Inferring interdependencies in climate networks constructed at inter-annual, intra-season and longer time scales

J. Ignacio Deza[1a], Marcelo Barreiro[2], and Cristina Masoller[1]

1 Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya,
  Colom 11. E-08222, Terrassa, Barcelona, Spain.
2 Instituto de Física, Facultad de Ciencias, Universidad de la República,
  Iguá, 4225, Montevideo, Uruguay.

Abstract. We study global climate networks constructed by means of ordinal time series analysis. Climate interdependencies among the nodes are quantified by the mutual information, computed from time series of monthly-averaged surface air temperature anomalies, and from their symbolic ordinal representation (OP). This analysis allows identifying topological changes in the network when varying the time-interval of the ordinal pattern. We consider intra-season time-intervals (e.g., the patterns are formed by anomalies in consecutive months) and inter-annual time-intervals (e.g., the patterns are formed by anomalies in consecutive years). We discuss how the network density and topology change with these time scales, and provide evidence of correlations between geographically distant regions that occur at specific time scales. In particular, we find that an increase in the ordinal pattern spacing (i.e., an increase in the timescale of the ordinal analysis), results in climate networks with increased connectivity on the equatorial Pacific area. On the contrary, the number of significant links decreases when the ordinal analysis is done with a shorter timescale (by comparing consecutive months), and interpret this effect as due to more stochasticity in the time-series in the short timescale. As the equatorial Pacific is known to be dominated by El Niño-Southern Oscillation (ENSO) on scales longer than several months, our methodology allows constructing climate networks where the effect of ENSO goes from mild (monthly OP) to intense (yearly OP), independently of the length of the ordinal pattern and of the thresholding method employed.

1 Introduction

Complex networks appear in almost all fields of science, examples being the internet, social interactions, food webs, biochemical reactions, brain functional networks and so forth [1–5]. Many systems lead naturally to the concept of networks of interacting elements, where one can define nodes and assign links among them depending on the

---

a e-mail: juan.ignacio.deza@upc.edu

(in principle, very complex) features of the system under study. By using the network approach it is then possible to extract relevant information about a system without over simplifying it and without having to handle the full scale detailed model which can obscure the interpretations. It also provides of a mathematical framework for analyzing data, taking advantage of the well-known graph theory.

This is precisely the situation in the fast-developing field of climate networks. As a novel approach to information analysis and data mining, the complex network method has been successfully employed; several properties of the network topology have been investigated and related to specific climate phenomena [6–16].

Since the atmosphere connects geographically far away regions through waves and advection of heat and momentum, this long-range coupling makes the network modeling approach of the Earth's climate extremely attractive. By covering the Earth' surface with a regular grid of points (nodes), and by assigning links between any pair of nodes depending on their climate interdependency, the network approach has been able to extract novel and meaningful information about the Earth climate. The interdependency of the climate in two nodes, i and j, has been quantified in terms of the statistical similarity of the time series of climatological variables, say surface air temperature (SAT) anomalies, $x_i(t)$ and $x_j(t)$, recorded in nodes i and j. The usual measures employed for this quantification are the Pearson coefficient (the absolute value of the cross-correlation) and the mutual information.

Indeed, the authors of Ref. [14] provided a complete background for constructing climate networks using the Pearson cross correlation and the mutual information as measures for determining the links' strength. Special emphasis was placed in assessing the statistical significance of the links. Once the network was constructed, a number of measures were used to characterize it, including edge density, clustering coefficient, average path length, and giant component size. The application of this novel approach to specific climatic phenomena or on specific areas in the world [17–23] has also been successful. For example, in Ref. [19] the authors showed a connection between the climate network topology and the predictability of surface temperature in the tropical area affected by ENSO. Two networks were constructed using data from the El Niño and from La Niña years from the past 60 years. Comparing these networks using graph theory measures, the authors found that the El Niño network has a lower number of links, smaller degree distribution, lower clustering coefficient and shorter characteristic path length than La Niña, and concluded that, because of this reduced connectivity, the temperature field in the area was less predictable during El Niño years. This result was verified using alternative variance temperature data.

The stability of climate networks with time was studied in Refs. [20, 21]. In Ref. [20] a climate network was constructed from daily surface temperature anomalies and, by computing correlations over a short time window, two types of links were identified, referred to as "blinking", which appear and disappear, and "robust", which are present over long time intervals. Using this methodology the authors were able to detect a sharp lowering of the total number of the links of the network happening during the El Niño years, and argued that this lowering is a very sensitive method for determining ENSO periods. Tracking the blinking links in other parts of the globe revealed a response to "El Niño" phenomenon even in zones where the mean temperature is not affected by it. In [21] the study was extended to examine climate networks constructed from other daily data. Daily temperature or geopotencial height data present high variability and are hard to predict, but the authors found long lasting links between the nodes yielding to a robust network pattern. Part of this robustness was attributed to the geographical regular embedding of the network; however, physical robust coupling between different locations was also found, with the coupling on the equator being significantly less pronounced than in the extratropics.

Recently, two of us used the framework of nonlinear time-series analysis to construct global climate networks that take into account characteristic time scales in the climate [24]. Specifically, climate interdependencies were quantified via symbolic analysis of monthly-averaged SAT anomalies. Two methods were employed to transform anomaly time-series into symbolic sequences: ordinal patterns [25,26] and binary representations. These symbolic transformations take into account the order in which the anomalies occur (i.e., increasing or decreasing with respect to past and future values), but do not take into account the magnitude of the anomaly value.

In climatological data oscillatory patterns occur, which tend to repeat from time to time, with various time scales. Ordinal pattern time-series analysis allows selecting specific scales, and the calculation of the mutual information of the ordinal sequence, allows quantifying the similarity of two time-series in the selected time scale. For example, by varying the interval covered by the pattern in an intra-season or in an inter-annual time-scale, one can construct climate networks with different topologies that can then be interpreted in terms of known climate processes, with shorter or longer memory. Different network topologies were found when the symbolic pattern was formed by comparing SAT anomalies in consecutive months (i.e., taking into account memory effects with intra-season time-scales) and when the symbolic pattern was formed by comparing SAT anomalies in consecutive years (i.e., taking into account longer memory effects with inter-annual time-scales).

To quantify climate interdependencies the mutual information (MI) has been employed, which is computed from single and joint probabilities, p_i, p_j and p_{ij}, associated with either the full time-series $x_i(t)$ and $x_j(t)$ of SAT anomalies in nodes i and j, or with their ordinal representations, $s_i(t)$ and $s_j(t)$.

When the full time-series are considered, the probabilities p_i, p_j and p_{ij} are approximated in terms of histograms computed with a certain number of bins, N_{bin}, which is limited only by the length of the time series. In the symbolic approach, the number of bins is defined by the number of possible patterns, which in turn is determined by the number symbols in the ordinal pattern. If, for example, we consider patterns formed by 3 (or 4) symbols, then there are 3!=6 (or 4!=24) possible patterns, and this is the number of bins for computing the probabilities associated with the symbolic sequences. We remark that even when the number of symbols in the pattern is kept fixed (let's say 3), the symbolic method still allows for varying the time-interval of the pattern by considering the anomalies in 3 consecutive months (e.g., January, February, March; February, March, April; etc), in 3 consecutive years (e.g., January 2010, January 2011, January 2012; February 2010, February 2011, February 2012; etc.), and also in other time-scales, for example, by considering anomalies in 3 months equally spaced covering a 1-year period (e.g., in this case the ordinal pattern is formed by comparing SAT anomalies in Januaries, Mays, and Septembers; Februaries, Junes and Octobers; etc.).

The goal of the present work is to extend the analysis of [24] in two ways: first, by examining the influence of the methodology used for quantifying statistical similarity, and second, by analyzing the influence of the criteria used for thresholding (i.e., for disregarding the links that are not statistically significant). Regarding the first goal, we compare the MI values computed from the full SAT time-series and from its OP representation, covering different time intervals, and study the influence of the number of N_{bin} used for computing the probabilities associated either to the full SAT time-series or to its OP representation. Regarding the second goal, we consider two thresholding methods, the first one is based on pre-defining the number of links in the network (i.e., the threshold is such that it gives a network with a certain number of links), and the second thresholding method is based on the statistical significance of MI values, which is determined in terms of the distribution of MI values computed from shuffled data.

Links with mutual information values below the threshold are considered not significant and disregarded. Here we show that for low N_{bin} values the number of significant links tends to increase, as compared with larger N_{bin}.

This paper is organized as follows. Section 2 presents the network construction methods, the data used and introduces the filtering techniques. Section 3 discusses the behavior of the networks depending on the construction method employed, with a special focus on the analysis of different time scales involved. Finally, section 4 presents conclusions and discussions.

2 Network construction

We analyze the monthly-averaged SAT anomalies (reanalysis data of the National Center for Environmental Prediction/National Center for Atmospheric Research, NCEP/NCAR) [27]. The anomalies are calculated as the actual temperature values minus the monthly average, and are normalized by the standard deviation. The data cover are given on a grid over the Earth's surface with latitudinal and longitudinal resolution of $2.5°$, resulting in $N = 10226$ grid points or network nodes. The data cover the period from January 1949 to December 2006, and thus, in each node, we have a time series of 696 data points.

The climate network properties will depend on the methodology employed to infer the presence of connections between two nodes, i.e., the procedure used to include a particular link in the network and to filter out those correlations that may have occurred merely by chance.

As in Refs. [15,24], to quantify climate interdependencies we use the mutual information (MI), that is a nonlinear symmetric measure computed in terms of the probabilities, $p_i(m)$ and $p_j(n)$, that characterize the time series in two nodes, i and j, as well as of their joint probability $p_{ij}(m,n)$.[28–30] The time-series of temperature anomalies in any two nodes, $x_i(t)$ and $x_j(t)$, is transformed into discrete time-series series, $s_i(t)$ and $s_j(t)$, which have N_{bin} possible values: $s^1, \ldots, s^{N_{bin}}$. This transformation is done either by using a certain number of bins or by a symbolic transformation (as described below). Then, the probabilities of the values s^m (with $m = 1 \ldots N_{bin}$) in nodes i and j are denoted as $p_i(m)$ and $p_j(m)$ (with $m = 1 \ldots N_{bin}$) and are calculated in terms of the frequency of appearance of s^m in the time series $s_i(t)$ and $s_j(t)$. The joint probability $p_{ij}(m,n)$ is calculated in terms of the frequency of simultaneous appearance of s^m in $s_i(t)$ and s^n in $s_j(t)$.

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

M_{ij} is a measure of the degree of statistical interdependency: if the two time series are independent, $p_{ij}(m,n) = p_i(m)p_j(n)$ and $M_{ij} = 0$. We calculate the mutual information from Eq. (1) with different probabilities associated to the time-series of SAT anomalies: the usual histogram of values (referred to as MIH), and the probabilities of the symbolic ordinal patterns (OPs) [25], referred to as MIOPDL with D being the length of the pattern, which will be either 3 or 4 symbols.

The OPs are calculated from the SAT time series by noting the value of a data point relative to its neighboring values in the series. For instance, OPs of length 3 are formed by 3 symbols in the following way: if a value $(x_i(2))$ is higher than the previous one $(x_i(1))$ but lower than the next one $(x_i(3))$, gives the pattern '123', the opposite case $(x_i(1) > x_i(2) > x_i(3))$ gives the pattern '321', etc. With 3 symbols there are 3! = 6 different patterns. This symbolic transformation allows to detect correlations in the sequence of values (which are not taken into account with histograms of values

that do not take into account the order in which the values appear in the time series), but has the drawback that does not take into account the information about the relative magnitudes.

Ordinal patterns do not need to be constructed only with immediately adjacent data points. We can construct them with data points that are separated in time, and in this way we can consider different time scales. For example, a separation of 12 months allows grouping together individual months of the year, thus reducing the dependence of the magnitude of the anomaly on the seasonal cycle.

In this paper we consider patterns of length 3 and 4 formed by:

- consecutive months (i.e., by comparing three or four consecutive values in the time series); we refer to the MI computed in this way as MIOP3L01 and MIOP4L01 respectively;

- months in consecutive years (i.e., by comparing $x_i(t)$, $x_i(t+12)$, and $x_i(t+24)$); we will refer to the MI computed in this way as MIOP3L12 (OPs of length 3) and MIOP4L12 (OPs of length 4).

- equally spaced months covering a one-year period. For patterns of length 3, this is done by comparing $x_i(t)$, $x_i(t+4)$, and $x_i(t+8)$; for patterns of length 4, by comparing $x_i(t)$, $x_i(t+3)$, $x_i(t+6)$, and $x_i(t+9)$. We refer to the MI computed in these ways as MIOP3L04 and MIOP4L03 respectively.

The next step is to filter the links that are considered not significant, i.e., the links between pairs of nodes i and j such that the M_{ij} value is low enough to be consistent with a random value. The detection of weak significant links requires a careful analysis of the statistical significance of weak correlations, which is challenging task in the case of short and noisy data (see the review by M. Paluš, Ref. [31] where tests for inferring nonlinearity and coupling in atmospheric data, sunspot numbers and brain signals were discussed in detail).
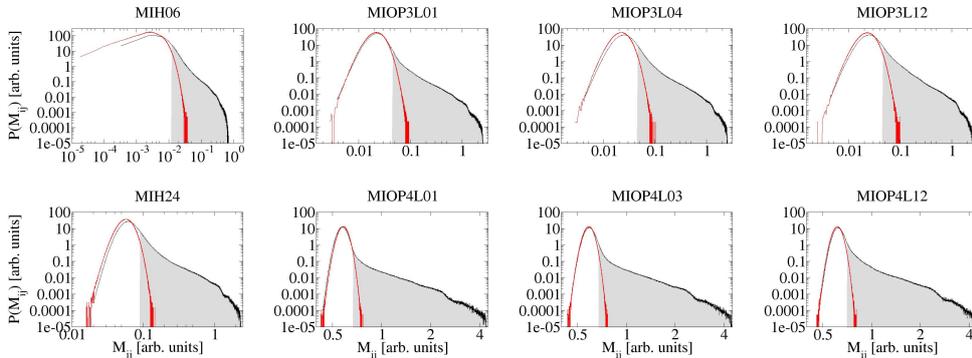

Fig. 1. Distribution of the mutual information values, M_{ij} with $i, j \in [1, N = 10226]$, computed from the original time-series (black line) and from the surrogate time-series (red line) in a double logarithmic scale. In the top row the MI is computed from Eq. (1) with probabilities defined over 6 bins; in the bottom row, over 24 bins. The shadowed area indicates the statistically significant values, which are above a threshold calculated as $\mu + 3\sigma$, with μ and σ being the mean value and the standard deviation of the distribution of M_{ij} values computed from surrogate data. The different methods used to compute the probabilities are described in the text: histograms of values, MIH06 and MIH24 (first column); OPs formed with consecutive months, MIOP3L01 and MIOP4L01 (second column); OPs formed with equally spaced months covering a one-year period, MIOP3L04 and MIOP4L03 (third column); and OPs formed with months in consecutive years, MIOP3L12 and MIOP4L12 (fourth column).

Here we use random re-sampling to test the null hypothesis of independent time-series, for which a Gaussian distribution of MI values is expected: we calculated the M_{ij}^s values (with the supra-indice s representing surrogate data) and found they are Gaussian-like distributed for both MIH and MIOPs (The exception is MIH with $N_{bin} = 6$, see the top-left panel in Fig. 1; however, we verified that this distribution is similar to the one obtained with Gaussian random numbers with small positive mean, when the negative values are disregarded). Therefore, we computed the mean value, μ, and standard deviation, σ, of the M_{ij}^s distribution and chose the significance threshold as $\tau = \mu + 3\sigma$, and a link between nodes i and j was considered significant if the value of M_{ij} was above τ. This criterion gives about 99.87% of confidence that the links have MI values that could not be due to chance. In Fig. 1 the gray section indicates the values that satisfy this condition.

The final step is to represent the network with a commonly used measure, namely, the *area-weighted connectivity* (AWC)[13,15,19,24]. This is done by plotting the number of links every node has, taking into account that the nodes represent geographic regions with different area (points near the poles representing a smaller area than points near the equator). We plot two-dimensional maps in which the color code indicates the AWC of the nodes, which is the fraction of the total area of the Earth to which a node i is connected,

$$AWC_i = \frac{\sum_j^N A_{ij} \cos(\lambda_j)}{\sum_j^N \cos(\lambda_j)}, \qquad (2)$$

where λ_i is the latitude of node i and $A_{ij} = 1$ if nodes i and j are connected (i.e., if the value of M_{ij} is larger than τ and zero otherwise.

We remark that the AWC plots provide information about the size of the area to which a node is connected, but do not indicate to which nodes the node is connected. To provide some information about these connections we chose a node with large connectivity located in the region characterized by ENSO (we refer to this node as X) and plot in color code the value of M_{Xj} if j is connected of X, and in white if j is not connected.

3 Results and discussion

Figures 2 and 3 present the results of the four methods of analysis, considering 6 bins (Fig. 2) and 24 bins (Fig. 3). Ordered from top to bottom, the network was constructed by computing the MIH (top row), the MIOP with OPs formed by consecutive months (second row), the MIOP with OPs covering a one-year period (third row) and with the same month in consecutive years (bottom row). Note that in order to cover a year period, the spacing on the third row in Fig. 2 is four months (as we are using three letters) and it is three months in Fig. 3 (which uses four letters). In both figures the left column presents the AWC plots, and the right column, the connectivity maps.

The AWC obtained here compares well tieh that calculated in [24] (left panel, Fig. 1), except that the network density is about double. This is explained by a different way to calculate the significance threshold. In [24] the threshold consisted in the largest MI value computed from shuffled data, while in Fig. 3 of this study the threshold is defined in terms of the standard deviation of the MI distribution calculated from shuffled data.

Considering the left column of Figs. 2 and 3, (showing the AWC plots for 6 bins/3 letters and 24 bins/4 letters respectively), one can observe highly connected spots on the first row which are present only in some of the other three maps. See, for

example, the highly connected green spot in the Labrador Sea, which is also seen in the second and to a lesser extent in the third row; but is not present in the plot on the fourth row. 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 increasing its density. The passage of storms is in turn related to the state of the North Atlantic Oscillation, a preferred pattern of atmospheric variability in the north Atlantic basin. As 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 on ENSO activity (see second row of Figure 2).

In addition, in the first row of the same Fig. 2, we can notice highly connected areas in Africa, the equatorial Atlantic and western tropical north Atlantic which are not present in the short-time scale networks (second row) but that are seen in the long-time scale networks, (third and fourth rows). Thus, these connections arise because regions are connected on inter-annual, but not on monthly or seasonal time scales.

On interannual time scales El Niño teleconnections include a decrease of the northern trade winds that reduce the heat loss in the western tropical Atlantic, as well as a tropospheric warming over most of the tropical band, thus inducing warming over several regions, including the equatorial Atlantic and the Indian ocean [32]. The air-sea interaction in the equatorial Atlantic leads to inter-annual modes of variability which can interact constructively or not with the anomalies induced by El Niño, resulting in less number of links compared to those in the Indian ocean [33]. Therefore, the OP symbolic method allows to see how the network topology is modified by climate processes acting on different time scales.

In spite of the fact that we have used the same significance criterion for the four networks presented above, the network density (defined as ρ = number of links present in the network over the total possible number of links, which is equal to $10226 \times 10225/2$ because the network is symmetric and self-links are not included) is significantly different. Other climate network construction methods that are based on quantifying global similarities or interdependencies in time-series (using, e.g., MIH or cross-correlation as a similarity measures) lack this capability and are therefore not suitable for uncovering climate phenomena that take place in specific time scales.

On interannual time scales ENSO also influences climate over the northeastern Pacific and the south Pacific, including the Antarctic Peninsula, probably through the propagation of Rossby waves (Figure 2, last row). The latter Pacific link has been recently suggested as the mechanism responsible for the temperature trend in the Antarctic Peninsula during the last 30 years [34].

According to these results the tropical region becomes connected on seasonal time scales, while the extratropics become connected to the equatorial Pacific only when considering interannual time scales. These teleconnection patterns evidence the propagation of Kelvin and Rossby waves from the equatorial Pacific on different time scales. While the overall picture is similar in Figs. 2 and 3, it is clear that the use of 6 bins represents more adequately the known atmospheric processes, probably due to the shortness of the time series that prevents weak links to be declared significant when using 24 bins to calculate the mutual information.

To analyze the influence of the significance criterion, in Figs. 4 and 5 we present the networks when the threshold to define the links is such that all the networks have the same number of links.We have chosen a density of links of 0.027%, because all the networks presented in Fig. 2 and 3 have a density equal or larger than this value. This is the approach used in [24]. In this previous work, it was found that the short-time scale network MIOP4L01 was more uniformly connected than the year-time scale network MIOP4L12. The present analysis confirms this result and extends it to the

case where the density depends on the significance test (Figs 2 and 3). This suggests
that on short time scales there is no dominant phenomenon that interconnects remote
regions. Instead, temperature anomalies seem to be governed by regional patterns of
atmospheric internal variability.

4 Conclusions

We have studied different aspects of the global climate associated to short time scales
(of a few months) and to longer time scales (of a few years) using the frameworks
of climate networks and nonlinear time-series analysis. The goal was to examine the
influence on the network topology of i) the methodology used for quantifying the
degree of statistical similarity in two nodes, and ii) the significance criteria used for
thresholding, to define the links.

We employed the mutual information calculated from probabilities that were de-
fined over i) a small number of bins (6, Figs. 2 and 4), and ii) a large number of bins
(24, Figs. 3 and 5); also, the probabilities were computed directly from histograms of
SAT anomaly values, and from the symbolic, ordinal representation of the anomaly
time series. This was done in order to compare the dependence of the network topolo-
gies on the number of letters of the ordinal representation.

The latter method allowed considering different time-scales when transforming
the anomalies time series into ordinal patterns. We also considered the influence of
thresholding, and defined the threshold i) in terms of surrogate data (significance
test) (Figs. 2, 3) and ii) to obtain a network with a given link density (Figs. 4, 5).
As the equatorial Pacific is known to be dominated by ENSO on scales longer than
several months, our method allows us to obtain networks where the effect of ENSO
goes from mild (monthly OP) to intense (yearly OP) independently of the number of
letters used for the ordinal patterns and of the thresholding method.

We have found that an increase in the ordinal pattern spacing (therefore an in-
crease in the timescale), generates a growth of the connectivity on the equatorial Pa-
cific area (figures 2–5, left column, rows 2–4). We have also found that this increase
in connectivity is associated with an increase on the teleconnections from points sit-
uated on this area. (right column of the same figures). This result is consistent with
previous work [24].

We also observe that the number of significant links is smaller when the ordinal
patterns are constructed with a shorter timescale (by comparing consecutive months)
and interpret this effect as due to more stochasticity in the time-series in the short
timescale. We find that, when the networks are set at a fixed link density (figures 4,
5) the networks – constructed with 6 and with 24 bins respectively – are remarkably
similar. We argue that 24 bins show a tradeoff of resolution versus data length, equiv-
alent to the 6 bins MI, but more flexible in its essence as its statistical significance can
be increased by the use of filtering techniques, better estimators or longer time series.
Moreover, our results indicate that the significance of the links in climate networks
should be carefully examined in order to avoid disregarding weak but significant links.

A main goal of our work was to try to determine an optimal thresholding method-
ology, resulting in climate networks that i) contain only truly relevant connections
(and the links that represent random correlations are filtered) and ii) do not disregard
the weak links that are significant (i.e., the links representing statistically significant
deviations from random correlations are not filtered). While in [24] the significance
threshold used was the maximum MI value obtained from shuffled time-series, here
we considered a more "tolerant" threshold (the mean value plus 3 standard devia-
tions of the MI distribution) that results in networks with a higher number of links,
as compared to those in [24]. While in this way we risk including links that are not

significant, the networks obtained are consistent with those in [24] and we found that, by increasing the number of links, we obtained networks that display in more detail the complexity of the atmospheric teleconnections. Another conclusion of our work is that, since Figs. 4 and 5 are very similar, for a fixed link density the main features of the network are independent of the bin number used for computing the mutual information, which confirms the robustness of the climate networks constructed with this thresholding methodology.

Acknowledgments

References

1. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang, Phys. Rep., **424**, 4 (2006)
2. R. Albert, A. L. Barabasi, Phys. Rev. Lett., **85**, 24 (2000)
3. A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, C. S. Zhou, Phys. Rep., **469**, 3 (2008)
4. A. Barrat, M. Barthèlemy, A. Vespignani, Phys. Rev. Lett., **92**, 228701 (2004)
5. S. Bialonski, M.T. Horstmann, K. Lehnertz, Chaos, **20**, 013134 (2010)
6. A.A. Tsonis, P.J. Roebber, Physica A, **333**, 497–504 (2004)
7. A.A. Tsonis, K. L. Swanson, P. J. Roebber, B. Am. Meteorol. Soc., **87**, 585–595 (2006)
8. K. Steinhaeuser, N. V. Chawla, A. R. Ganguly, Stat. Anal. Data Mining, **5**, 4 (2011).
9. M. Paluš, D. Hartman, J. Hlinka, M. Vejmelka, Nonlinear Proc. Geoph., **18**, 751 (2011)
10. Y. Zou, J. F. Donges and J. Kurths, Complex Systems and Complexity Science **8**, 27–38 (2011)
11. N. Marwan, J.F. Donges, Y. Zou, R.V. Donner, J. Kurths, Phys. Lett. A, **373**, 4246–4254 (2009)
12. R.V. Donner, J. F. Donges, Acta Geophys., **1**, 35 (2012)
13. J. Heitzig, J. F. Donges, Y. Zou, N. Marwan, J. Kurths, Eur. Phys. J. B, **85**, 1–22 (2012)
14. J. F. Donges, Y. Zou, N. Marwan, J. Kurths, Eur. Phys. J. Spec. Top., **174**, 157–179 (2009)
15. J. F. Donges, Y. Zou, N. Marwan, J. Kurths, Europhys. Lett. **87**, 48007 (2009)
16. J. F. Donges, H.C.H. Schultz, N. Marwan, Y. Zou, J. Kurths, Eur. Phys. J. B, **84**, 635–651 (2011)
17. A. Gozolchiani, S. Havlin, K. Yamasaki, Phys. Rev, Lett., **107**, 148501 (2011)
18. N. Malik, B. Bookhagen, N. Marwan, J. Kurths, Clim. Dynam., **39**, 3 (2012)
19. A. A. Tsonis and K. L. Swanson, Phys Rev. Lett., **100**, 228502 (2008)
20. K. Yamasaki, A. Gozolchiani, S. Havlin, Phys. Rev. Lett. **100**, 228501 (2008)
21. Y. Berezin, A. Gozolchiani, O. Guez, S. Havlin Sci. Rep., **2**, 666 (2012)
22. O. Guez, A. Gozolchiani, Y. Berezin, S. Brenner, S. Havlin, Europhys. Lett. **98**, 3 (2012)
23. A. A. Tsonis, K. Swanson, S. Kravtsov, Geophys. Res. Lett., **34**, 13 (2007)
24. M. Barreiro, A.C. Marti, C. Masoller, Chaos, **21**, 201 (2011)
25. C. Bandt, B. Pompe, Phys. Rev. Lett., **88**, 174102 (2002)
26. C. Yinhe, W. Tung, J. B. Gao, V. A. Protopopescu, L. M. Hively, Phys. Rev. E, **70**, 4 (2004)
27. 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, D. Joseph, B. Am. Meteorol. Soc., **77**, 437 (1996)
28. A. Kraskov, H.Stögbauer, P. Grassberger, Phys. Rev. E, **69**, 6 (2004)
29. C.J. Cellucci, A.M. Albano, P.E. Rapp, Phys. Rev. E, **71**, 6 (2005)

30. S. Khan, S. Bandyopadhyay, A.R. Ganguly, S. Saigal, D.J. Erickson III, V. Pro-
    topopescu, G. Ostrouchov, Phys. Rev. E, **76**, 026209 (2007)
31. M. Paluš, Contemp. Phys., **48**, 6 (2007)
32. J.C.H. Chiang, A.H. Sobel, J. Climate, **15**, 2616-2631 (2002)
33. M. Barreiro, P. Chang, L. Ji, R. Saravanan, A. Giannini, Dynam. Atmos. Oceans, **39**,
    1 (2005)
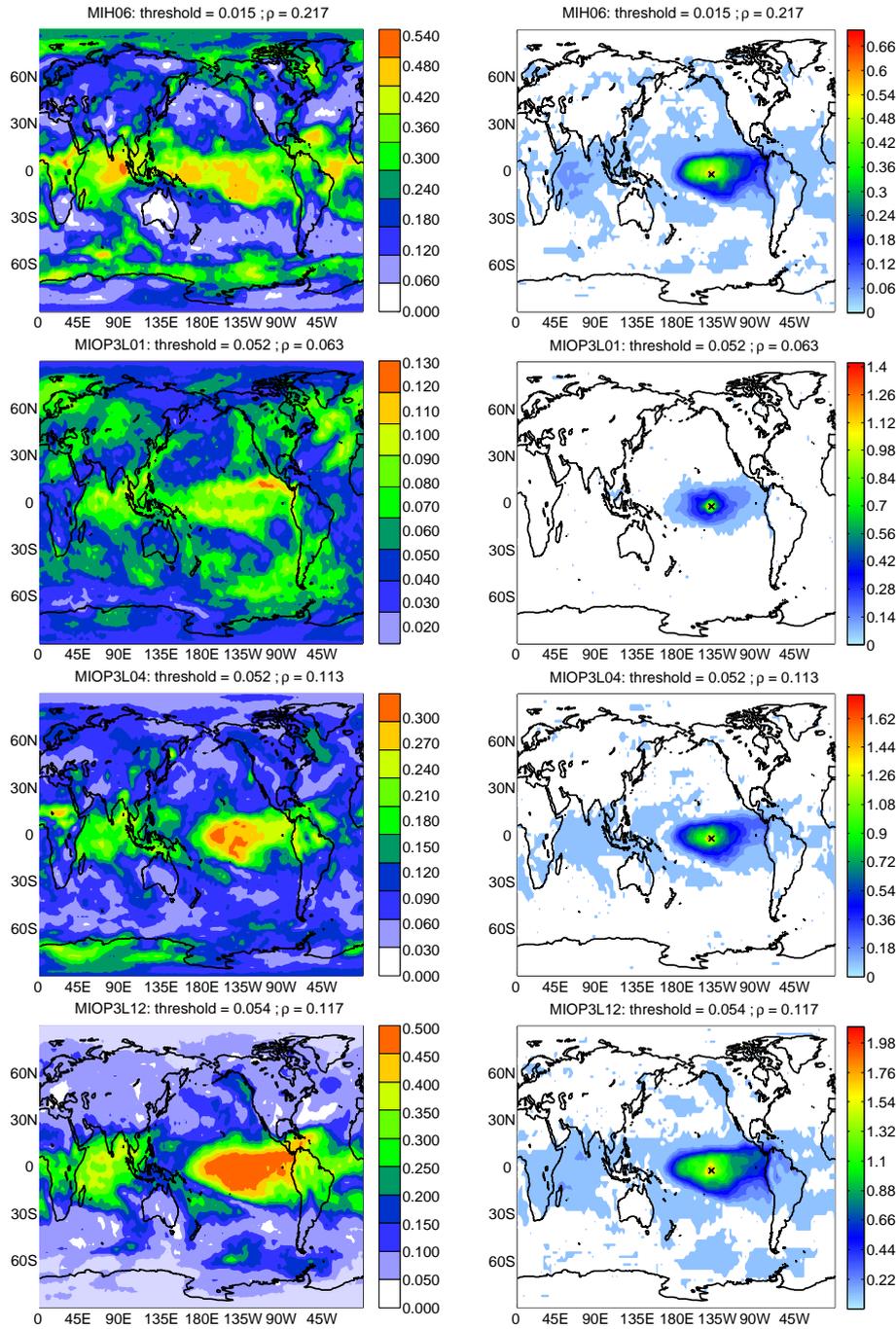34. Q. Ding, E.J. Steig, D.S. Battisti, M. Küttel, Nature Geosci., **4**, 6 (2011)

Fig. 2. AWC (left column) and connectivity maps (right column) using the different methods of network construction described in the text, computing the probabilities with 6 bins and using the significance threshold $\tau = \mu + 3\sigma$ to define the links. The methods are: histograms of anomaly values MIH (top row); OPs formed with three consecutive months (second row); OPs formed with three equally spaced months covering a one-year period (third row); and OPs formed with three months in consecutive years (bottom row).
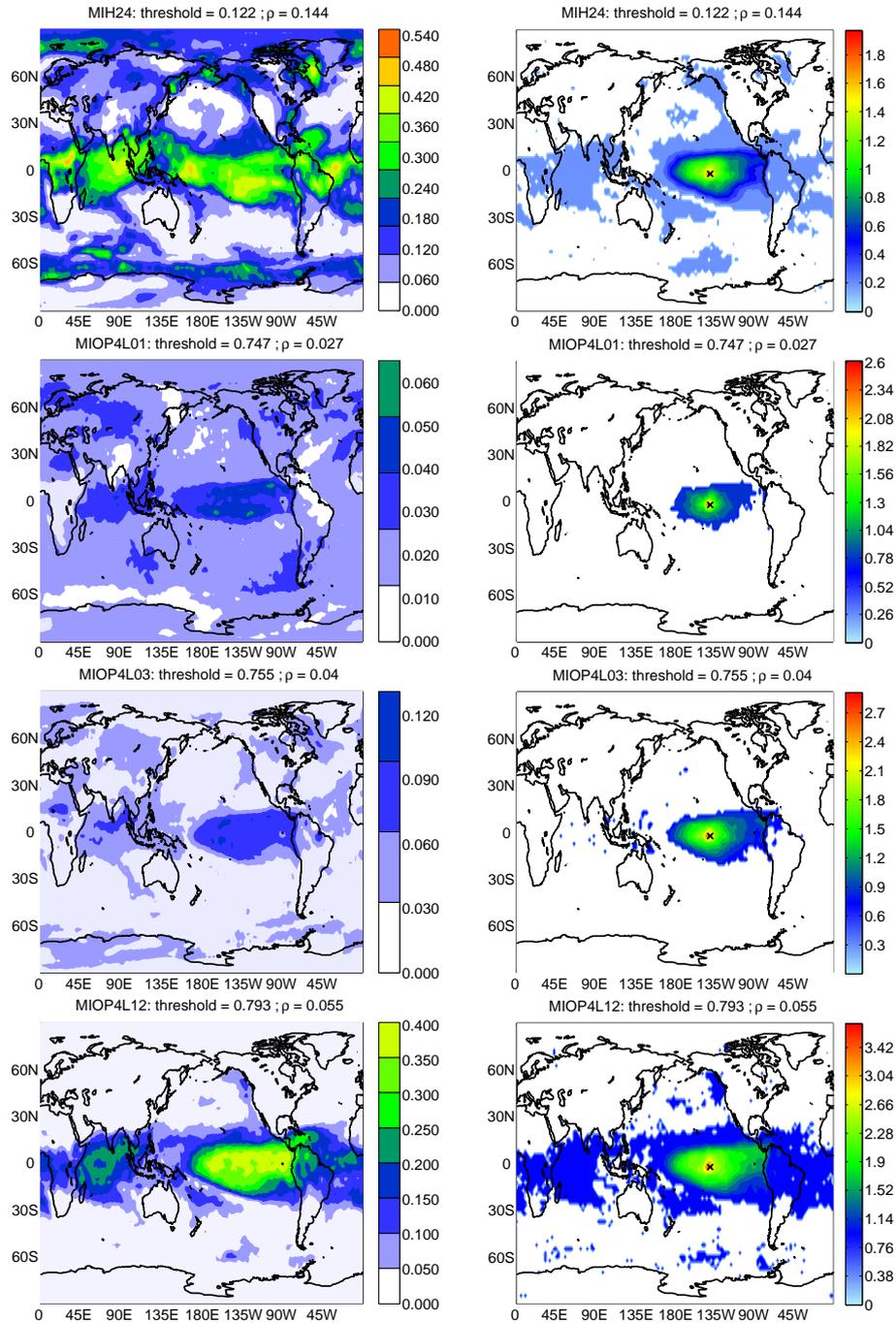
Fig. 3. As Fig. 2 but the probabilities are calculated with 24 bins: histograms of anomaly values MIH (top row); OPs formed with four consecutive months (second row); OPs formed with four equally spaced months covering a one-year period (third row); and OPs formed with four months in consecutive years (bottom row). In order to better compare the AWC plots, we have used the same color scale as in the left column in Fig. 2.
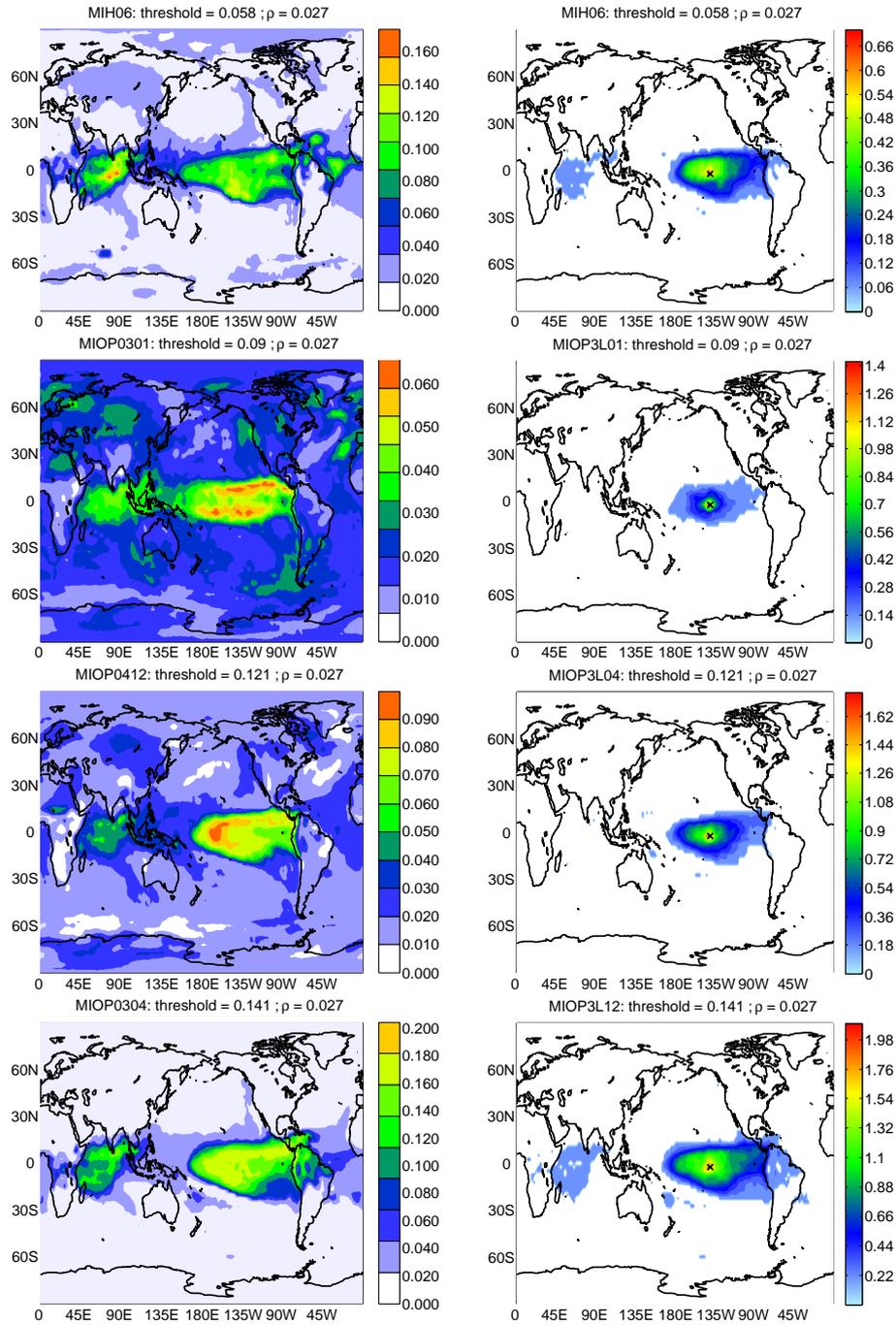
Fig. 4. As Fig.2 but with the threshold τ chosen such that the networks have the same link density (0.03%).

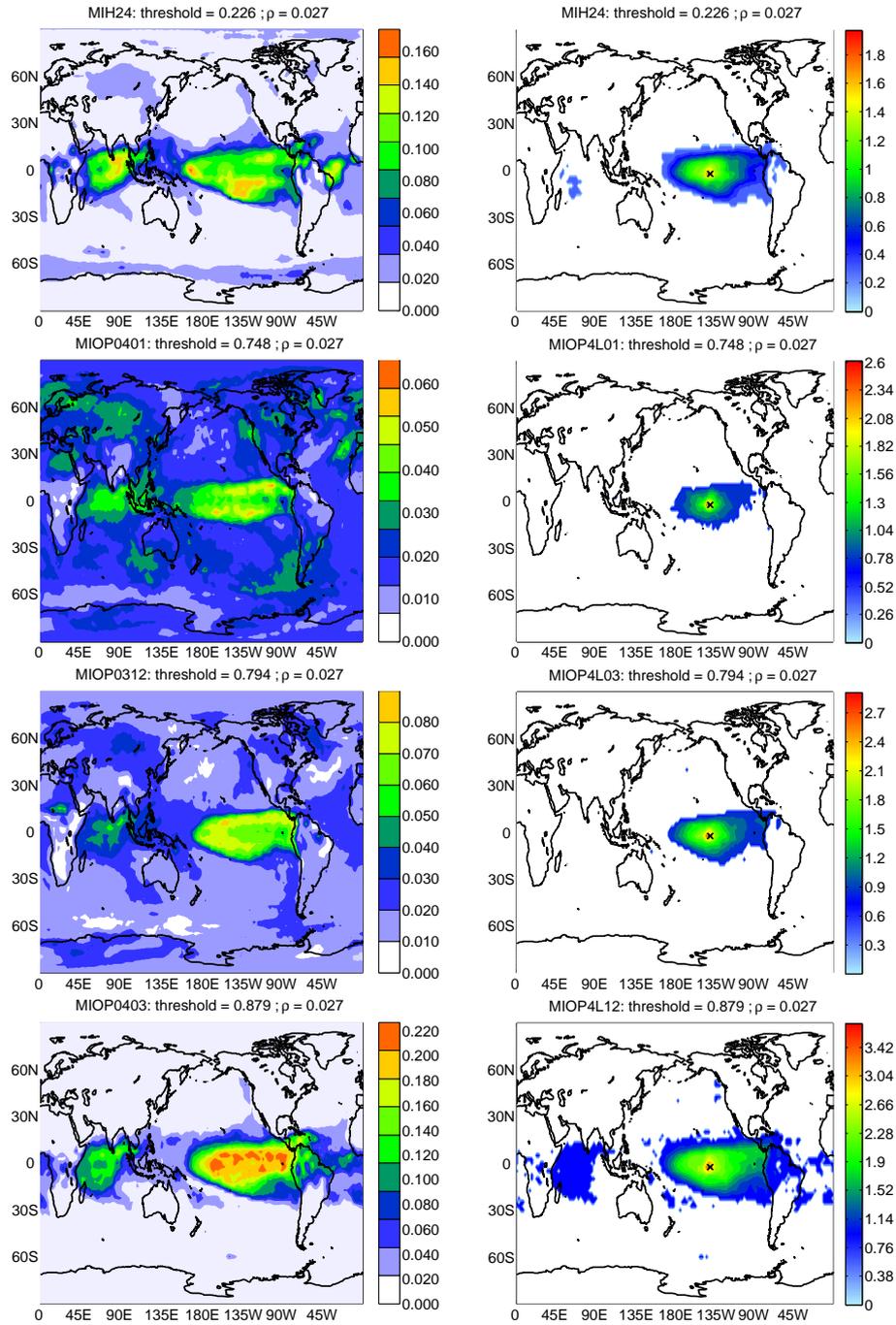**Fig. 5.** As Fig.3 but with the threshold τ chosen such that the networks have the same link density (0.03%). In order to better compare the AWC plots, we have used the same color scale as in the left column in Fig. 4.