Large-Scale Atmospheric Phenomena Under the Lens of Ordinal Time-Series Analysis and Information Theory Measures

J.I. Deza, G. Tirabassi, M. Barreiro, and C. Masoller

Abstract This review presents a synthesis of our work done in the framework of the European project *Learning about Interacting Networks in Climate* (LINC, climatelinc.eu). We have applied tools of information theory and ordinal time series analysis to investigate large scale atmospheric phenomena from climatological datasets. Specifically, we considered monthly and daily Surface Air Temperature (SAT) time series (NCEP reanalysis) and used the climate network approach to represent statistical similarities and interdependencies between SAT time series in different geographical regions. Ordinal analysis uncovers how the structure of the climate network changes in different time scales (intra-season, intra-annual, and longer). We have also analyzed the directionally of the links of the network, and we have proposed novel approaches for uncovering communities formed by geographical regions with similar SAT properties.

Keywords Climate networks • Nonlinear time series analysis • Climate communities • Information transfer

1 Introduction

Complex networks constitute the huge revolution in nonlinear science in the twentieth century because it provides a unified framework for the study of a wide range of real-world complex systems, such as the Internet, social networks, transport networks, ecological and metabolic networks, and even the human brain (Albert and Barabasi 2002; Newman 2003; Boccaletti et al. 2006).

For understanding and extracting information from observed data, various methods for mapping statistical interdependencies between time series into “functional”

J.I. Deza • G. Tirabassi • M. Barreiro
Instituto de Física, Facultad de Ciencias, Universidad de la República, Igua 4225, Barcelona, Spain

C. Masoller (✉)
Departament de Fisica, Universitat Politecnica de Catalunya, Colom 11, Terrassa, 08222, Barcelona, Spain
e-mail: cristina.masoller@upc.edu

A.A. Tsonis (ed.), *Advances in Nonlinear Geosciences*,
DOI 10.1007/978-3-319-58895-7_4

networks have been proposed. These methods for constructing complex networks from data are complemented by a careful analysis of the inferred network, in order to detect fake links, missing links, hidden nodes, etc. (Timme 2007; Serrano et al. 2009; Shandilya and Timme 2011; Yu and Parlitz 2011; Rubido et al. 2014; Tirabassi et al. 2015a, b).

Considering the complexity of the inter-relations between the different elements that constitute the climate system, it is clear that the analysis of observed climatological data from a complex network perspective has great potential for yielding light into relevant, previously unknown features of our climate.

Indeed, in the last two decades the research field of climate networks has provided important insight into complex phenomena in our climate (Tsonis and Roebber 2004; Tsonis and Swanson 2008; Yamasaki et al. 2008; Donges et al. 2009; Barreiro et al. 2011; Fountalis et al. 2014; Hlinka et al. 2014; Tirabassi et al. 2015a, b). Nowadays climate networks are a research field located at the triple intersection of three active areas in nonlinear science: network theory, time series analysis, and climate dynamics.

The European project *Learning about Interacting Networks in Climate* (LINC, climatelinc.eu) brought together researchers from these communities with the goals of training the new generation of researchers, developing cutting-edge science, and promoting new collaborations. Here we present a summary of some of our results developed within the LINC project.

2 Time-Scale Analysis of Climate Interactions

The work by Barreiro et al. (2011) was a first approach to characterize the climate network by means of recurrent oscillatory patterns, with various time scales, as described by using symbolic *ordinal analysis* (Bandt and Pompe 2002). By mapping these processes into a climate network, we found that the structure of the network changes drastically at different time scales.

The symbolic method of ordinal analysis first divides a time series $x(t)$ of length M 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" (OP) 201 because $v_2 < v_0 < v_1$. In this way, each vector has associated an OP composed by D symbols, and the symbol sequence comes from the comparison of neighboring values. With $D = 3$ the $3! = 6$ different patterns are (012, 021, 102, 120, 201, and 210). Last, the presence of recurrent oscillatory patterns in the time series is characterized by means of the probabilities of the ordinal patterns, computed from their frequency of occurrence in the time series.

A classical measure to investigate mutual interdependencies between time series is the mutual information (MI), which is computed from the probability distribution functions (PDFs) associated to the two time series, and the joint probability distribution. When using the ordinal probabilities to compute the MI, the PDF is computed with $D!$ bins, and the joint probability, with $D! \times D!$ bins. Therefore, to have a good statistics one must have enough data points in the time series, i.e., $M \gg D! \times D!$

A main advantage of ordinal analysis is that it allows selecting the time scale of the analysis by comparing L-lagged data points instead of consecutive data points. For example, in SAT reanalysis with monthly resolution, by comparing four data points separated by twelve months (i.e., (v_0, v_{12}, v_{24}, v_{36})) we can investigate recurrent oscillatory patterns with a characteristic time scale of 4 years.

When using an interdependency statistical measure, such as the MI, to define the links of the climate network, one must use an appropriate criterion to define which MI values are considered significant and represented as network links. Performing such significance analysis is a challenging task. A particularly important problem for climate networks is the fact that, due to physical proximity (i.e., due to the spatial embedding of the network), the strongest links are those between neighboring regions. Therefore, by using a high significance threshold, one ends up with a network in which long-distance links are scarce. On the other hand, by choosing a low significance threshold, a lot of "noise" is included in the network as fake links. Therefore, the challenge is how to select the threshold that provides the best compromise between the need to include relevant long-distance links that represent genuine atmospheric teleconnections, and the need to limit the proliferation of noisy links.

The networks obtained from Surface Air Temperature (SAT, NCEP/NCAR monthly reanalysis covering the period January 1949 to December 2006) with ordinal patterns formed by comparing SAT anomalies in the same month during four consecutive years (i.e., $D = 4$ and $L = 12$) are shown in Fig. 1 (Barreiro et al. 2011). In this figure, the networks obtained with different MI significance thresholds are shown. One can notice that in this "inter-annual" time scale the

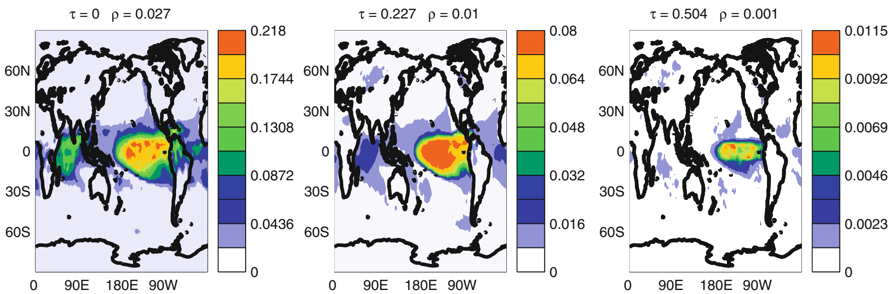

Fig. 1 Climate networks constructed by computing the mutual information (MI) from the probabilities of ordinal patterns of length $D = 4$ defined by comparing SAT anomalies (NCEP/NCAR monthly reanalysis) in consecutive years ($L = 12$). The color-code is such that the *white* (*red*) regions indicate the geographical areas with zero (largest) area weighted connectivity. The significance threshold increases from *left* to *right*. Adapted from Barreiro et al. (2011)

dominant atmospheric connections are located in the tropical Pacific and Indian Ocean areas, mainly associated with El Niño phenomenon. One can also notice that, as expected, the connectivity of the network decreases as the MI significance threshold is increased. For the highest threshold considered (shown in the right panel, here the threshold is selected such that the density of the network is 0.1% of the total possible links) the El Niño—Indian Ocean teleconnection is significantly weakened with respect to the lower threshold network (shown in the left panel, here the threshold is chosen equal to the maximum MI value obtained from surrogated data, which gives a network with 2.7% of the total links).

Figure 2 summarizes the effect of the lag L used to define the ordinal patterns (Deza et al. 2013). When the OPs are defined in terms of consecutive months (top row) the network links are mainly local. In the seasonal time scale (middle row) the tropical region becomes connected. Clearly, the extra-tropics become connected to the equatorial Pacific through atmospheric teleconnection processes only when considering inter-annual time scales (bottom row).

Figure 3 displays the climate network when the mutual information is computed with the classical approach, i.e., computing the PDFs from the histograms of values in the time series (i.e., without taking into account the ordering of the data points). We note that the network looks as a "superposition" of spatial structures which are present only in some of the maps shown in Fig. 2. See, for example, the highly connected green spot in the Labrador Sea, which is also seen in Fig. 2a and to a lesser extent in Fig. 2b; but is not present in Fig. 2c. The Labrador Sea is one of the most important regions of deep water formation in the north Atlantic. The formation of this water occurs in wintertime and depends on the passage of extratropical storms that cool the surface. The passage of storms is in turn related to the state of the North Atlantic Oscillation. As a result, there is a clear connection of the Labrador Sea with the rest of the north Atlantic mainly on seasonal time scales and is mostly independent of ENSO activity.

3 Climate Communities

Many natural systems can be represented by networks with modular structure in the form of communities of densely interconnected nodes. In the context of climate networks, climate communities can be understood as a set of geographical regions that share some common property (dynamical or statistical) of the climate in those regions.

The existence of such regions is expected because of the physical processes that govern our climate (ocean and atmospheric processes, solar forcing, vegetation, human activity, etc.), act in a similar way in distant regions (having similar effects), and therefore, distant regions can have similar climate. Examples include tropical rainforests, dry and arid regions, maritime regions, etc.

The methodology for constructing climate networks described in the previous section is not appropriate for detecting such community structure (i.e., regions which have similar climatic properties) because, as seen in Figs. 2 and 3, the short-

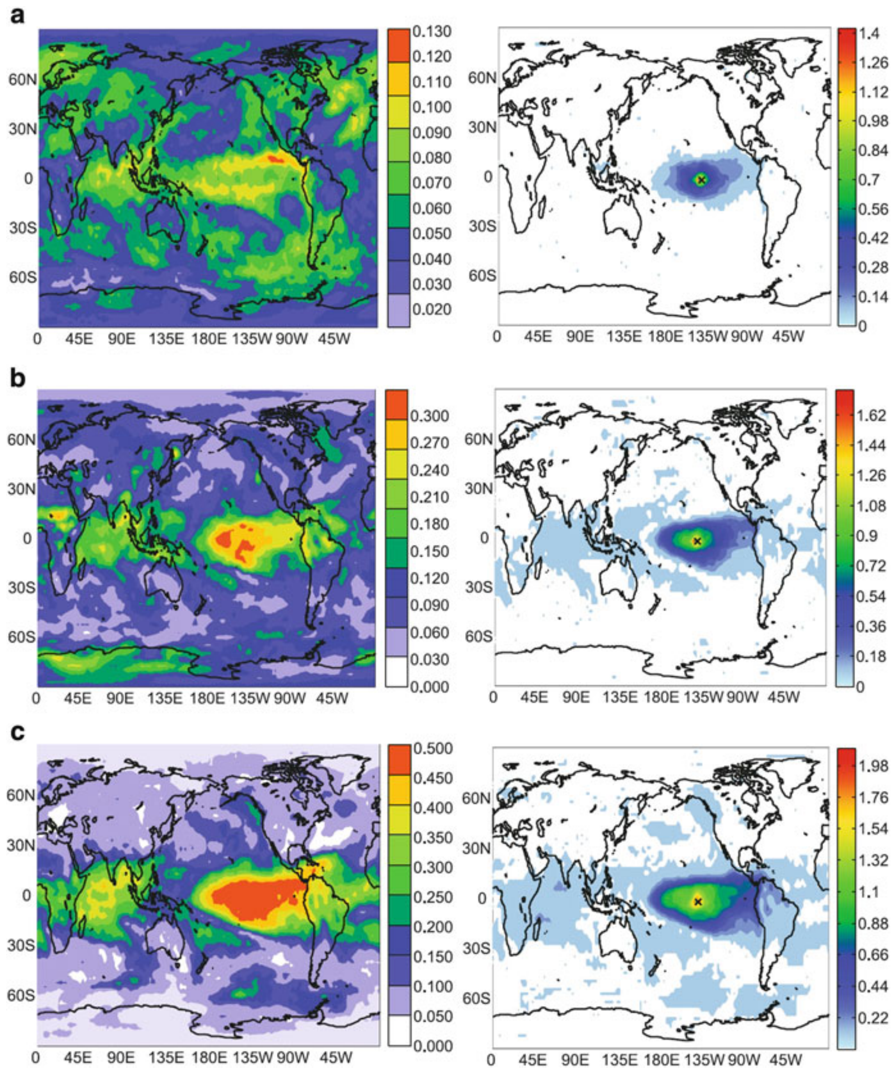

Fig. 2 Area weighted connectivity (*left column*) and connectivity maps (mutual information values of the significant links of the node indicated by an *x*, *right column*) using $D = 3$ OPs formed with three consecutive months ($L = 1$, *top row*), OPs formed with three equally spaced months covering a one-year period ($L = 3$, *middle row*); and OPs formed with 3 months in consecutive years ($L = 12$, *bottom row*). Adapted from Deza et al. (2013)

distance links between neighboring nodes dominate, and the northern and southern hemispheres are only indirectly or weakly connected. Therefore, in this network, areas of tropical rainforests, for example, which are located in different hemispheres won't be identified as belonging to the same community, because there won't be links that interconnect them.

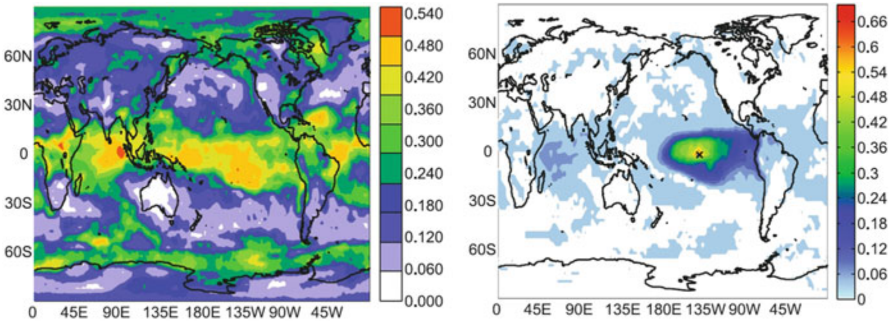

Fig. 3 As Fig. 2 but when the mutual information is computed from the PDFs of SAT anomaly data values. Adapted from Deza et al. (2013)

Recently we proposed a novel methodology to construct the network that allows overcoming this problem (Tirabassi and Masoller 2016). The methodology, based on ordinal analysis, allows to group together regions that share similar properties of the symbolic dynamics.

The main idea is to assign a high (low) weight to the link between two regions, if the ordinal transition probabilities (TPs) that describe the statistics of the symbolic sequence are very similar (very different) in the two regions. In other words, the symbolic sequences are mutually compared in terms of the probability of pattern "A" being followed by pattern "B." Then, a significance threshold is used to keep only the regions that have very similar transition probabilities. The third step was to run the Infomap community detection algorithm (Rosvall and Bergstrom 2007) in order to identify the groups of densely interconnected regions.

Figure 4 summarizes the results of the analysis. Panel a displays the communities uncovered when the network is constructed with the classical approach (in this case, the similarity measure used is the Pearson cross-correlation coefficient) and panel b displays the communities uncovered by means of the novel approach, based on the similarity of the ordinal transition probabilities.

By using the classical approach with a threshold $W = 0.5$ (Tsonis and Roebber 2004), Infomap algorithm uncovers 8604 communities, but only 20 are composed by more than two nodes. Figure 4a displays the largest 16 communities. The detected communities include the central-east equatorial Pacific dominated by El Niño, the tropical western Pacific, Indian Ocean, and tropical north Atlantic regions controlled mainly by the exchange of heat fluxes with the atmosphere, and the equatorial Atlantic cold tongue dominated by dynamical air–sea interaction. The other communities are small and some may be just noise.

In contrast, with the ordinal approach the community structure inferred, shown in Fig. 4b, divides the world in eight areas that share similar climatic properties, as measured by similar symbolic transition probabilities. There are five macro-communities: extratropical continents and southern ocean characterized by large SAT variability (indicated with number 0), northern oceans (2), regions of tropical

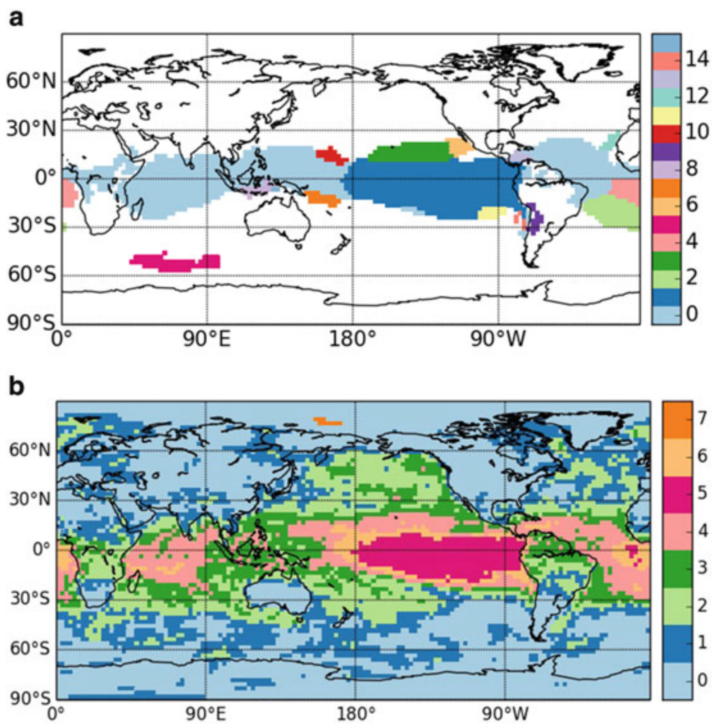

Fig. 4 Community structure uncovered by Infomap algorithm. The different communities are indicated with different colors. (**a**) The network is constructed by using the Pearson cross-correlation coefficient as a measure of dynamical similarity. (**b**) The network is constructed by calculating the similarity of the ordinal transition probabilities. In panel (**a**), for clarity, only the largest 16 communities are shown. Adapted from Tirabassi and Masoller (2016)

deep convection such as the western Pacific warm pool, Amazon and Congo basins (3), tropical oceans dominated air–sea heat fluxes (4) and ENSO basin (5). Then, there are also two boundary communities, indicated with numbers 1 and 6, which are placed at the communities interfaces.

Both methodologies identify the region dominated by the El Niño dynamics as a community, but there are differences in the rest. Compared to the communities calculated with the classical approach the new methodology is able to separate better in terms of processes dominating the SAT variability. For example, the new methodology (1) identifies the central equatorial Atlantic as having a similar behavior to El Niño, which is consistent with the literature (Zebiak 1993); (2) separates the behavior of SAT over the maritime continent from that of the Indian and tropical Atlantic oceans, consistent with a different rainfall regime, (3) considers the tropical north and south Atlantic as belonging to the same community, which is consistent because temperature is strongly controlled by air–sea heat fluxes.

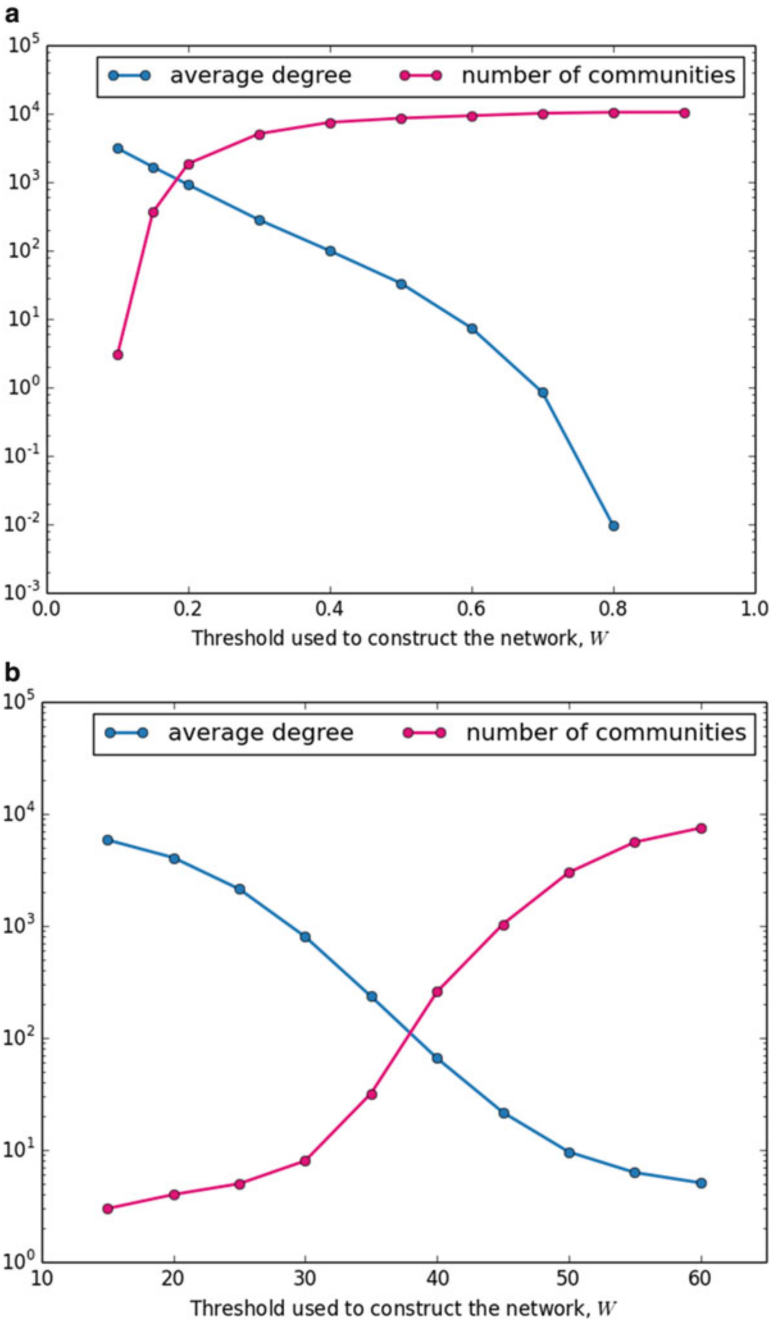

Fig. 5 Network average degree and number of communities vs. the threshold used to construct the network, W. In panel (**a**) the network is constructed by using the Pearson cross-correlation coefficient. For the community structure shown in Fig. 4a, the threshold used was $W = 0.5$ (as in

As discussed before, in order to construct a climate network, the links weights have to be pruned by using an adequate threshold. Decreasing the threshold leads to a more connected network, while increasing it results in a sparser one. The number of communities depends on the number of connections, which in turn depends on the threshold. In order to uncover a coherent, well-defined community structure, the threshold has to be carefully chosen. Figure 5 displays the number of communities and the average degree as a function of the threshold. It can be seen that there is a negative correlation between them. The fragmentation of the network into smaller communities (as community seven in Fig. 4b) can be due to the removal of relevant links that keep the bigger communities together. Thus, to obtain a meaningful community structure, we selected *ad hoc* a threshold that provided the best compromise between the need to limit the small-communities-proliferation and the need to include in the network only the relevant links.

4 Net Direction of Climate Interactions

A main drawback of the methodology discussed in the previous sections for inferring the climate network is that it uses a symmetric similarity measure (the mutual information or the Pearson correlation coefficient) that yield non-directed networks. In these networks the presence of a link indicates inter-dependency but the direction of the underlying interaction is not inferred. For improving the understanding of climate phenomena and its predictability, it is of foremost importance not only to be able to infer the presence of a link between two nodes but also to infer the direction of this interaction.

Deza et al. (2015) used a methodology that allows inferring directed interactions via an analysis of the net direction of information transfer. A measure was used—based on conditional mutual information—that quantifies the amount of information in a time-series $x(t)$, contained in τ time units in the past of another time series $y(t)$. The resulting network was found to be in full agreement with state-of-the-art knowledge in climate phenomena, validating in this way the methodology used. No assumptions about physical processes were made, except for the appropriate setting of the parameter τ.

Fig. 5 (continued) Tsonis and Roebber 2004). In panel (**b**) the network is constructed by calculating the similarity of the ordinal transition probabilities. For uncovering the communities shown in Fig. 4b, the threshold used was $W = 30$. It can be observed that with the first approach, a very low threshold needs to be used to uncover a small set of communities. However, using a low threshold has the strong disadvantage of including in the network many links which are not significant. In contrast, with the novel approach (by constructing the network considering the similarities of the transition probabilities), the variation of the number of communities with the threshold is more gradual, which allows uncovering a small set of communities by using a threshold that is not too low. Adapted from Tirabassi and Masoller (2016)

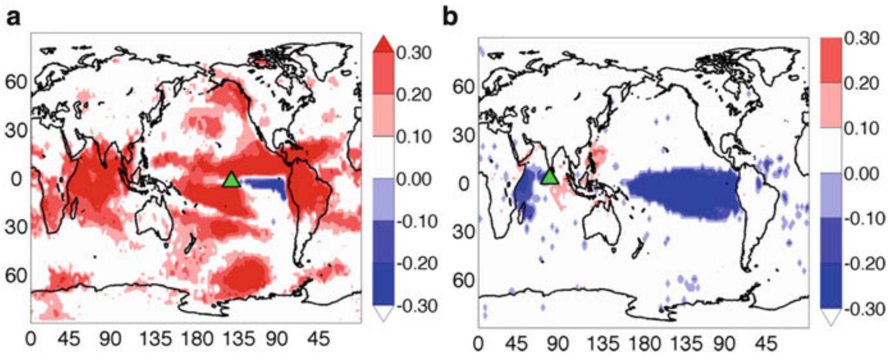

Fig. 6 Directionality of the links in a node in the central pacific (**a**) and in a node in the Indian Ocean (**b**) indicated with a *triangle*. The color code indicates the directionality index: outgoing links are shown in *red* while incoming links are shown in *blue*. The time scale of information transfer is $\tau = 30$ days. Adapted from Deza et al (2015)

The directionality measure and the statistical significance analysis are discussed in detail in Deza et al. (2015). Here we present two examples that illustrate the directional structure of the network. Figure 6 displays the directionality of the links of two nodes in the tropics (indicated with triangles) computed from SAT reanalysis data with daily resolution and parameter $\tau = 30$ days. The color code in this figure indicates the Directionality Index (DI): outgoing links are shown in red, while the incoming links are shown in blue.

Figure 6a shows, as expected, the central Pacific influenced by the eastern Pacific (in blue) and influencing the global network, with many regions in the tropics and in the extra-tropics in red. Reciprocally, Fig. 6b shows that the blue links come to the node in the Indian Ocean from a well-defined region in the central Pacific Ocean. In addition, few red outgoing links connect the node in the Indian Ocean to other regions. A main drawback of the directionality index used is that it does not distinguish indirect from direct information transfer. Therefore, the red areas influenced by the node in the Indian Ocean can be an artifact in the sense that these regions might be directly influenced by El Niño region.

Figure 7 displays the influence of the parameter τ that characterizes the time scale of the information transfer from one node to another. As an example, a region in southeastern South America is considered (indicated with a triangle). For synoptic time scales of a few days, the directionality index uncovers the existence of a wave train propagating with a southwest-northeast direction. Moreover, there is a clear separation line between regions with incoming and outgoing links. This configuration is characteristic of the progression of a front through the reference point. As the parameter τ increases, the extratropical wave train associated with synoptic time scales fades and only blue links to the tropics remain, consistent with an influence of the equatorial Pacific on the region on longer time scales, likely related to ENSO.

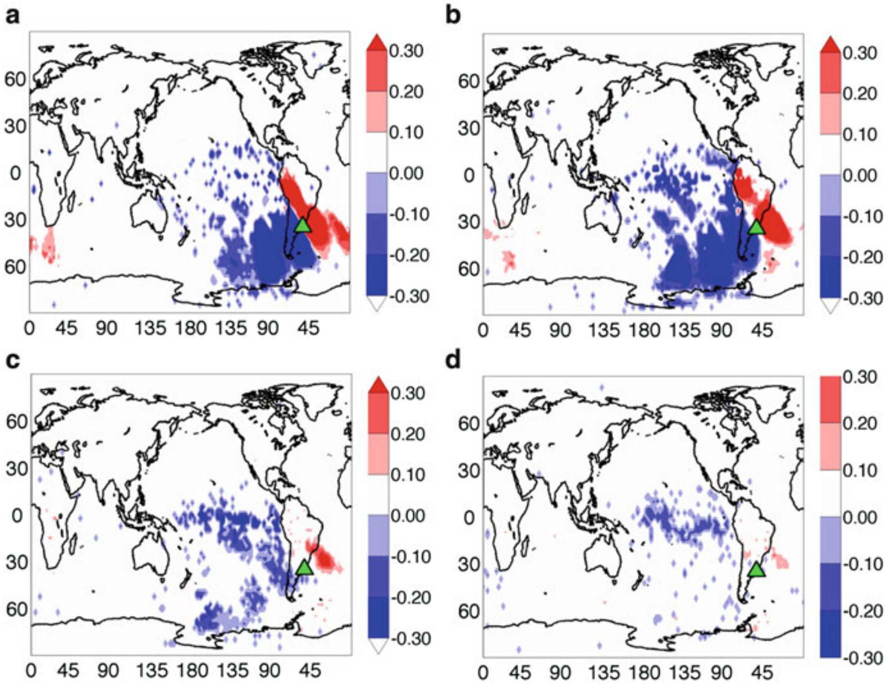

Fig. 7 Directionality of the links in a node in southeastern South America, indicated with a *triangle*. The color code indicates the directionality index: outgoing links are shown in *red* while incoming links are shown in *blue*. The time scale of information transfer τ is (**a**) 1 day, (**b**) 3 days, (**c**) 7 days, and (**d**) 30 days. Adapted from Deza (2015)

5 Conclusions

We have shown that symbolic time series analysis based on ordinal patterns and information theory measures, applied to surface air temperature anomalies (reanalysis data with monthly or daily resolution) are powerful tools for uncovering the large-scale structure of the climate network.

A main advantage of the ordinal methodology is that, by varying the dimension of the pattern and the year–month comparison, one can uncover memory processes with different time scales, and depending on the time scale considered, the climate network can change completely. Overall we found that on seasonal time scales the extra-tropical regions, mainly over Asia and North America, present strong connectivity, while in inter-annual time scales, the tropical Pacific clearly dominates.

A novel methodology for inferring the community structure of the climate network was proposed. Constructing the climate network by taking into account the similarity of the ordinal transition probabilities in different regions allowed to identify communities formed by geographical regions where the climate variability displays similar statistics of ordinal patterns. Five macro-communities were identified: extratropical continents, northern oceans, tropical convective regions, tropical oceans, and ENSO basin.

The analysis of the net directionality of the links revealed variability patterns consistent with well-known features of the global climate dynamics. For example, in the extra-tropics, the link direction revealed wave trains propagating from west to east, in both hemispheres. A drawback of the directionality index employed is that it does not distinguish direct from indirect interactions.

Ongoing and future work is aimed at exploring the suitability of other techniques of time series analysis, such as Hilbert analysis (Zappalà et al. 2016), other directionality measures (partial directed coherence and directed partial correlation, Tirabassi et al. 2017), and measures of distances between time series and entropy measures (Arizmendi et al. 2017) for gaining additional information from climatological datasets.

Acknowledgements This work was supported by the LINC project (FP7-PEOPLE-2011-ITN, Grant No. 289447), Spanish MINECO (FIS2015-66503-C3-2-P) and ICREA ACADEMIA.

References

Albert, R., and A.L. Barabasi. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74: 47–97.

Arizmendi, F., M. Barreiro, and C. Masoller. 2017. Identifying large-scale patterns of nonlinearity and unpredictability in atmospheric data. *Scientific Reports* 7: 45676.

Bandt, C., and B. Pompe. 2002. Permutation entropy: a natural complexity measure for time series. *Physical Review Letters* 88: 174102.

Barreiro, M., A.C. Marti, and C. Masoller. 2011. Inferring long memory processes in the climate network via ordinal pattern analysis. *Chaos* 21: 013101.

Boccaletti, S., V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. 2006. Complex networks: structure and dynamics. *Physics Reports* 424: 175–308.

Deza, J.I., M. Barreiro, and C. Masoller. 2013. Inferring interdependencies in climate networks constructed at inter-annual, intra-season and longer time scales. *The European Physical Journal Special Topics* 222: 511–523.

Deza, J.I., M. Barreiro, and C. Masoller. 2015. Assessing the direction of climate interactions by means of complex networks and information theoretic tools. *Chaos* 25: 033105.

Donges, J.F., Y. Zou, N. Marwan, and J. Kurths. 2009. The backbone of the climate network. *EPL* 87: 48007.

Fountalis, I., A. Bracco, and C. Dovrolis. 2014. Spatio-temporal network analysis for studying climate patterns. *Climate Dynamics* 42: 879–899.

Hlinka, J., D. Hartman, M. Vejmelka, D. Novotna, and M. Palus. 2014. Non-linear dependence and teleconnections in climate data: sources, relevance, nonstationarity. *Climate Dynamics* 42: 1873–1886.

Newman, M.E.J. 2003. The structure and function of complex networks. *SIAM Review* 45: 167–256.

Rosvall, R., and C.T. Bergstrom. 2007. An information-theoretic framework for resolving community structure in complex networks. *PNAS* 104: 7327–7331.

Rubido, N., A.C. Martí, E. Bianco-Martínez, C. Grebogi, M.S. Baptista, and C. Masoller. 2014. Exact detection of direct links in networks of interacting dynamical units. *New Journal of Physics* 16: 093010.

Serrano, M.A., M. Boguña, and A. Vespignani. 2009. Extracting the multiscale backbone of complex weighted networks. *PNAS* 106: 6483–6488.

Shandilya, S.G., and M. Timme. 2011. Inferring network topology from complex dynamics. *New Journal of Physics* 13: 013004.

Timme, M. 2007. Revealing network connectivity from response dynamics. *Physical Review Letters* 98: 224101.

———. 2016. Unravelling the community structure of the climate system by using lags and symbolic time-series analysis. *Scientific Reports* 6: 29804.

Tirabassi, G., C. Masoller, and M. Barreiro. 2015a. A study of the air–sea interaction in the South Atlantic Convergence Zone through Granger causality. *International Journal of Climatology* 35: 3440.

Tirabassi, G., R. Sevilla-Escoboza, J.M. Buldú, and C. Masoller. 2015b. Inferring the connectivity of coupled oscillators from time series statistical similarity analysis. *Scientific Reports* 5: 10829.

Tirabassi, G., L. Sommerlade, and C. Masoller. 2017. Inferring directed climatic interactions with renormalized partial directed coherence and directed partial correlation. *Chaos* 27: 035815.

Tsonis, A.A., and P. Roebber. 2004. The architecture of the climate network. *Physica A* 333: 497–504.

Tsonis, A.A., and K.L. Swanson. 2008. Topology and predictability of El Nino and La Nina networks. *Physical Review Letters* 100: 228502.

Yamasaki, K., A. Gozolchiani, and S. Havlin. 2008. Climate networks around the globe are significantly affected by El Nino. *Physical Review Letters* 100: 228501.

Yu, D., and U. Parlitz. 2011. Inferring network connectivity by delayed feedback control. *PloS One* 6: e24333.

Zappalà, D.A., M. Barreiro, and C. Masoller. 2016. Global atmospheric dynamics investigated by using Hilbert frequency analysis. *Entropy* 18: 408.

Zebiak, S.E. 1993. Air–sea interaction in the equatorial Atlantic region. *Journal of Climate* 6: 1567.