

# Inferring long memory processes in the climate network via ordinal pattern analysis

Marcelo Barreiro,<sup>1</sup> Arturo C. Marti,<sup>1</sup> and Cristina Masoller<sup>2</sup>

<sup>1</sup>*Instituto de Física, Facultad de Ciencias, Universidad de la República, Iguá, 4225, Montevideo, Uruguay*

<sup>2</sup>*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Colom 11, E-08222 Terrassa, Barcelona, Spain*

(Received 15 November 2010; accepted 3 January 2011; published online 25 February 2011)

We use ordinal patterns and symbolic analysis to construct global climate networks and uncover long- and short-term memory processes. Data analyzed are the monthly averaged surface air temperature (SAT field), and the results suggest that the time variability of the SAT field is determined by patterns of oscillatory behavior that repeat from time to time, with a periodicity related to intraseasonal oscillations and to El Niño on seasonal-to-interannual time scales. © 2011 American Institute of Physics. [doi:10.1063/1.3545273]

**We analyze climatological data from a complex networks perspective, using techniques of nonlinear time series symbolic analysis. Specifically, we employ ordinal patterns and binary representations to analyze monthly averaged surface air temperature (SAT) anomalies. By computing the mutual information of the time series in regular grid points covering the Earth's surface and then performing global thresholding, we construct climate networks that uncover short-term memory processes, as well as long ones (5–6 yr). Our results suggest that the time variability of the SAT anomalies is determined by patterns of oscillatory behavior that repeat from time to time with a periodicity related to intraseasonal variations and to El Niño on seasonal to interannual time scales. The present work is located at the triple intersection of three highly active interdisciplinary research fields in nonlinear science: symbolic methods for nonlinear time series analysis, network theory, and nonlinear processes in the earth climate. While a lot of effort is being done in order to improve our understanding of natural complex systems, with many different methods for mapping time series to network representations being investigated and employed in complex systems such as the human brain, our work is the first one aimed at characterizing the global climate network in terms of oscillatory patterns that tend to repeat from time to time, with various time scales. By mapping these processes into a global network, using ordinal patterns and binary representations, we find that the structure of the network changes drastically at different time scales.**

## I. INTRODUCTION

Complex networks have been intensively studied in the last years because they represent many real systems such as the Internet; ecological, social, and metabolic networks; genes; cells; and the brain.<sup>1</sup> Global climate modeling is also a hot topic nowadays because of its huge economic and social impact for future generations. Giving the complexity

of the inter-relations between the different elements that constitute our environment, it is important to analyze climatological data from a complex network perspective. However, despite the intensive effort in research done in these two interdisciplinary and fascinating fields, just very few studies have combined both.<sup>2–8</sup> These studies have shown that network theory can yield light into interesting, previously unknown features of our climate.

Tsonis and Swanson<sup>4</sup> and Yamasaki *et al.*<sup>3</sup> have shown that the climate network is significantly affected by El Niño, as during El Niño years many links of the network are broken. Tsonis and Swanson<sup>4</sup> constructed cross-correlation-based networks of the SAT field for El Niño and for La Niña years and investigated their structure. They found that the El Niño network possesses significantly fewer links and lower clustering coefficient and characteristic path length than the La Niña network. They conjectured and verified that, because El Niño network is less communicative and less stable than La Niña one, during El Niño years temperature predictability is lower compared to La Niña years. Using a different approach, Yamasaki *et al.*<sup>3</sup> arrived at a similar conclusion. They developed a method that allows to follow time variations of the network structure by observations of fluctuations in the correlations between nodes. The method allows to distinguish between the two qualitatively different groups of network links, blinking links that appear and disappear in a short time, and robust links that represent long lasting relations between temperature fluctuations in two regions. Assuming that broken links are due to structural changes in the network, by tracking these changes in several zones a strong response to El Niño was revealed, even in geographical regions where the mean temperature is not affected by El Niño.

Donges *et al.*<sup>6</sup> compared the structural properties of networks constructed by using, as a measure of dynamical similarity between regions, linear, and nonlinear measures: the linear Pearson correlation coefficient and the nonlinear mutual information. They analyzed two sets of data: the SAT anomalies obtained from large-scale climate simulations by the coupled atmosphere-ocean general circulation models

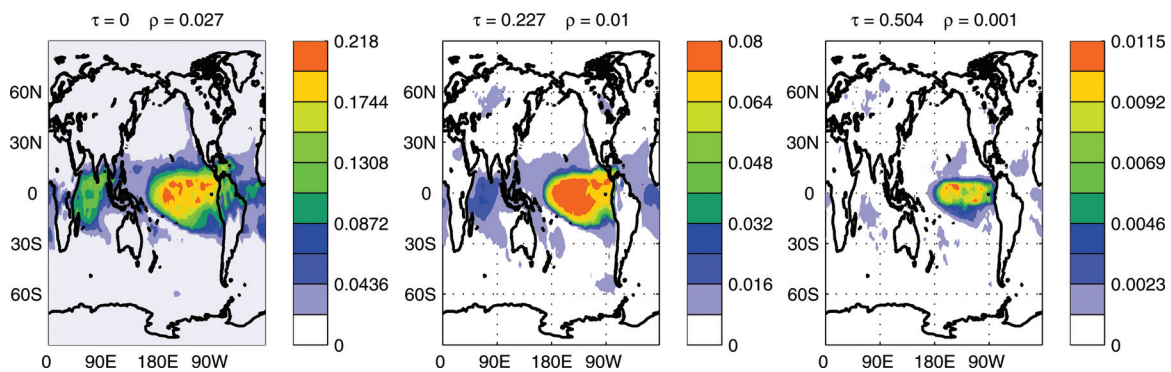


FIG. 1. (Color online) Zero-threshold (left), low-threshold (center), and high-threshold (right) networks constructed by computing the mutual information from ordinal patterns of length  $D = 4$  defined by comparing SAT anomalies in consecutive years. The 2D plots are color-coded such that the white (red) regions indicate the geographical areas with zero (largest) area weighted connectivity. In each panel, the values of the threshold,  $t$ , and of the edge-density,  $\rho$ , are indicated.

and the SAT anomalies reanalysis data sets. A high degree of similarity using the two approaches (linear and nonlinear similarity measures) was found on the local and on the mesoscopic topological scales; however, important differences were uncovered on the global scale, particularly in the betweenness centrality field. In Ref. 7, Donges *et al.* employed the mutual information to reveal wavelike structures of high-energy flow that could be traced back to global surface ocean currents. Their results point to the major role of the oceanic surface circulation in coupling and stabilizing the global temperature field in the long-term mean.

When computing the mutual information, in order to detect patterns and correlations in the variability of two nodes, a critical issue is defining probability distribution functions (PDFs) that fully take into account the temporal order in which the SAT anomalies occur in the time series. Histogram-based PDFs do not take into account this temporal order and, thus, are not optimal for capturing subtle correlated oscillatory patterns. Alternatively, one can use time-delay embedding techniques to represent the time series as a trajectory in a high-dimensional space; however, the information provided by the mutual information is strongly dependent on the embedding technique, the time-delay, and the phase space partition.<sup>9</sup>

An alternative methodology, originally proposed by Bandt and Pompe (BP) (Ref. 10) allows to define probability

distribution functions that fully take into account the time ordering of the SAT anomalies. The BP method is based on comparing values in the time series to construct “ordinal patterns.” By computing the PDF of the possible ordinal patterns, various information-theory quantifiers, such as the permutation entropy, the mutual information, complexity measures, etc., can be computed. The BP method has been successfully employed to analyze time series generated from physical, biological, and social systems (see, e.g., Ref. 11 and references therein).

When employing the BP methodology, the precise values of the SAT anomalies are neglected (as the method is based on comparing relative values in the time series); however, as we will show, with the BP method, one can identify patterns of oscillatory behavior that tend to repeat from time to time, with various time scales. A drawback of the BP method is that, in order to capture long memory processes, long time series are needed to compute the PDFs of the ordinal patterns with good statistics. The SAT data available (described in the next section) limited us to construct ordinal patterns of maximum length 5, which allows to consider time scales up to 5 yr or 5 months. To overcome this limitation we employed “binary representations,” by which the time series of SAT anomalies were transformed into sequences of 0 and 1 s. These binary representations allowed to consider processes with longer time scales, up to 6 yr or 6 months. We will show in what follows that ordinal patterns and binary representations are tools that, when employed within a complex network perspective, are very powerful for the analysis of climatological data. By revealing long- and short-term memory processes, they provide additional information to that obtained from conventional time series analysis, and thus, they help to a better understanding of our complex climate.

This article is organized as follows: Sec. II presents the description of the data analyzed and a summary of the methodology employed. Section III presents the results obtained with ordinal patterns and binary representations, and a comparison with the methodologies previously employed by other authors (i.e., the linear cross-correlation<sup>4</sup> and the nonlinear histogram-based mutual information<sup>6</sup>). Section IV contains a discussion of the results and the conclusions.

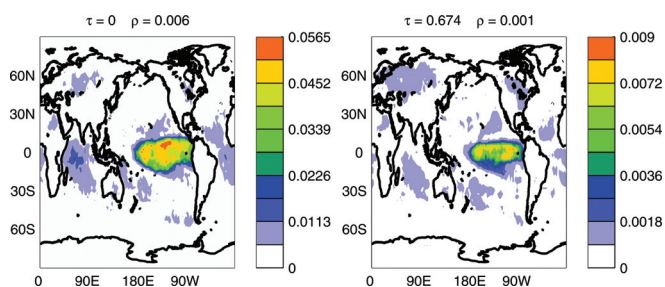


FIG. 2. (Color online) Zero-threshold (left) and high-threshold (right) networks constructed by computing the mutual information from ordinal patterns of length  $D = 5$  defined by comparing SAT anomalies in consecutive years. The 2D plots are color-coded such that the white (red) regions indicate the geographical areas with zero (largest) area weighted connectivity. The weaker links lose a bit of memory (compare the zero-threshold networks with  $D = 4$  and  $D = 5$ ) while the strong links do not, as the high-threshold networks are the nearly same for  $D = 4$  and  $D = 5$ .

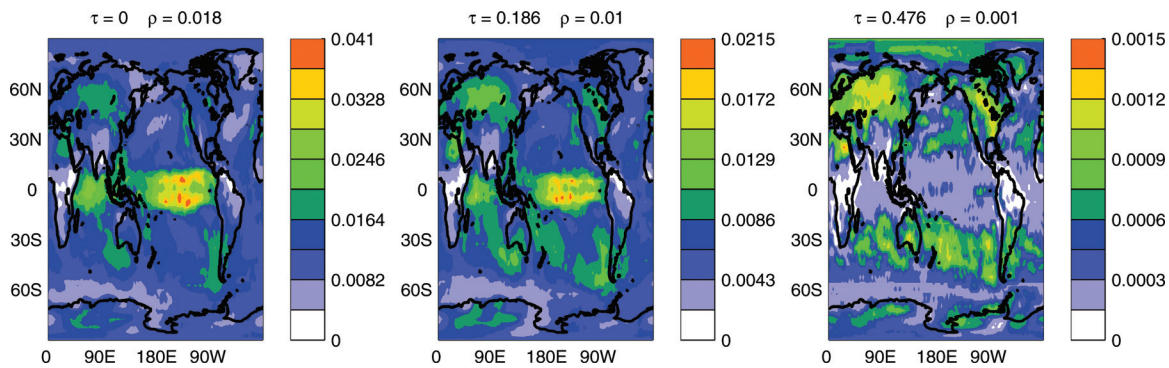


FIG. 3. (Color online) As Fig. 1 but  $D = 4$  ordinal patterns defined by comparing SAT anomalies in consecutive months. The 2D plots of the area weighted connectivity are color-coded such that the white (red) regions indicate the geographical areas with zero (largest) area weighted connectivity.

## II. DATA AND METHODS

We present the analysis of the monthly averaged surface air temperature [SAT field, reanalysis data from the National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) (Ref. 12)]. As in Refs. 2, 4, 6, and 7, anomaly values are considered (i.e., the actual temperature value minus the monthly average).

The data covers a regular grid over the earth's surface with latitudinal and longitudinal resolution of  $2.5^\circ$ . These  $N = 10226$  grid points are considered the nodes (or vertices) of a network (or graph), and the existence of a link (or edge) between any two nodes depends on the "weight" of the link that measures the degree of statistical similarity between the climate dynamics in those two nodes. The data covers the period January 1949 to December 2006, and therefore in each grid point  $i$  ( $i = 1 \dots N$ ), we have  $M = 696$  data points,  $\{x_i(t), t = 1 \dots M\}$ .  $W = \{w_{ij}, i, j = 1 \dots N\}$  is the matrix that contains the weights that characterize the links between any two nodes. Because we do not attempt to uncover directionality in the couplings among the nodes, we will consider a symmetric measure of statistical similarity that results in symmetric weights.

In Refs. 2 and 4 these weights were quantified with the absolute value of the linear cross-correlation coefficient; in Refs. 6 and 7, with the mutual information, a nonlinear measure that is a function of the probability density functions (PDFs) that characterize the time series in the two nodes,  $p_i(m)$  and  $p_j(n)$ , as well as of the joint probability,  $p_{ij}(m, n)$

$$W_{ij} = M_{ij} = \sum_{m,n} p_{ij}(m, n) \log \frac{p_{ij}(m, n)}{p_i(m)p_j(n)}. \quad (1)$$

The mutual information, which can also be written as

$$M_{ij} = S_i + S_j - S_{ij}, \quad (2)$$

where  $S_i = -\sum p_i \log p_i$ ,  $S_j = -\sum p_j \log p_j$ , and  $S_{ij} = -\sum p_{ij} \log p_{ij}$  indicates the amount of information of  $\{x_i(t)\}$  we obtain by knowing  $\{x_j(t)\}$ , and vice versa.  $M_{ij}$  measures the degree of statistical interdependence of the time series; if they are independent,  $p_{ij}(m, n) = p_i(m)p_j(n)$  and  $M_{ij} = 0$ .

To uncover correlated "patterns" of oscillatory behavior in the SAT anomalies, we employ the methodologies referred to as ordinal patterns and symbolic analysis, which

are based on comparing consecutive values in the time series, to compute the PDFs in Eq. (1). We begin by presenting the ordinal pattern methodology.<sup>10</sup>

First, in each grid point  $i$ , the time series  $\{x_i(t)\}$  is divided into  $M - D$  overlapping vectors of dimension  $D$ . Then, each element of a vector is replaced by a number from 0 to  $D - 1$ , in accordance with its relative magnitude in the ordered sequence (0 corresponding to the smallest and  $D - 1$  to the largest value in each vector). For example, with  $D = 3$ , the vector  $(v_0, v_1, v_2) = (6.8, 11.5, 11)$  gives the ordinal pattern 201 because  $v_2 < v_0 < v_1$ . In this way, each vector has associated an "ordinal pattern" (OP) composed by  $D$  symbols, and the symbol sequence comes from a comparison of neighboring values. Last, one computes the PDF of the  $D!$  possible ordinal patterns. For example, with  $D = 3$  the  $3! = 6$  different patterns are (012, 021, 102, 120, 201, and 210), and thus, the PDF is calculated with six bins. To have a good statistics one must have  $M - D \gg D!$  (i.e., # of OPs in the time series  $\gg$  # of possible OPs).

Because in each time series we have  $M = 696$  data points, to compute the PDFs with good statistics, we limit to consider only  $D = 4$  and  $D = 5$ . Ordinal patterns of  $D \leq 3$  do not provide good resolution for computing the mutual information, Eq. (1), because the PDFs are calculated with very few bins (for  $D = 6$ , there are only six ordinal patterns and, thus, only six bins).

With climatological data, meaningful ordinal patterns can be formed either by comparing consecutive years or

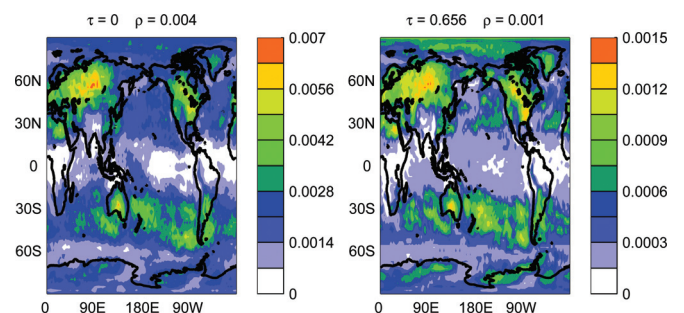


FIG. 4. (Color online) Zero-threshold (left) and high-threshold (right) networks constructed by computing the mutual information from ordinal patterns of length  $D = 5$  defined by comparing SAT anomalies in consecutive months. The 2D plots of the area weighted connectivity are color-coded such that the white (red) regions indicate the geographical areas with zero (largest) area weighted connectivity.



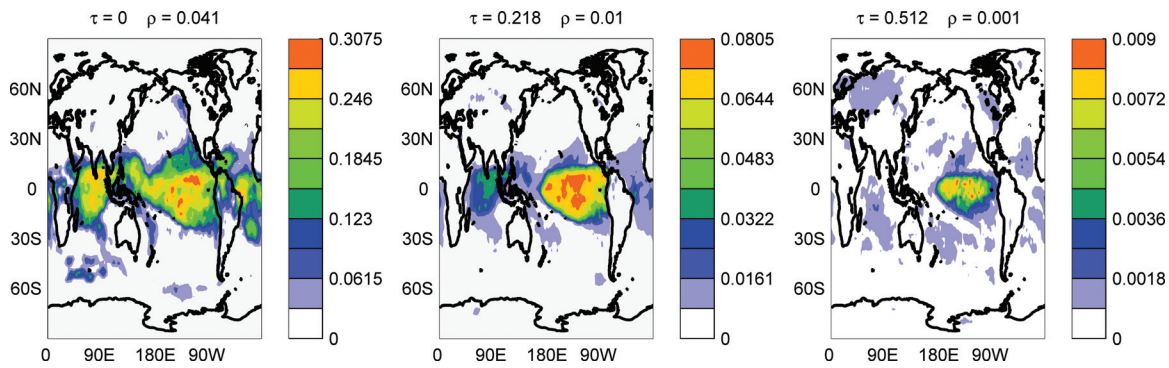


FIG. 5. (Color online) As Fig. 1 but employing binary representation. Zero-threshold (left), low-threshold (center), and high-threshold (right) networks constructed by computing the mutual information from patterns of length  $D = 4$  defined by comparing SAT anomalies in consecutive years. The 2D plots are color-coded such that the white (red) regions indicate the geographical areas with zero (largest) area weighted connectivity. In each panel, the values of the threshold,  $t$ , and of the edge-density,  $\rho$ , are indicated.

consecutive months. Specifically, if we use  $D = 3$ , when comparing consecutive years, the OPs in node  $i$  are defined by  $(x_i(t), x_i(t + 12), x_i(t + 24))$ ,  $t = 1, \dots, M - 24$ ; when comparing consecutive months, they are defined by  $(x_i(t), x_i(t + 1), x_i(t + 2))$ ,  $t = 1, \dots, M - 2$ .

To decide whether there is a link between two nodes, we perform global thresholding,<sup>13</sup> i.e., we define a threshold  $\tau$  (which is the same for all pairs of nodes) and assume that there is a link between  $i$  and  $j$  if the weight of the link is above the threshold, i.e.,  $w_{ij} \geq \tau$ .

Clearly, a careful selection of the threshold is crucial for uncovering the backbone of the network.<sup>6</sup>

We use the following procedure: first, we check that we only take into account significant network connections. To do this, we compute the weight matrix  $W$  from randomly shuffled time series in each node. The random elements of this  $10\,226 \times 10\,226$  matrix have a very narrow PDF which, in principle, allows the use of the maximum matrix element,  $w_{\max}$ , as a significant limit. Then, we compute  $W$  with the original time series and consider that there is a significant

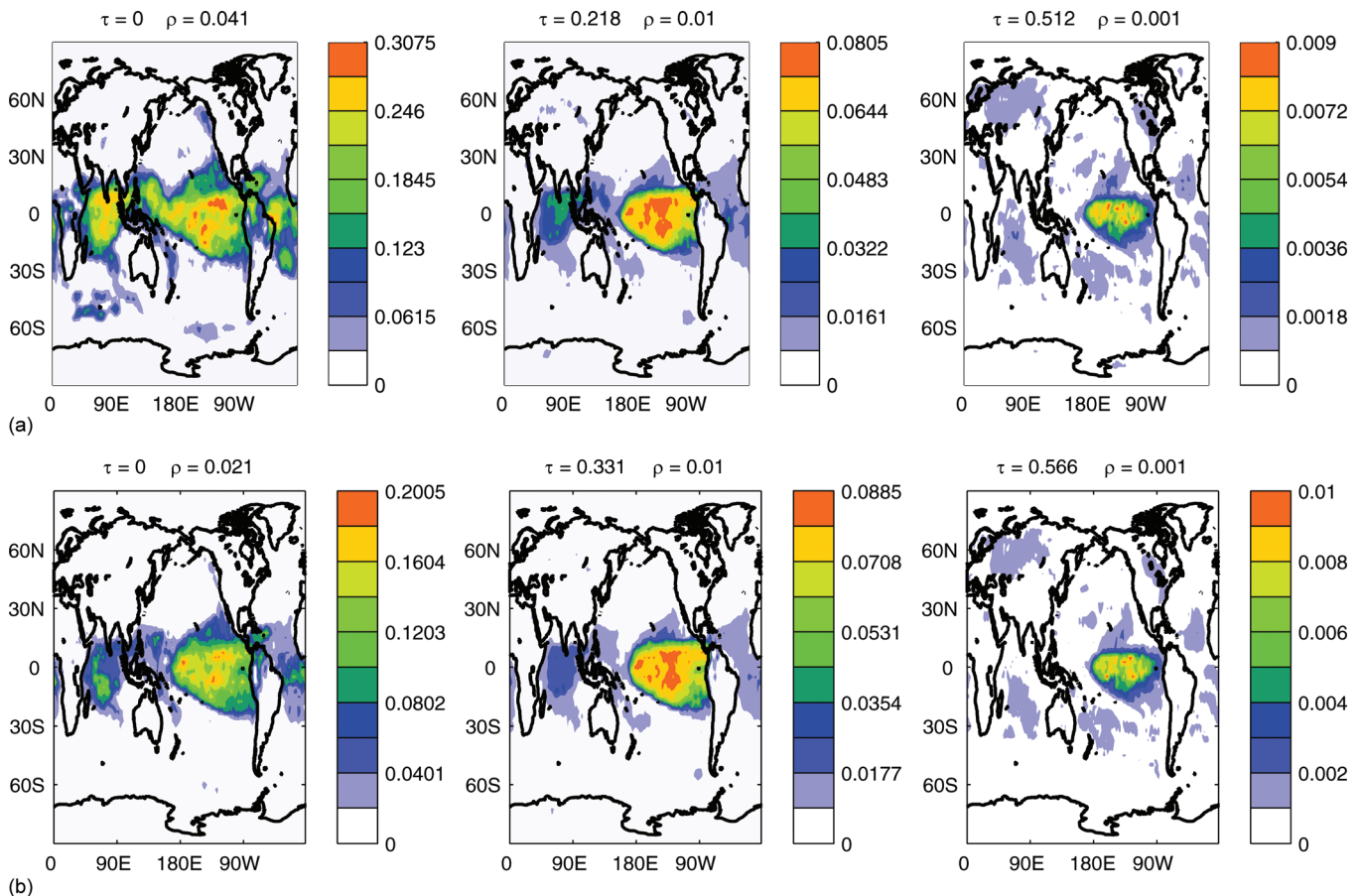


FIG. 6. (Color online) As Fig. 5 but with  $D = 5$ .



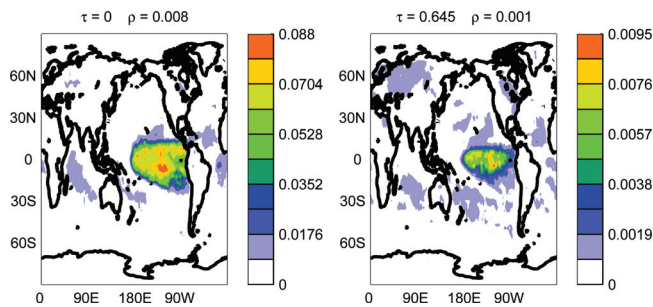


FIG. 7. (Color online) As Fig. 5 but with  $D = 6$ . Comparing with Figs. 5 and 6 one can see that there is a good agreement with the results obtained previously with ordinal patterns: the weaker links tend to lose memory, while the strongest links do not.

link between the nodes  $i$  and  $j$  if  $w_{ij} > w_{\max}$ , otherwise, we set  $w_{ij} = 0$ . While there are several methods to eliminate non-significant links and the evaluation of statistical significance is still an open problem (see, e.g., the discussion in Ref. 6), this procedure is computationally cost-efficient and we will show in what follows that allows to uncover meaningful climate networks. The drawback is that it is a rather strong test that eliminates weak but significant links, and as a result, the networks tend to be very sparse. The final step is to choose a threshold  $t$  to select the strongest links. As in Ref. 6, we chose  $t$  such that the resulting networks have a predetermined number of links. In the following, we present results for networks that have 1% of the total possible links (which will be referred to as “low-threshold” networks) and networks containing 0.1% of the total possible links (referred to as “high-threshold” networks). For easy comparison and to visualize the effect of thresholding, we also present the networks containing *all the significant links*, which will be referred to as the “zero-threshold” networks.

The networks are represented graphically as two-dimensional maps by plotting the *area-weighted connectivity*,<sup>2,4,6,7</sup> which is the fraction of the total area of the earth to which each node  $i$  is connected

$$AWC_i = \frac{\sum_j A_{ij} \cos(\lambda_i)}{\sum_j \cos(\lambda_j)}, \quad (3)$$

where  $\lambda_i$  is the latitude of node  $i$  and  $A_{ij} = 1$  if nodes  $i$  and  $j$  are connected (i.e., if  $w_{ij} \geq \tau$ ) and 0 otherwise. The cosine

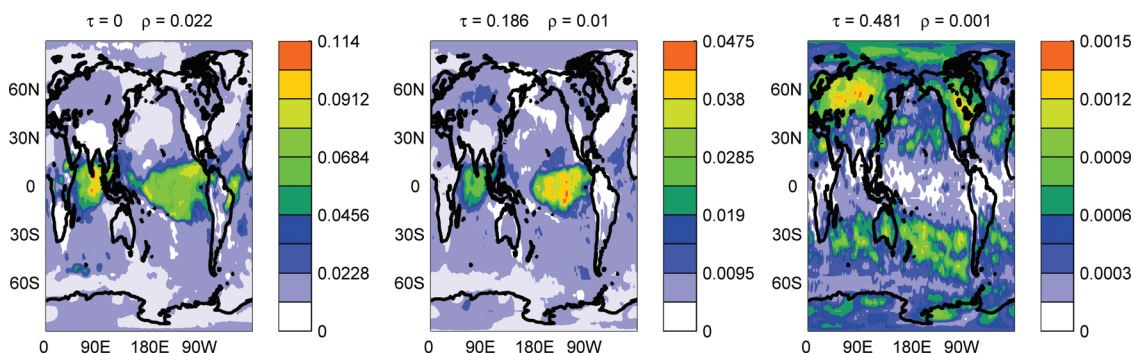


FIG. 8. (Color online) As Fig. 5 but the networks constructed with binary representation comparing anomalies in  $D = 4$  consecutive months. The 2D plots of the area weighted connectivity are color-coded such that the white (red) regions indicate the geographical areas with *zero* (largest) area weighted connectivity.

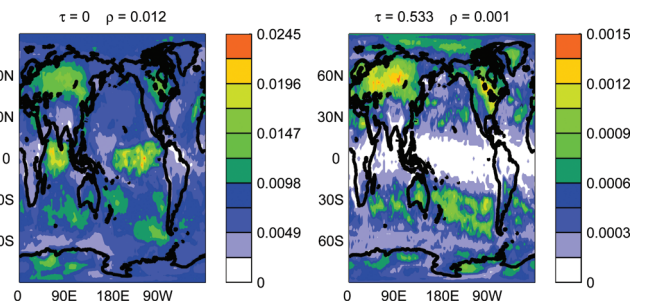


FIG. 9. (Color online) As Fig. 8 but with  $D = 5$ .

terms correct for the fact that in a surface spherical network defined on a regular planar grid, the nodes correspond to regions of different area.

### III. RESULTS

#### A. Ordinal patterns analysis

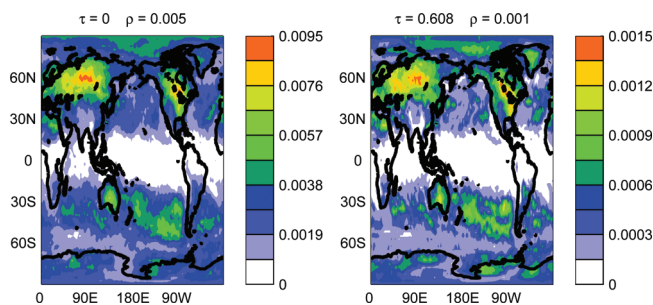
The networks obtained when the ordinal patterns are defined by comparing SAT anomalies in consecutive years and in consecutive months are displayed in Figs. 1–4. In each panel, the values of the threshold,  $t$ , and of the edge-density,

$$\rho = \frac{\sum_{ij} A_{ij}}{N(N-1)}, \quad (4)$$

are indicated.

For consecutive years, the networks with  $D = 4$  (Fig. 1) and  $D = 5$  (Fig. 2) are very similar, showing highest connectivity for the tropical region.

For zero-threshold (left panels in Figs. 1 and 2), the tropical Pacific shows the largest connectivity, particularly in the central and eastern side of the basin; the tropical Atlantic and Indian oceans follow. In the extra tropics, there are patches of high connectivity off the western coast of Canada in the Northern Hemisphere (N.H.) and in the south Pacific in the Southern Hemisphere (S.H.). This connectivity structure is more pronounced for  $D = 5$ , although some of the weak links lose memory. These characteristics hint to El Niño as a fundamental player in setting up these connections.<sup>14</sup> The El Niño phenomenon occurs on interannual time

FIG. 10. (Color online) As Fig. 8 but with  $D = 6$ .

scales and consists in an anomalous warming of the eastern equatorial Pacific. This warming in turn warms up the local atmosphere and influences other tropical regions through the excitation of equatorial Kelvin and Rossby waves. Changes in the precipitation associated with El Niño also induce stationary Rossby waves in the northern and southern extra tropics that generate long range connections called atmospheric teleconnection patterns. Examples of these structures are the Pacific-North American pattern that affects the northern Pacific and North America and the Pacific-South American pattern that propagates in the southern Pacific toward South America. These anomalous structures connect the tropical Pacific with remote locations and affect the local climate by changing, for example, the advection of heat or moisture into a region.

For non-zero-thresholds (center and right panels in Figs. 1 and 2), only the strongest links remain and the networks clearly show again an El Niño-like structure in the tropical Pacific. Secondary maxima in the Indian and Atlantic oceans are present in the low-threshold networks but are significantly weakened in the high-threshold ones. The continental regions have overall very low connectivity, which translates in the low predictability of surface temperature anomalies on interannual time scales. Within these continental regions, the largest connectivity is seen over Asia and North America, the latter maximum being perhaps due to the Pacific North American pattern induced by El Niño.<sup>15</sup>

The networks obtained when the ordinal patterns are defined by comparing temperature anomalies in consecutive months (Figs. 3 and 4) present, for zero- and low-threshold, similar features as for consecutive years, although the net-

works are more homogeneous. There is a maximum in the equatorial Pacific, a secondary maximum in the Indian ocean and extra tropical maxima over Asia, North America, and southern subtropics. On the other hand, the high-threshold network shows that the strongest links are located in the extra tropics. We speculate that this could be a result of the modulation of the temperature variance by the seasonal cycle, which is strongest over the northern hemisphere continental masses and has a minimum in the tropical band. This network structure is also seen when using ordinal patterns formed by five consecutive months, Fig. 4.

## B. Binary representations

To capture longer memory processes one should use larger  $D$  values; however, for  $D = 6$ , there are  $6! = 720$  possible ordinal patterns, and because we have time series with less than 700 data points, there is not enough data to calculate ordinal patterns PDFs with good statistics.

As discussed in the Introduction, a solution to overcome this problem is employing “binary representations,” by which the time series  $\{x_i(t)\}$  is transformed into a sequence  $\{v_i(t)\}$  of 0 and 1 s, using the following rule:  $v_i(t) = 0$  if  $x_i(t) \geq 0$  and  $v_i(t) = 1$  otherwise (because the  $x_i$  values are temperature anomalies, we are taking into account whether the SAT field is above or below its monthly averaged value). We can then define “binary patterns” of dimension  $D$  (e.g., for  $D = 3$ , the possible patterns are 000, 001, 010, 100, 011, 110, 101, and 111) and compute their PDF. The number of different patterns is  $2^D$ , and thus, we can calculate PDFs of patterns of  $D = 6$  ( $2^6 = 64$ ) with good enough statistics. Patterns with  $D \leq 3$  do not provide good resolution for computing the mutual information, Eq. (1), because the PDFs are calculated with very few bins. Therefore, in the following, we consider  $D = 4, 5$ , and 6.

Figures 5–10 present the results when the binary patterns are defined by consecutive years and months.

For consecutive years, Figs. 5–7, the networks obtained when using binary representations are very similar to those found with ordinal patterns. The tropical regions are quite uniformly well connected (although a Pacific maximum is clear) while the extra tropics show localized regions of high connectivity likely due to atmospheric teleconnections forced from the tropics, particularly for low density networks.

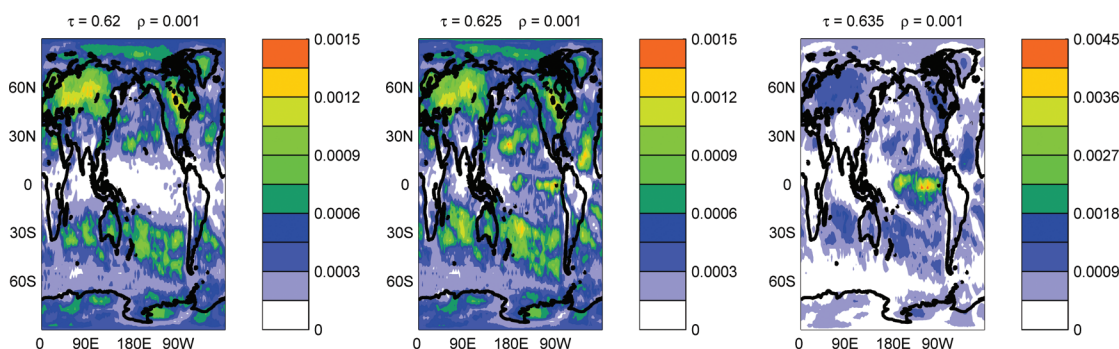


FIG. 11. (Color online) High-threshold networks constructed with binary representations, with patterns of  $D = 6$  covering time-intervals of 1 yr (left), 2 yr (center), and 3 yr (right). The 2D plots of the area weighted connectivity are color-coded such that the white (red) regions indicate the geographical areas with zero (largest) area weighted connectivity. See text for details.

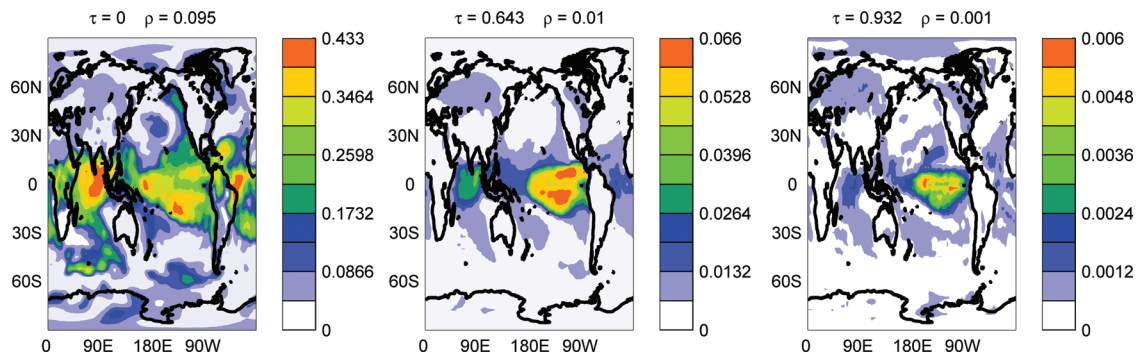


FIG. 12. (Color online) Zero-threshold network (left) and non-zero-threshold networks (center and right) constructed by estimating the weights with the absolute value of the cross-correlation coefficient. The 2D plots of the area weighted connectivity are color-coded such that the white (red) regions indicate the geographical areas with zero (largest) area weighted connectivity.

The networks obtained for consecutive months, Figs. 8–10, show that as the threshold or as  $D$  increases there are overall similar changes in structure as those seen for ordinal patterns, Figs. 3 and 4: for short memory processes (or for low threshold), the maximum connectivity is in the tropics, while for longer time scales (or for higher threshold), the extra tropics show largest number of links. As discussed before, this could be the result of the modulation of the temperature variance by the seasonal cycle, which would be the main process that connects grid points in very low density networks (grid points that are connected by very strong links). Our results could also hint at the role of land surface conditions such as snow or soil humidity in increasing the persistence of surface temperature anomalies over the northern hemisphere continents. Overall, these results agree well with the fact that temperature teleconnections from the tropical Pacific tend to last no much longer than a season in the different parts of the world.

As a way to test the interpretation of the above presented results, in terms of the symbolic methodology of time series analysis capturing two different time scales of the Earth's climate, seasonal and interannual, we constructed binary patterns of fixed dimension,  $D=6$ , that cover three different time intervals

(i) covering 1-yr, the patterns are composed as

$$[x_i(t), x_i(t+2), x_i(t+4), \dots, x_i(t+10)];$$

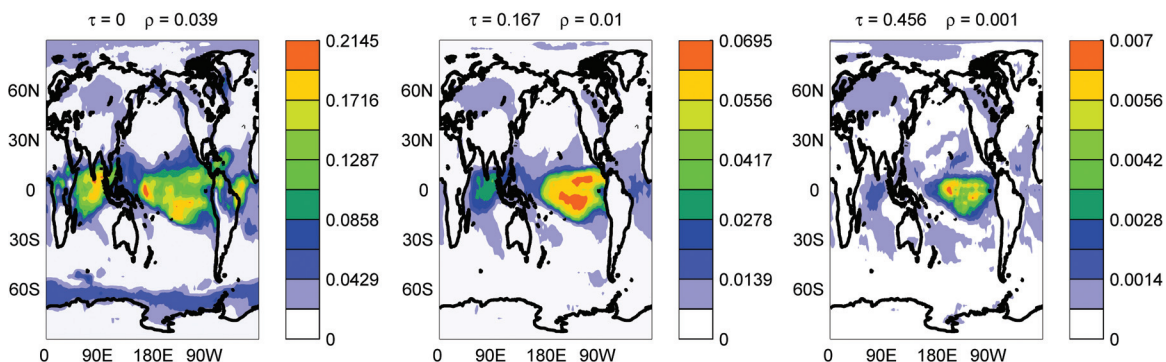


FIG. 13. (Color online) Zero-threshold network (left) and non-zero-threshold networks (center and right) constructed by estimating the weights with the mutual information, calculating the PDFs from histograms of SAT anomalies. The 2D plots of the area weighted connectivity are color-coded such that the white (red) regions indicate the geographical areas with zero (largest) area weighted connectivity.

(ii) covering 2-yr, the patterns are composed as

$$[x_i(t), x_i(t+4), x_i(t+8), \dots, x_i(t+20)];$$

(iii) covering 3-yr, the patterns are composed as

$$[x_i(t), x_i(t+6), x_i(t+12), \dots, x_i(t+30)].$$

The results are presented in Fig. 11, where one can see how the network changes. For a time interval of 1 yr, the extra tropics have the largest number of links and there are very few in the tropical region. On the other hand, for a time interval of 3 yr, the extra tropics keep about the same number of connections while in tropical Pacific “El Niño” stands out (note the different color scales in the panels in Fig. 11). In summary, confirming the results previously found with ordinal patterns and binary representations composed by consecutive years and by consecutive months, on intraseasonal time scales the extra tropical connections dominate, while on interannual scales, the “El Niño” is the key player in setting up teleconnections worldwide.

As it was previously discussed, the significance of the network links was tested in comparison with links computed from surrogate time series in each node. Since surrogate data does not preserve the autocorrelation properties of the original time series, to further test the validity of the previously presented results, we did the following test: we computed the mutual information using the original time series in node  $i$  and the time-inverted series in node  $j$ . The results show no



significant spatial structure in the area weighted connectivity plots (not shown).

### C. Comparison with other measures

It is interesting to compare the results obtained using ordinal patterns and binary representations with those obtained using conventional techniques of time series analysis, as the linear cross-correlation coefficient (as in Ref. 4) and the mutual information, computing the PDFs from standard histograms of amplitude values (as in Refs. 6 and 7). Figure 12 displays the zero-, low-, and high-threshold networks when the weights are calculated with the absolute value of the cross-correlation coefficient and Fig. 13, when they are calculated with mutual information, with the PDFs calculated from histograms of temperature anomaly values. In this case, the PDFs were computed employing 32 bins, and in each time series, the values of the SAT anomalies were renormalized such that each time series has zero mean and standard deviation equal to one (as in Refs. 6 and 7, for easier comparison).

The 2D plots of the area weighted connectivity are similar to those previously reported in Ref. 6 and also, to those seen in Figs. 1, 2, 5, and 6, where the ordinal patterns and the binary representations are formed by comparing consecutive years. “El Niño” is the main feature uncovered. There are also regions with relatively high number of links in the northern hemisphere continents and southern subtropics, but the high connectivity in the extra tropics seen previously in Figs. 3, 4, 8–10 is not observed. In other words, employing the cross-correlation coefficient or the histogram-based mutual information uncovers mainly the interannual network. These methodologies fail to separate the two distinct time scales (intraseasonal and interannual) that are clearly seen when using symbolic analysis and the time series are transformed in sequences of patterns by comparing consecutive years or consecutive months.

### IV. CONCLUSIONS

Concluding, we have shown that ordinal patterns and symbolic analysis applied to anomalies of the surface air temperature are powerful tools for the analysis of the large-scale topology of the climate network. The success of these methods is based on an appropriate partition of the phase space that results in ordinal patterns and binary representations having PDFs that characterize the diversity of patterns present in the climate dynamics.

A main advantage of the methodology proposed here is that by varying the dimension of the pattern and the year-month comparison, one can uncover memory processes with different time scales. We found that both, monthly and

yearly, patterns reveal long memory processes, and that depending on the time scale considered the climate network can change completely.

The fact that ordinal patterns and symbolic analysis give meaningful information indicates that the time variability of the anomaly SAT field is strongly determined by patterns of oscillatory behavior that tend to repeat from time to time.

Overall we found that on seasonal time scales the extra tropical regions, mainly over Asia and North America, present the strongest links while in interannual time scales, the tropical Pacific clearly dominates.

### ACKNOWLEDGMENTS

The authors thank the two anonymous referees for their very useful comments and suggestions. M.B. acknowledges support from the European Community’s Seventh Framework Programme (FP7/2007-2013) under Grant Agreement N 212492 (CLARIS LPB. A Europe-South America Network for Climate Change Assessment and Impact Studies in La Plata Basin). C.M. acknowledges support from the ICREA Academia programme, the Ministerio de Ciencia e Innovacion, Spain, project FIS2009-13360, and the Agencia de Gestio d’Ajuts Universitaris i de Recerca (AGAUR), Generalitat de Catalunya, through project 2009 SGR 1168.

<sup>1</sup>A. Arenas, A. Diaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, *Phys. Rep.* **469**, 93 (2008).

<sup>2</sup>A. A. Tsonis, K. L. Swanson, and P. J. Roebber, *Bull. Amer. Meteorol. Soc.* **87**, 585 (2006).

<sup>3</sup>K. Yamasaki, A. Gozolchiani and S. Halvin, *Phys. Rev. Lett.* **100**, 228501 (2008).

<sup>4</sup>A. A. Tsonis and K. L. Swanson, *Phys. Rev. Lett.* **100**, 228502 (2008).

<sup>5</sup>A. A. Tsonis, K. L. Swanson, and G. Wang, *J. Climate* **21**, 2990 (2008).

<sup>6</sup>J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, *Eur. Phys. J. Spec. Top.* **174**, 157 (2009).

<sup>7</sup>J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, *Eur. Phys. Lett.* **87**, 48007 (2009).

<sup>8</sup>S. Bialonski, M. T. Horstmann, and K. Lehnertz, *Chaos* **20**, 013134 (2010).

<sup>9</sup>R. Q. Quiroga, T. Kreuz, and P. Grassberger, *Phys. Rev. E* **66**, 041904 (2002).

<sup>10</sup>C. Bandt and B. Pompe, *Phys. Rev. Lett.* **88**, 174102 (2002).

<sup>11</sup>J. M. Amigo, *Permutation Complexity in Dynamical Systems: Ordinal Patterns, Permutation Entropy and All That* (Springer, Berlin, Germany, 2010).

<sup>12</sup>E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, B. Reynolds, R. Jenne, and D. Joseph, *Bull. Amer. Meteorol. Soc.* **77**, 437 (1996).

<sup>13</sup>V. M. Eguíluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian, *Phys. Rev. Lett.* **94**, 018102 (2005).

<sup>14</sup>K. E. Trenberth, G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, *J. Geophys. Res.* **103**, 14291 (1998).

<sup>15</sup>C. F. Ropelewski and M. S. Halpert, *Mon. Wea. Rev.* **114**, 23522362 (1986).