation, ρ_i, as:

$$\rho(x,y) = \frac{1}{N-1} \sum_{i=1}^{N} \left(\overline{\frac{x_i - \mu_x}{\sigma_x}} \right) \left(\overline{\frac{y_i - \mu_y}{\sigma_y}} \right) \tag{A1}$$

For our purposes, we performed the analysis by lagging both series from 1 to 9 months, and defined a measure of feedback through the ratio between correlations for each lag of each covariance matrix between pairs of variables.

Nevertheless, correlation analyses just quantify the degree of linear association between variables. We overcame such limitation by using causality analyses to quantify the connectivity between variables. Such causality or information transfer has been defined within the framework of information theory [69,84-86]. Three basic tenets of the information transfer are the following: (1) Causality implies correlation but correlation does not imply causality, (2) Causality implies directionality, which means that the transfer of information detects the direction of information transfer between two systems, and (3) Asymmetry is a basic property of information transfer.

We consider a feedback as a bi-directional causality between two variables. To estimate the information flow, [69] proposes a system of two stochastic differential equations:

$$\frac{\partial x_k}{\partial t} = F_x(x,y,t) + b_{xx}\omega_x + b_{xy}\omega_y \tag{A2}$$

$$\frac{\partial y_k}{\partial t} = F_y(x,y,t) + b_{yx}\omega_x + b_{xy}\omega_y \tag{A3}$$

where ω_x and ω_y represent white noise; while b_{xy} and $F_{x,y}$ are arbitrary functions of x, y and t. [69] proved that it is possible to measure the causality between two series x and y through the absolute rate of information flow $(T_{Y \to X})$ in terms of the Shannon entropy, such that:

$$T_{Y \to X} = -E\left(\frac{1}{\delta_x} \frac{\partial(F_x \delta_x)}{\partial x} \right) + \frac{1}{2}E\left(\frac{1}{\delta_x} \frac{\partial^2 (b_{xx}^2 + b_{xy}^2)\delta_x}{\partial x} \right) \tag{A4}$$

where δ_x is the marginal density of x, and E represents the expected value [69]. By applying eqn. (A4) to two series x and y, the maximum likelihood estimator of eqn. (A4) is given by,

$$T_{Y \to X} = \frac{C_{xx}C_{xy}C_{y,dx} - C_{xy}^2 C_{x,dy}}{C_{xx}^2 C_{yy} - C_{xx}C_{xy}^2} \tag{A5}$$

where C_{xx}, C_{xy} y C_{yy} denote the possible covariance between series. To estimate $C_{x,dy}$ and $C_{y,dx}$ it is necessary to estimate the covariance between pairs of series (x, $\dot{y}_n$) and (y, $\dot{x}_n$), where $\dot{x}_n$ and $\dot{y}_n$ are the discrete approximations of $\frac{dx}{dt}$ and $\frac{dy}{dt}$, through the Eulerian formulation:

$$\dot{x}_n = \frac{x_{n+\kappa} - x_n}{\kappa \Delta t} \tag{A6}$$

with $\kappa = 1$. In some cases, such as deterministic chaotic processes, $\kappa = 2$.

The rate of information flow ($T_{Y \to X}$) can be zero or different than zero. If $T_{Y \to X} = 0$ then y does not cause x. There are two possible ways to interpret the information flow depending on the sign: a positive sign of $T_{Y \to X}$ implies that y acts upon x by augmenting its uncertainty, while a negative sign indicates that y stabilizes x by diminishing its entropy. This type of interpretation is rather confusing because it goes against the ordinary algebraic interpretation. Diverse similar metrics do exist within the context of information theory, such as the Interaction Information, whose positive (negative) values determine redundancy (synergy) in the interaction between two series [54,55].

Derivation of eqn. (A4) uses the concept of Shannon or Absolute Entropy [59]. For two-variable systems, [85] proved that the flow of information, $T_{y \to x}$, is the same for both absolute and relative entropy. The latter is more suitable to study predictability, owing to its invariance properties under non-linear interactions [54,56,84].

A measure of relative flow of information [86], in terms of the marginal entropy, can be given as,

$$Z_{y \to x} = \left| T_{y \to x} \right| + \left| \frac{dH_y^*}{dt} \right| + \left| \frac{dH_y^{noise}}{dt} \right| \tag{A7}$$

The maximum likelihood estimators suitable to solve eqn.(A7), are:

$$p = \frac{C_{yy}C_{x,dx} - C_{xy}C_{y,dx}}{\det(C)} \tag{A8}$$

$$q = \frac{C_{yy}C_{x,dx} - C_{xy}C_{y,dx}}{\det(C)} \tag{A9}$$

Additionally, the two components of entropy (eqn. A7), estimated in terms of p and q are [86]

$$\left| \frac{dH_y^*}{dt} \right| = p \tag{A10}$$

$$\left| \frac{dH_y^{Noise}}{dt} \right| = \frac{\Delta t}{2C_{xx}}(C_{dx,dx} + p^2 C_{xx} + q^2 C_{yy} - 2pC_{dx,x} - 2qC_{dx,y} + 2pqC_{xy}) \tag{A11}$$

Finally, the Relative Flow of Information ($\tau_{y \to x}$) is defined as:

$$\tau_{y \to x} = \frac{\left| Z_{y \to x} \right|}{T_{y \to x}} \tag{A12}$$

If $\tau_{y \to x} = 100$, the variation of x is 100% due to the flow of information from y; and if $\tau_{x \to y} = 0$, then y is not the cause. Nevertheless, $\tau_{y \to x}$ evalutes the degree influence of y on x relative to other processes.

Appendix B Connectivity of Graphs at Interannual Time Scale

We use graphs to infer structural relations and cycles in the LAFs over TropSA [52]. To that end, graphs must be represented through matrices. There is a mathematical connection between graphs and the algebraic properties of such matrices [53,55,78,87]. From this point of view, a graph, Γ_k, has two important matrices associated: the adjacency matrix (W) and the Laplacian matrix (L).

The adjacency matrix, $W_{[i \times j]}$ (eqn A13), is a square matrix whose dimensions, rows (i) and columns (j), represent the nodes of a graph. A null entry in this matrix indicates that there is no edge between nodes i and j, and thus:

$$W_{ij} = \begin{cases} w_{ij} : & \text{weight in the edge } (i \, , \, j) \\ 0 \ : & \text{there is no edge between } i \text{ and } j \end{cases}, \tag{A13}$$

where w_{ij} is the correlation, $\rho(x \, , \, y)$, and/or the relative causality, $\tau_{x \to y}$, between the time series x_k and y_k of each MCS_k. If the graph is unidirectional, the adjacency matrix, $W \in \Re^{i \times j}$ is a symmetric matrix of $|N_k| \times |N_k|$, with $|N_k|$ being the number of variables under analysis.

Figure A1 basically represents a directed graph connecting three nodes x_k, y_k and z_k using linear (ρ) and non-linear (τ) coupling metrics. Particularly, this directed graph is an extension of the two nodes scheme presented in Figure A1 (Right). Each directed graph associates its respective adjacency matrix $W_{[i \times j]}^{\rho}$ (correlations) and $W_{[i \times j]}^{\tau}$ (causalities). In each adjacency matrix, we head rows and columns with the variable associated to each node (X, Y, Z), and thus we emphasize how this matrix establishes all possible connections between these three nodes.

For example, for the case of the correlation-based graph (Figure A1-Left), the edge $\rho(x_k, y_k)$ connects the nodes in the X-to-Y direction, while the edge $\rho(y_k, x_k)$ connects the nodes in the Y-to-X direction. Likewise, in the case of non-linear coupling graphs, the edge $\tau_{x \to y}$ connects the nodes in the X-to-Y direction, while the edge $\tau_{y \to x}$ connects the nodes in the *Y*-to-*X* direction (Figure A1-Right). If this method is applied to the interactions between the *X* and *Z* nodes and between the Y and Z nodes of this graph, we find 6 possible interactions between the three nodes. The diagonal of the adjacency matrix is null because we are not analyzing auto-correlations or auto-causalities.

To study graph properties [88], we initially use the spectral graph theory by applying a Singular Value Decomposition on the adjacency matrix $W_{[i \times j]}$ (eqn. A14):

$$W_{[i \times j]} = U_{[i \times i]} \Sigma_{[i \times j]} V_{[j \times j]}^T = \sum_{h=1}^{i} u_h \sigma_h v_h^T = u_1 \sigma_1 v_1^T + u_2 \sigma_2 v_2^T + ... + u_i \sigma_i v_j^T \tag{A14}$$

This SVD of $W_{[i \times j]}$ provides important information about the structural relations and cycles among variables of the LAF graphs. As the structure of $W_{[i \times j]}$ is designed as to evaluate all the possible interactions among variables, the singular values of eqn. (A14) represent the amount of variance explained by each interaction mode. Also, the largest entries of the main singular vectors u_1 and v_1 of eqn. A14 define the most important variables of each mode of interaction [88].

$$W^{\rho}_{[ixj]} = \begin{matrix} & X & Y & Z \\ X \\ Y \\ Z \end{matrix} \begin{pmatrix} 0 & \rho_t(x_k,y_k) & \rho_t(x_k,z_k) \\ \rho_t(y_k,x_k) & 0 & \rho_t(y_k,z_k) \\ \rho_t(z_k,x_k) & \rho_t(z_k,y_k) & 0 \end{pmatrix}$$

$$W^{\tau}_{[ixj]} = \begin{matrix} & X & Y & Z \\ X \\ Y \\ Z \end{matrix} \begin{pmatrix} 0 & \tau_{x\to y} & \tau_{x\to z} \\ \tau_{y\to x} & 0 & \tau_{y\to z} \\ \tau_{z\to x} & \tau_{z\to y} & 0 \end{pmatrix}$$
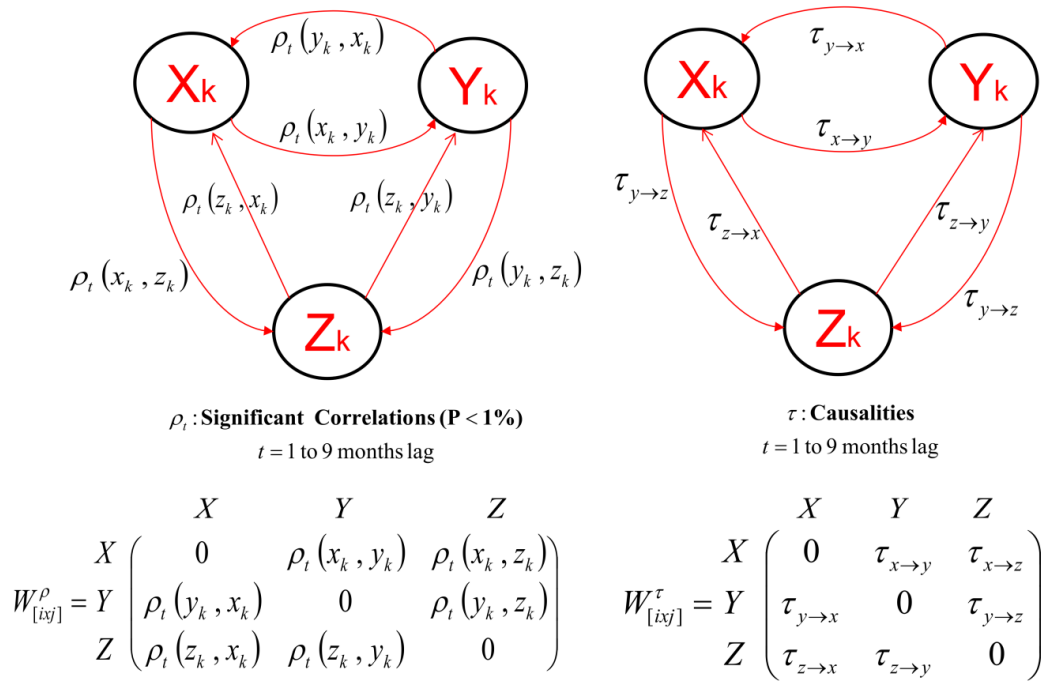
Figure A1. Graphs among three variables (X, Y, Z) with links denoting lagged correlations (top left) and causalities (top right) between variables/nodes. Bottom panels include the structure of the graph's adjacency matrix, W, to illustrate the node-to-node connection that is established in each graph.

Furthermore, a diagonal matrix of degree $D \in \Re^{ixj}$ is defined, which quantifies the number of edges per variable; that is the number and cumulative weight of the connections of each node, as

$$D_i = \sum_{\{j|(i,j)\in N_k\}} w_{ij} \tag{A15}$$

From eqns. A14 and A15 the Laplacian or Kirchhoff matrix of the graph, $L \in \Re^{ixj}$, is estimated as the difference between the degree matrix and the adjacency matrix, $W_{[ixj]}$. This implies that the sum of all columns of the Laplacian matrix, $L(\Gamma_k)$, adds to zero, so that:

$$L = D - W \tag{A16}$$

$$L_{ij} = \begin{cases} d, & \text{if } i = j \\ -w_{ij}, & \text{if } (i,j) \\ 0, & \text{if there is no edge between } i, j \end{cases} \tag{A17}$$

In this case, we use the spectral graph theory by using eigen-decomposition on the Laplacian matrix to estimate the eigen-values (λ). These eigen–values contain measures of the connectivity of each graph Γ_k constructed to assess LAFs over TropSA.

Given a graph, Γ_k, and its Laplacian matrix, $L(\Gamma_k)$, the eigen-values and eigen-vectors are estimated from

$$L = V\Lambda V^{-1}, \tag{A18}$$

where $\Lambda = \{ \lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_i \}$ is a diagonal matrix containing the spectrum of eigen-values of the Laplacian matrix, and the matrix $V = \{ \boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_i \}$ contains the eigen-vectors.

The second non-null eigenvalue (λ_2) of the Laplacian matrix is the algebraic connectivity of the associated graph. This metric was introduced by [78], also known as the Friedler Value. The larger the value of λ_2 the larger is the connectivity of Γk. Besides, smaller values of λ_2 indicates that Γk increases its modularity.

Also, the eigen-vector associated with the algebraic connectivity (λ_2) is known as the Friedler vector (v2). The sign of this eigen-vector can establish a thick partition in two groups of the connected variables of the graph, Γk.

Likewise, [87] indicate the following important characteristics of the Laplacian matrix L(Γk):

1. L(Γk) is symmetrical,
2. The set of eigen-values and eigen-vectors of L(Γk) are real and non-negative,
3. Because of the special structure of the matrix L(Γk), where all entries in each column add up to zero, the smallest eigen-value of this matrix is always zero $\lambda_0 = 0$ [87].

4. The graph $\boldsymbol{\Gamma}_k$ has h connected components if and only if $\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{i-1} = 0$ [78]. This means that the multiplicity of the null eigen-values ($\lambda = 0$) is the number of components in which the graph $\boldsymbol{\Gamma}_k$ can be decomposed. The number of null eigen-values corresponds to the number of groups in which the graph can be decomposed.

References

1. Wohl, E., Barros, A., Brunsell, N., Chappell, N. A., Coe, M., Giambelluca, T., ... & McDonnell, J. The hydrology of the humid tropics. Nature Climate Change **2012**, 2(9), 655-662.

2. Trenberth, K. E., 1999: Atmospheric moisture recycling: Role of advection and local evaporation. J. Climate, 12, 1368-1381.

3. Trenberth, K. E., and D. J. Shea, 2005: Relationships between precipitation and surface temperature. *Geophys. Res. Lett.*, **32**, L14703, doi:10.1029/2005GL022760.

4. Taylor, C. M., de Jeu, R. A., Guichard, F., Harris, P. P., & Dorigo, W. A. (2012). Afternoon rain more likely over drier soils. *Nature*, *489*(7416), 423-426.

5. Seneviratne, S. I., & Stöckli, R. (2008). The role of land-atmosphere interactions for climate variability in Europe. In *Climate Variability and Extremes during the Past 100 years* (pp. 179-193). Springer Netherlands.

6. Kolstad, EW, Barnes, EA, & ... (2017). Quantifying the Role of Land–Atmosphere Feedbacks in Mediating Near-Surface Temperature Persistence. *Quarterly Journal of the ….* Wiley Online Library.

7. Betts, A. K., Ball, J. H., Beljaars, A., Miller, M. J., & Viterbo, P. A. (1996). The land surface-atmosphere interaction: A review based on observational and global modeling perspectives. *Journal of Geophysical Research: Atmospheres*, *101*(D3), 7209-7225.

8. Schär, C., D. Lüthi, U. Beyerle, and E. Heise, 1999: The soil-precipitation feedback: A process study with a regional climate model. *J. Climate*, **12**, 722-741.

9. Xue, Y., De Sales, F., Li, W. P., Mechoso, C. R., Nobre, C. A., & Juang, H. M. (2006). Role of land surface processes in South American monsoon development. *Journal of climate*, *19*(5), 741-762.

10. Berg, A., and Coauthors, 2015: Interannual Coupling between Summertime Surface Temperature and Precipitation over Land: Processes and Implications for Climate Change. *J. Climate*, **28**, 1308–1328.

11. Knist, S., Goergen, K., Buonomo, E., Christensen, O. B., Colette, A., Cardoso, R. M., ... & Kartsios, S. (2017). Land-atmosphere coupling in EURO-CORDEX evaluation experiments. *Journal of Geophysical Research: Atmospheres*.

12. Baisya, H., Pattnaik, S., & Rajesh, P. V. (2017). Land surface-precipitation feedback analysis for a landfalling monsoon depression in the Indian region. *Journal of Advances in Modeling Earth Systems*.

13. Zemp, D. C., Schleussner, C. F., Barbosa, H. M., Hirota, M., Montade, V., Sampaio, G., ... & Rammig, A. (2017). Self-amplified Amazon forest loss due to vegetation-atmosphere feedbacks. *Nature Communications*, *8*, 14681.

14. Koster, R. D., and M. J. Suarez, 1999: A simple framework for examining the interannual variability of land surface moisture fluxes. *J. Climate*, **12**, 1911-1917.

15. Koster, R. D., P. A. Dirmeyer , Z. Guo, G. Bonan, E. Chan, P. Cox, and T. Yamada, 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138-1140.

16. Orlowsky, B., & Seneviratne, S. I. (2010). Statistical analyses of land–atmosphere feedbacks and their possible pitfalls. *Journal of Climate*, *23*(14), 3918-3932.

17. Catalano, F., Alessandri, A., De Felice, M., Zhu, Z., & Myneni, R. B. (2015). Observationally based analysis of land–atmosphere coupling. *Earth System Dynamics Discussions*, *6*(2).

18. Wei, J., & Dirmeyer, P. A. (2012). Dissecting soil moisture-precipitation coupling. *Geophysical Research Letters*, *39*(19).

19. Gimeno, L., and Coauthors, 2012: Oceanic and terrestrial sources of continental precipitation. *Rev. Geophys.*, **50**, 4003, doi: 10.1029/2012RG000389

20. Roy, S. B. (2009). Mesoscale vegetation-atmosphere feedbacks in Amazonia. *Journal of Geophysical Research: Atmospheres*, *114*(D20).

21. Konings, A. G., Dekker, S. C., Rietkerk, M., & Katul, G. G. (2011). Drought sensitivity of patterned vegetation determined by rainfall-land surface feedbacks. *Journal of Geophysical Research: Biogeosciences*, *116*(G4).

22. Gentine, P., Entekhabi, D., & Polcher, J. (2011). The diurnal behavior of evaporative fraction in the soil–vegetation–atmospheric boundary layer continuum. *Journal of Hydrometeorology*, *12*(6), 1530-1546.

23. Bagley, J. E., A. R. Desai, K. J. Harding, P. K. Snyder, and J. A. Foley, 2014: Drought and Deforestation: Has land cover change influenced recent precipitation extremes in the Amazon? *J. Climate*, **27**, 345-361, doi:10.1175/JCLI-D-12-00369.1

24. Boers, N., Marwan, N., Barbosa, H. M., & Kurths, J. (2017). A deforestation-induced tipping point for the South American monsoon system. *Scientific Reports*, *7*.

25. Yin, L., R. Fu, Y.-F. Zhang, P. A. Arias, D. N. Fernando, W. Li, K. Fernandes, and A. R. Bowerman, 2014: What controls the interannual variation of the wet season onsets over the Amazon?, *J. Geophys. Res. Atmos.*, **119**, 2314–2328, doi:10.1002/2013JD021349.

26. Gentine, P., Holtslag, A. A., D'Andrea, F., & Ek, M. (2013). Surface and atmospheric controls on the onset of moist convection over land. *Journal of Hydrometeorology*, *14*(5), 1443-1462.

27. Taylor, C. M. (2015). Detecting soil moisture impacts on convective initiation in Europe. *Geophysical Research Letters*, *42*(11), 4631-4638.

28. Salati, E., Dall'Olio, A., Matsui, E., & Gat, J. R. (1979). Recycling of water in the Amazon basin: an isotopic study. *Water resources research*, 15(5), 1250-1258.

29. Lettau, H., Lettau, K., & Molion, L. C. B. (1979). Amazonia's hydrologic cycle and the role of atmospheric recycling in assessing deforestation effects. *Monthly Weather Review*, 107(3), 227-238.

30. Eltahir, E.A.B., Bras, R.L. Precipitation Recycling. *Reviews of Geophysics*, v.34, n.3, p.367-379, 1996.

31. van der Ent, R. J., H. H. Savenije, B. Schaefli, and S. C. Steele-Dunne, 2010: Origin and fate of atmospheric moisture over continents. *Water Resour. Res.*, **46,** doi: 10.1029/2010WR009127

32. Medvigy, D., Walko R., and Roni Avissar. (2011) Effects of Deforestation on Spatio-temporal Distributions of Precipitation in South America. *Journal of Climate* 24:8, 2147-2163.

33. Zemp, DC, C.-F. Schleussner, H. M. J. Barbosa, R. J. van der Ent, J. F. Donges, J. Heinke, G. Sampaio, and A. Rammig (2014), On the importance of cascading moisture recycling in South America. *Atmos. Chem. Phys.*, 14, 13337–13359,

34. Zemp, D. C., Schleussner, C. F., Barbosa, H. M., Hirota, M., Montade, V., Sampaio, G., ... & Rammig, A. (2017). Self-amplified Amazon forest loss due to vegetation-atmosphere feedbacks. *Nature Communications*, *8*, 14681.

35. Rocha et al., 2015; Rocha, V.M., Correia, F.W.S., and Fonseca, P.A.M., (2015), Reciclagem de precipitação na Amazônia: Um estudo de revisão. *Rev. Bras. Meteorol.*, v. 30, n. 1, 59-70.

36. Sörensson, A. A., & Menéndez, C. G. (2011). Summer soil–precipitation coupling in South America. *Tellus A*, *63*(1), 56-68.

37. Barreiro, M., Díaz, N., & Renom, M. (2014). Role of the global oceans and land–atmosphere interaction on summertime interdecadal variability over northern Argentina. *Climate dynamics*, *42*(7-8), 1733-1753.

38. Ruscica, R. C., Sörensson, A. A., & Menéndez, C. G. (2014). Hydrological links in Southeastern South America: soil moisture memory and coupling within a hot spot. *International Journal of Climatology*, *34*(14), 3641-3653.

39. Ruscica, R. C., Sörensson, A. A., & Menéndez, C. G. (2015). Pathways between soil moisture and precipitation in southeastern South America. *Atmospheric Science Letters*, *16*(3), 267-272.

40. Wang, H., & Fu, R. (2004). Influence of cross-Andes flow on the South American low-level jet. *Journal of climate*, *17*(6), 1247-1262.

41. Boers, N., Bookhagen, B., Marwan, N., & Kurths, J. (2016). Spatiotemporal characteristics and synchronization of extreme rainfall in South America with focus on the Andes Mountain range. *Climate Dynamics*, *46*(1-2), 601-617.

42. Collini, E. A., Berbery, E. H., Barros, V. R., & Pyle, M. E. (2008). How does soil moisture influence the early stages of the South American monsoon?. *Journal of Climate*, *21*(2), 195-213.

43. Ma, H. Y., Mechoso, C. R., Xue, Y., Xiao, H., Wu, C. M., Li, J. L., & De Sales, F. (2011). Impact of land surface processes on the South American warm season climate. *Climate dynamics*, *37*(1-2), 187-203.

44. Poveda, G., and O. J. Mesa, 1997: Feedbacks between hydrological processes in tropical South America and large scale oceanic atmospheric phenomena, *J. Climate*, **10**, 2690-2702.

45. Barreiro, M. and N. Diaz, 2011: Land-atmosphere coupling in El Niño influence over South America. *Atm. Sc. Lett.*, **12**, 351-355, doi: 10.1002/asl.348

46. Builes-Jaramillo, A., Marwan, N., Poveda, G., & Kurths, J. (2017). Nonlinear interactions between the Amazon River basin and the Tropical North Atlantic at interannual timescales. *Climate Dynamics*, 1-19.

47. Brubaker, K. L., & Entekhabi, D. (1996). Analysis of feedback mechanisms in land-atmosphere interaction. *Water Resources Research*, *32*(5), 1343-1357.

48. Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., ... & Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3), 125-161.

49. Gentine, P., Entekhabi, D., & Polcher, J. (2011). The diurnal behavior of evaporative fraction in the soil–vegetation–atmospheric boundary layer continuum. *Journal of Hydrometeorology*, *12*(6), 1530-1546.

50. Seneviratne, S. I., Lüthi, D., Litschi, M., & Schär, C. (2006). Land–atmosphere coupling and climate change in Europe. *Nature*, *443*(7108), 205-209.

51. Haghighi, E. and Kirchner, J. W. (2017), Near-surface turbulence as a missing link in modeling evapotranspiration-soil moisture relationships. *Water Resour. Res.* doi:10.1002/2016WR020111

52. Runge, J., Petoukhov, V., & Kurths, J. (2014). Quantifying the strength and delay of climatic interactions: the ambiguities of cross correlation and a novel measure based on graphical models. *Journal of Climate*, *27*(2), 720-739.

53. Donges, J. F., Petrova, I., Loew, A., Marwan, N., & Kurths, J. (2015). How complex climate networks complement eigen-techniques for the statistical analysis of climatological data. *Climate Dynamics*, *45*(9-10), 2407-2424.

54. Runge, J. (2015). Quantifying information transfer and mediation along causal pathways in complex systems. *Physical Review E, 92*(6), 062829.

55. Kretschmer, M., Coumou, D., Donges, J. F., & Runge, J. (2016). Using causal effect networks to analyze different Arctic drivers of midlatitude winter circulation. *Journal of Climate*, *29*(11), 4069-4081.

56. Shannon, Claude E. (October 1948). "A Mathematical Theory of Communication". Bell System Technical Journal 27 (4): 623–656. doi:10.1002/j.1538-7305.1948.tb00917.x.

57. Parrondo, J. M., Horowitz, J. M., & Sagawa, T. (2015). Thermodynamics of information. *Nature physics*, *11*(2), 131-139.

58. Horowitz, J. M., & Esposito, M. (2014). Thermodynamics with continuous information flow. *Physical Review X*, 4(3), 031015.

59. Espinoza Villar, J. C., Ronchail, J., Guyot, J. L., Cochonneau, G., Naziano, F., Lavado, W., ... & Vauchel, P. (2009). Spatio-temporal rainfall variability in the Amazon basin countries (Brazil, Peru, Bolivia, Colombia, and Ecuador). *International Journal of Climatology*, 29(11), 1574-1594.

60. Schneider, U., A. Becker, F. Finger, A. Meyer-Christoffer, M. Ziese, and B.Rudolf, 2013: GPCC's new land-surface precipitation climatology based on quality-controlled in-situ data and its role in quantifying the global water cycle, *Theor. Appl. Climatol.*, doi 10.1007/s00704-013-0860-x.

61. Dee, D.P. et al. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J.R. Meteorol. Soc.* 137, 553-597. doi: 10.1002/qj.828

62. Albergel, and Coauthors, 2013: Skill and Global Trend Analysis of Soil Moisture from Reanalyses and Microwave Remote Sensing. *J. Hydrometeor*, **14**, 1259–1277, doi: 10.1175/JHM-D-12-0161.1

63. Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., ... & De Rosnay, P. (2015). ERA-Interim/Land: a global land surface reanalysis data set. *Hydrology and Earth System Sciences*, *19*(1), 389-407

64. Nicolai-Shaw, N., Gudmundsson, L., Hirschi, M., & Seneviratne, S. I. (2016). Long-term predictability of soil moisture dynamics at the global scale: Persistence versus large-scale drivers. *Geophysical Research Letters*, *43*(16), 8554-8562.

65. Mayer, M., K. E. Trenberth, L. Haimberger, and J. T. Fasullo 2013: The Response of Tropical Atmospheric Energy Budgets to ENSO. *J. Climate*, **26**, 4710-4724.

66. Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3), 211-218.

67. Bjornsson, H., and S. A. Venegas, 1997: A manual for EOF and SVD analyses of climatic data. *CCGCR Report*, *97*(1).

68. von Storch, H., & Zwiers, F. W. (2001). *Statistical analysis in climate research*. Cambridge university press.

69. Liang, X. S. (2014). Unraveling the cause-effect relation between time series. *Physical Review E*, *90*(5), 052150.

70. Poveda, G., Jaramillo, A., Gil, M. M., Quiceno, N., & Mantilla, R. I. (2001). Seasonally in ENSO-related precipitation, river discharges, soil moisture, and vegetation index in Colombia. *Water resources research*, 37(8), 2169-2178.

71. Poveda, G., D. M. Alvarez, and O. A. Rueda, 2011: Hydroclimatic variability over the Andes of Colombia associated with ENSO: A review of climatic processes and their impact on one of the Earth´s most important biodiversity hotspots, *Clim. Dyn.*, **36**, 2233-2249.

72. Poveda, G., L. Jaramillo, and L. F. Vallejo, 2014: Seasonal precipitation patterns along pathways of South American low-level jets and aerial rivers. *Water Resour. Res.*,**50**, *doi: 10.1002/2013WR014087*

73. Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., & Dolman, A. J. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, *15*(2), 453.

74. Trenberth, K. E., Dai, A., Rasmussen, R. M., & Parsons, D. B. (2003). The changing character of precipitation. *Bulletin of the American Meteorological Society*, 84(9).

75. Marengo, J. A., Nobre, C. A., Tomasella, J., Oyama, M. D., Sampaio de Oliveira, G., De Oliveira, R., ... & Brown, I. F. (2008). The drought of Amazonia in 2005. *Journal of Climate*, 21(3), 495-516.

76. Marengo, J. A., Tomasella, J., Alves, L. M., Soares, W. R., & Rodriguez, D. A. (2011). The drought of 2010 in the context of historical droughts in the Amazon region. *Geophysical Research Letters*, *38*(12).

77. Espinoza, J. C., Ronchail, J., Frappart, F., Lavado, W., Santini, W., & Guyot, J. L. (2013). The major floods in the Amazonas river and Tributaries (western Amazon basin) during the 1970–2012 Period: a focus on the 2012 flood*.*Journal of Hydrometeorology*, *14*(3), 1000-1008.

78. Friedler, M. (1973). Algebraic connectivity of graphs. *Czechoslovak mathematical journal*, *23*(2), 298-305.

79. Wang, H., & Fu, R. (2004). Influence of cross-Andes flow on the South American low-level jet. Journal of climate, 17(6), 1247-1262.

80. Mapes, B. E., Warner, T. T., & Xu, M. (2003). Diurnal patterns of rainfall in northwestern South America. Part III: Diurnal gravity waves and nocturnal convection offshore. Monthly Weather Review, 131(5), 830-844.

81. Vera, C., Higgins, W., Amador, J., Ambrizzi, T., Garreaud, R., Gochis, D., ... & Nogues-Paegle, J. (2006). Toward a unified view of the American monsoon systems. *Journal of climate*, 19(20), 4977-5000.

82. Grimm, A. M., & Tedeschi, R. G. (2009). ENSO and extreme rainfall events in South America. *Journal of Climate*, 22(7), 1589-1609.

83. Mo, K. C., & Berbery, E. H. (2011). Drought and persistent wet spells over South America based on observations and the US CLIVAR drought experiments. *Journal of Climate*, 24(6), 1801-1820.

84. Schreiber, T. (2000). Measuring information transfer. *Physical review letters*,*85*(2), 461.

85. Liang, X. S. (2013). The Liang-Kleeman information flow: Theory and applications. *Entropy*, *15*(1), 327-360.

86. Liang, X. S (2015). Normalizing the causality between time series. *Physical Review E*, *92*(2), 022126.

87. Pothen, A., Simon, H. D., & Liou, K. P. (1990). Partitioning sparse matrices with eigenvectors of graphs. *SIAM journal on matrix analysis and applications*, *11*(3), 430-452.

88. Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, *46*(5), 604-632.