

Discerning connectivity from dynamics in climate networks

M. Paluš¹, D. Hartman¹, J. Hlinka¹, and M. Vejmelka¹

¹Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2,
182 07 Prague 8, Czech Republic

Received: 1 August 2011 – Revised: 18 October 2011 – Accepted: 20 October 2011 – Published: 24 October 2011

Abstract. The bias due to dynamical memory (serial correlations) in an association/dependence measure (absolute cross-correlation) is demonstrated in model data and identified in time series of meteorological variables used for construction of climate networks. Accounting for such bias in inferring links of the climate network markedly changes the network topology and allows to observe previously hidden phenomena in climate network evolution.

1 Introduction

Statistical physics met graph theory in order to describe structure and function of complex systems comprising a large number of interacting elements. Such complex networks (Boccaletti et al., 2006) are increasingly used as a model and analysis paradigm for multivariate time series recorded from complex, potentially nonlinear geophysical processes including the Earth's atmosphere and climate. Climate networks (Tsonis and Roebber, 2004; Tsonis et al., 2006) are usually, although not exclusively, constructed using gridded time series of meteorological variables preprocessed, e.g. in the NCEP/NCAR reanalysis project (Kalnay et al., 1996). Monthly (Tsonis and Swanson, 2008; Donges et al., 2009a) or daily (Yamasaki et al., 2008; Gozolchiani et al., 2008) surface air temperature data are frequently used, however, equipotential heights (Tsonis et al., 2008; Donges et al., 2011), sea surface temperature, humidity, precipitation and related data (Steinhaeuser et al., 2011; Malik et al., 2011) and other meteorological data are also analysed. Individual grid-points, characterised by time series of a chosen meteorological variable, are considered as nodes (vertices) of the climate network, while links (edges) are inferred from some,

mostly statistical association between the time series related to the two nodes at the edge's end-points. The most common association measure is Pearson correlation (Tsonis and Roebber, 2004; Tsonis and Swanson, 2008), however, also more general, nonlinear measures are tested, e.g. measures derived from information theory (Donges et al., 2009b; Barreiro et al., 2011), or phase synchronisation analysis (Yamasaki et al., 2009). The influence of the choice of an association measure on the topology of the climate network has been tested (Donges et al., 2009a,b), however, one important point hitherto has been neglected. In this study we demonstrate how dynamics, or temporal complexity of time series influences association measures such as absolute correlations. Higher regularity or dynamical memory causes an upward bias in association measures. We demonstrate that accounting for such bias leads to dramatic changes in the topology of climate networks, in particular, the role of the North Atlantic Oscillation in the connectivity of global climate networks is sharply increased at the cost of the role of the El Niño Southern Oscillation.

2 Dynamics and connectivity

Consider complex, dynamic processes $\{Y_i(t)\}$ evolving in time t . For simplifying the notation we will drop the index i in general considerations about a process; and use it only when we will need to distinguish processes, say Y_1 and Y_2 . A series of measurements done on such a process in consecutive instants of time $t = 1, 2, \dots$ is usually called a time series $\{y(t)\}$.

Let the dynamics at two nodes of a network be characterised by dynamic processes $\{Y_1(t)\}$ and $\{Y_2(t)\}$, respectively. The existence of a link between the two nodes is inferred from a measure of dependence $d(Y_1, Y_2)$ which characterises coordination or, in a special case, synchronisation between the processes $\{Y_1(t)\}$ and $\{Y_2(t)\}$. In the network



Correspondence to: M. Paluš
(mp@cs.cas.cz)

context such dependence measures are also called connectivity measures. The most common measure of dependence, coming from statistics, is the linear (Pearson) correlation:

For N observations $y_1(t)$, $y_2(t)$, $t = 1, \dots, N$, we define means

$$\bar{y}_i = \frac{1}{N} \sum_{t=1}^N y_i(t) \quad (1)$$

and standard deviations

$$\sigma_i^2 = \frac{1}{N-1} \sum_{t=1}^N (y_i(t) - \bar{y}_i)^2 \quad (2)$$

and compute the normalised observations

$$\tilde{y}_i(t) = \frac{y_i(t) - \bar{y}_i}{\sigma_i}. \quad (3)$$

The correlation coefficient between $\{y_1(t)\}$ and $\{y_2(t)\}$ is

$$c(Y_1, Y_2) = \frac{1}{N-1} \sum_{t=1}^N \tilde{y}_1(t) \tilde{y}_2(t). \quad (4)$$

Typically, for independent processes or systems $\{Y_1(t)\}$ and $\{Y_2(t)\}$, the correlation coefficient $c(Y_1, Y_2)$, or a general dependence measure $d(Y_1, Y_2)$ is equal to zero which means that there is no link between the two nodes represented by $\{Y_1(t)\}$ and $\{Y_2(t)\}$. For non-zero $d(Y_1, Y_2)$ some dependence, coordination or even synchronisation between the considered nodes exists and we will speak about *connectivity*. In this study we will not investigate the form of the observed connectivity, nor try to discern whether it is direct or indirect (Nawrath et al., 2010).

In practical computations using experimental data, estimates of connectivity measures are always different from zero. Then, it is necessary to decide whether the digression from zero of a connectivity measure is due to random effects or due to real connections between the studied processes. The answer to such questions can be obtained using ideas and tools of mathematical statistics. The term *statistical significance* is used for values of a statistical quantity (a connectivity measure in our case) which differ from a value specified by a so-called *null hypothesis* (independence, in our case) more than expected due to random effects. For more details and explanation of statistical testing see Paluš (2007) and references therein.

Some researchers, using the correlation coefficient to establish connectivity or dependence in general, use *critical values* for significance of correlations given in a number of statistical handbooks. Such tables of critical values for the correlation coefficient, however, are computed for the null hypothesis of independent, identically distributed (IID) processes, i.e. for observations $y_1(t)$, $y_2(t)$ which are realisations of white noise processes. A well-known example of such incorrect application of the correlation significance tables is the evaluation of correlations between sunspot numbers and the numbers of Republicans in the US Senate. Paluš

(2007) analyzes this example in detail, showing that the temporal dynamics of correlated data (the oscillatory behaviour of the sunspot numbers and also of the numbers of the Republican senators in 1960's–1980's) is the reason of the increased correlation coefficient, not any real dependence between these datasets.

2.1 Characterisation of dynamics by entropy rates

In order to understand the behaviour of dependence measures with respect to the dynamics of studied processes, let us consider that the temporal evolution of the studied process or system is not random, but that the state of the system at time t depends on the state in which the system was at time $t - \tau$. The strength of such a dependence per unit time delay τ , or, inversely, a rate at which the system “forgets” information about its previous states, or, in other words, generates uncertainty about its future, can provide an important quantitative characterisation of temporal complexity in the system's evolution.

A time series $\{y(t)\}$, which is a recording of (a part of) the temporal evolution of a system or process under study, is considered as a realisation of a stochastic process, i.e. an indexed sequence of random variables, characterised by the joint probability distribution function $p(y(1), \dots, y(n))$. Uncertainty in a stochastic variable is measured by its entropy. The rate at which the stochastic process “produces” uncertainty is measured by its entropy rate

$$h = \lim_{n \rightarrow \infty} \frac{1}{n} H(Y(1), \dots, Y(n)), \quad (5)$$

where $H(Y(1), \dots, Y(n))$ is the joint entropy of the n variables $Y(1), \dots, Y(n)$, see Paluš (1996) for details. The concept of entropy rates is common to the theory of stochastic processes as well as to information theory where the entropy rates are used to characterise information production by information sources (Cover and Thomas, 1991).

Alternatively, we can consider a network of coupled dynamical systems or oscillators. Then the time series $\{y(t)\}$ is considered as a projection of a trajectory of a dynamical system, evolving in some measurable state space. A. N. Kolmogorov, who introduced the theoretical concept of classification of dynamical systems by information rates, was inspired by information theory and generalized the notion of the entropy of an information source. The Kolmogorov-Sinai entropy is a topological invariant, suitable for the classification of dynamical systems or their states, and is related to the sum of the system's positive Lyapunov exponents (see Paluš (1996) and references therein). Thus, the concept of entropy rates is common to theories based on philosophically opposite assumptions (randomness vs. determinism) and is ideally applicable for the characterisation of complex geophysical processes, where possibly deterministic rules are always accompanied by random influences.

Let us consider in the following that $\{Y(t)\}$ is a zero-mean stationary Gaussian process with spectral density function

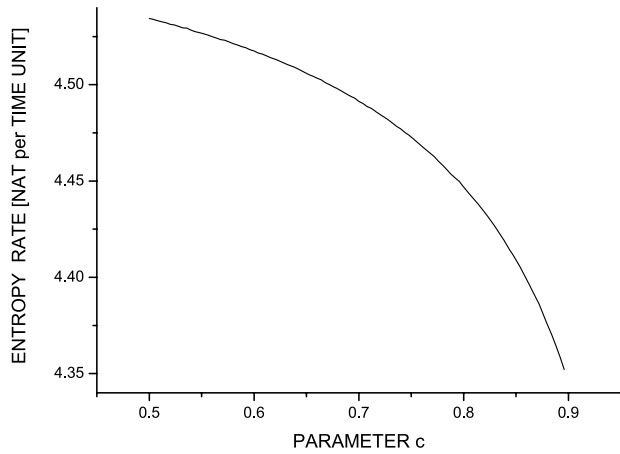


Fig. 1. Dependence of the entropy rate h_G for the autoregressive process (7) on the parameter c .

$f(\omega)$, where ω is a normalised frequency. Then its entropy rate h_G , apart from a constant term, can be expressed using $f(\omega)$ (see Paluš (1997) and references therein) as

$$h_G = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\omega) d\omega. \quad (6)$$

2.2 Dynamics and bias in connectivity measures

Let us consider the autoregressive process (ARP)

$$y_t = c \sum_{k=1}^{10} a_k y_{t-k} + \sigma e_t, \quad (7)$$

where $a_{k=1,\dots,10} = 0, 0, 0, 0, 0, .19, .2, .2, .2, .2$, $\sigma = 0.01$ and e_t are Gaussian deviates with zero mean and unit variance. For $c = 1$ this ARP has a long coherence time, for $c < 1$ the coherence time decreases and the entropy rate increases (Paluš, 1996). In particular, we can generate realisations of the ARP (7) with different values of c and thus with different entropy rates. The entropy rates of such ARP's monotonically decrease with increasing c . Figure 1 presents the entropy rates h_G for 100 ARP's with c increasing from 0.5 to 0.9. Now, let us study the distribution of cross-correlations between *independent* realisations of the process (7) for different values of the parameter c . For each c we generate 8192 process realisations, each realisation consisting of 16 384 samples. Figure 2 presents the histograms of cross-correlations between independent realisations of ARP (7) for three different values of c . The mean value is always correctly equal to zero, however, the variance increases with increasing c , i.e. with decreasing entropy rate. As a consequence, when considering *absolute* correlations, or any non-negative dependence measure, e.g. mutual information (Paluš et al., 1993; Paluš, 1995), its mean value has an increasing upward bias with decreasing entropy rates, i.e. with increasing dynamical memory of the studied processes (Fig. 3).

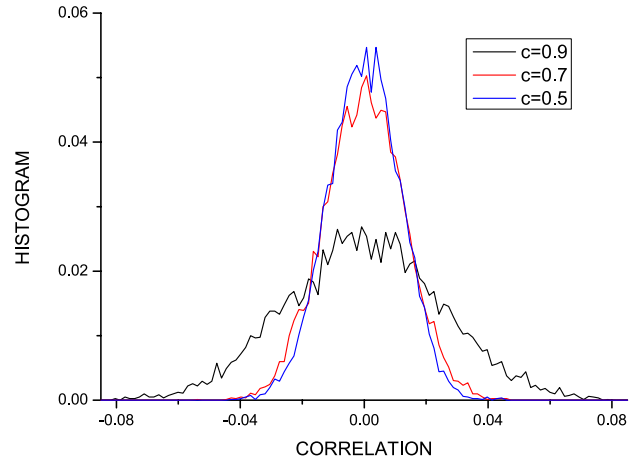


Fig. 2. Histograms of cross-correlations between independent realisations of the autoregressive process (7) for three different values of the parameter c .

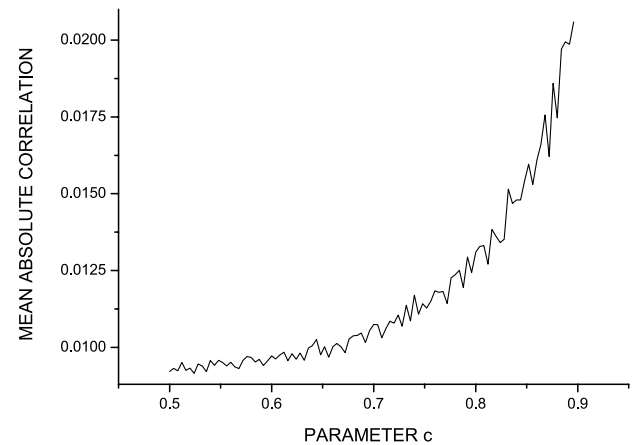


Fig. 3. Dependence of the mean absolute cross-correlation between independent realisations of the autoregressive process (7) on the parameter c .

2.3 Coping with dynamical bias: surrogate data

The above example demonstrates that the values of connectivity measures obtained from experimental data are only relative and the true connectivity can only be inferred in a proper statistical testing procedure. For experimental time series with arbitrary dynamics no analytical treatment is possible, but we can generate an empirical distribution of the used connectivity measure for independent realisations of so-called *surrogate data*. In general we can say that the surrogate data are numerically generated data which preserve all important statistical properties of the original data but the property which is tested for. A more realistic definition for practical purposes requires that the surrogate data replicate those data properties which influence the discriminating

statistic and are the primary source of its bias and variance irrespectively of the presence of the tested property. Of course, the surrogate data should not possess the tested property which is now the dependence between the processes $\{Y_1(t)\}$ and $\{Y_2(t)\}$. The dynamics, in particular, the entropy rates of the processes should be preserved. If a Gaussian process is a suitable model for the analysed data, it is satisfactory to generate independent realisations of a process with the same spectrum as the original data. This can be accomplished using Fourier transform (FT) surrogates which are constructed by computing the fast Fourier transform (FFT) of a time series, randomising the phase of each frequency component and then taking the inverse FFT. For details see Schreiber and Schmitz (2000); Paluš (2007) and references therein.

3 Dynamics and connectivity in climate networks

3.1 Data

In the following study we use monthly mean values of the near-Surface Air Temperature (SAT) from the NCEP/NCAR reanalysis data (Kalnay et al., 1996). We include the data up to the latitudes 87.5° in the grid of $2.5^\circ \times 2.5^\circ$ which leads to 10 224 grid points or network nodes. Temporal segments of 512 months starting in January 1958 enter the FFT for generating surrogate data, however, correlations are computed from 500 samples after discarding 6 points at each end of time series. Due to the nonstationary character of the data, shifting the chosen temporal segment through the available data (1948–2010), subtle local changes can be observed in the presented maps of network characteristics. However, the overall picture and the findings of this study are not affected. As the only pre-processing of the data, the annual cycle was removed by subtracting the mean values for each month in the year.

Since we will study possible effects of solar activity and large-scale atmospheric circulation phenomena (North Atlantic Oscillation, NAO; and El Niño Southern Oscillation, ENSO) on climate networks, the following data will also be used: the monthly NAO index, available with its description at <http://www.cru.uea.ac.uk/cru/data/nao/>; the SOI index, available at <http://www.cru.uea.ac.uk/cru/data/soi/>, and the solar 10.7 cm radio flux monthly means obtained from http://www.ukssdc.ac.uk/data_menu.html.

3.2 The network

Using the above SAT anomalies (i.e. digressions from the seasonal means, SATA in the following) we compute correlation coefficients $c_{i,j}$ for each pair of nodes $i, j = 1, \dots, N_N = 10\,224$. We use the matrix of absolute correlations $C_{i,j} = |c_{i,j}|$ in order to obtain the adjacency matrix $A_{i,j}$ of the binary network, defined as: $A_{i,j} = 1$ iff $C_{i,j} > c_T$, otherwise $A_{i,j} = 0$. $A_{i,i} = 0$ by definition.

The basic characterisation of connectivity of a node i is its degree, or degree centrality

$$k_i = \sum_{j=1}^{N_N} A_{i,j}, \quad (8)$$

giving the number of nodes to which the node i is connected. In the context of climate networks, a related quantity is defined, known as the area weighted connectivity

$$AWC_i = \frac{\sum_{j=1}^{N_N} A_{i,j} \cos(\lambda_j)}{\sum_{j=1}^{N_N} \cos(\lambda_j)}. \quad (9)$$

Since the reanalysis data are defined on a grid which is regular in angular coordinates, the geographic distances of the grid points depend on latitude λ_j . AWC corrects for this dependence and can be interpreted as the fraction of the Earth's surface area a vertex is connected to (Tsonis et al., 2006).

From the multitude of network characteristics (Boccaletti et al., 2006; Hartman et al., 2011), in this study we will use the key mesoscopic measure known as *betweenness centrality* (also called the *shortest path betweenness centrality*) defined as

$$BC_i = \sum_{j \neq i \neq k; j,k=1}^{N_N} \frac{\sigma_{j,k}(i)}{\sigma_{j,k}}, \quad (10)$$

where $\sigma_{j,k}$ denotes the number of all shortest paths between vertices j and k and $\sigma_{j,k}(i)$ denotes the number of the shortest paths between the same vertices that go through the vertex i (see Hartman et al. (2011) and references therein). This measure is mesoscopic in the sense that it locally characterises a property emerging in large-scale connectivity patterns. In particular, it expresses how important the characterised node is from a perspective of “information flow” from one node to any other node. In other words, how important this node is as a mediator between any pair of nodes.

3.3 Connectivity or dynamics?

We start the SATA climate network analysis in the line of Donges et al. (2009a): In converting the absolute correlations $C_{i,j}$ into the binary adjacency matrix $A_{i,j}$ we establish the threshold c_T so that we preserve only the “strongest” 0.5% of all possible links. We say that the network (edge) density is $\varrho = 0.005$, which corresponds to $c_T = 0.74515$. The related geographical distribution of AWC is mapped in Fig. 4a. The resulting pattern of the highest apparent connectivity in the tropical oceanic, particularly in the tropical Pacific areas has been obtained in this way not only for the absolute correlations, but also for mutual information (Donges et al., 2009a) and also for mutual information of higher-order symbolic sequences (Barreiro et al., 2011). Although several authors proposed an explanation of this tropical connectivity patterns by the large-scale atmospheric phenomenon of El Niño (Tsonis and Swanson, 2008; Yamasaki et al., 2008;

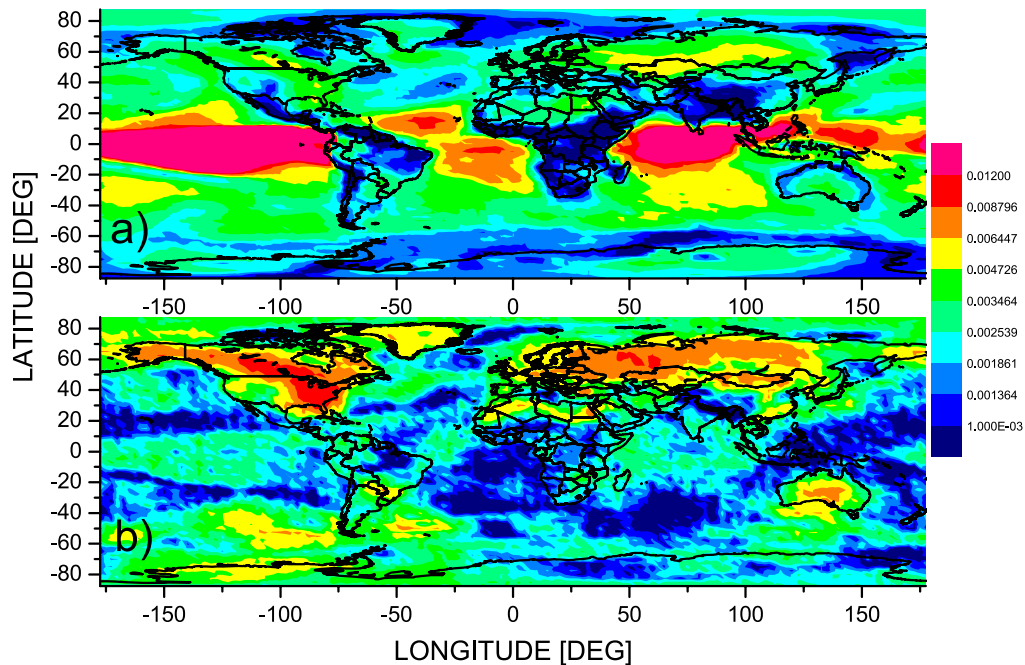


Fig. 4. Area weighted connectivity for SATA climate networks with density $\rho = 0.005$ obtained by uniform thresholding of the absolute correlations (a) and of the absolute correlation Z -scores related to independent univariate FT surrogate data (b).

Gozolchiani et al., 2008), we can ask whether the absolute correlation or mutual information provide an unbiased estimation of the climate network connectivity. Let us analyse the dynamics of the SATA time series using the entropy rate h_G defined in Eq. (6). The entropy rate h_G exactly characterises the dynamics of linear Gaussian processes. A more general version of a nonlinear entropy rate has been used for air temperature records by von Bloh et al. (2005), however, their data included the annual cycle. We consider the entropy rate h_G suitable for this study, since the instrumental air temperature records do not exhibit substantial nonlinearity (Paluš and Novotna, 1994). Moreover, the entropy rate h_G is able to distinguish different dynamical regimes of nonlinear systems (Paluš, 1997).

The entropy rate h_G has been computed for each node and is mapped in Fig. 5a. We can see that the lowest entropy rates and thus the strongest dynamical memories can be found in the SAT anomalies from the tropical, especially from the tropical Pacific areas. Moreover, we can compute the mean absolute correlations of the SATA time series in each node with independent realisations of a Gaussian process with the same entropy rate, realised as independent, univariate FT surrogate data for each node's SATA time series (Fig. 5b). Again we can see the strong correspondence between the entropy rate of a process and the bias in its absolute correlations with similar processes, as in Figs. 1 and 3. In order to better understand the effect of dynamics and, in our case, of the geographical position of the correlated nodes, we choose two pairs of nodes and compute histograms of cross-correlations

of their independent FT surrogate data. Figure 6 presents surrogate cross-correlation histograms for a pair of nodes from the low entropy rate area (orange and red lines, nodes with the latitude 0° , longitude 90° W and 10° S, 120° W) and a pair from the high entropy rate area (green and blue lines, nodes 60° N, 25° E and 60° N, 75° E). Considering the SATA time series from the low entropy rate (tropical) areas, the histogram is wider, giving a higher probability for a random occurrence of large cross-correlation values even without any real connectivity (dependence). The green and orange lines illustrate histograms obtained from the simple FT surrogate data, while the blue and red histograms were obtained using the amplitude-adjusted (AA) FT surrogate data. The AAFT surrogates are used to cope with non-Gaussianity of the analysed data (Schreiber and Schmitz, 2000). We can see that the possible effect of non-Gaussianity is negligible in comparison with the effect of different dynamics, reflected by the entropy rates.

In explanation of the apparently high connectivity in the tropics induced by the El Niño phenomenon, Tsonis and Swanson (2008) studied the climate networks for different types of tropical atmospheric dynamics in so called El Niño and La Niña periods. These can be defined by the values of the Southern Oscillation Index (SOI, Sect. 3.1). In this study we have divided the sorted SOI values into three equally populated bins. The bin around the median was discarded, while the values in the bin over the median define the positive SOI (ENSO+) episodes, and the bin under the median define the negative SOI (ENSO-) episodes. Now we can construct the

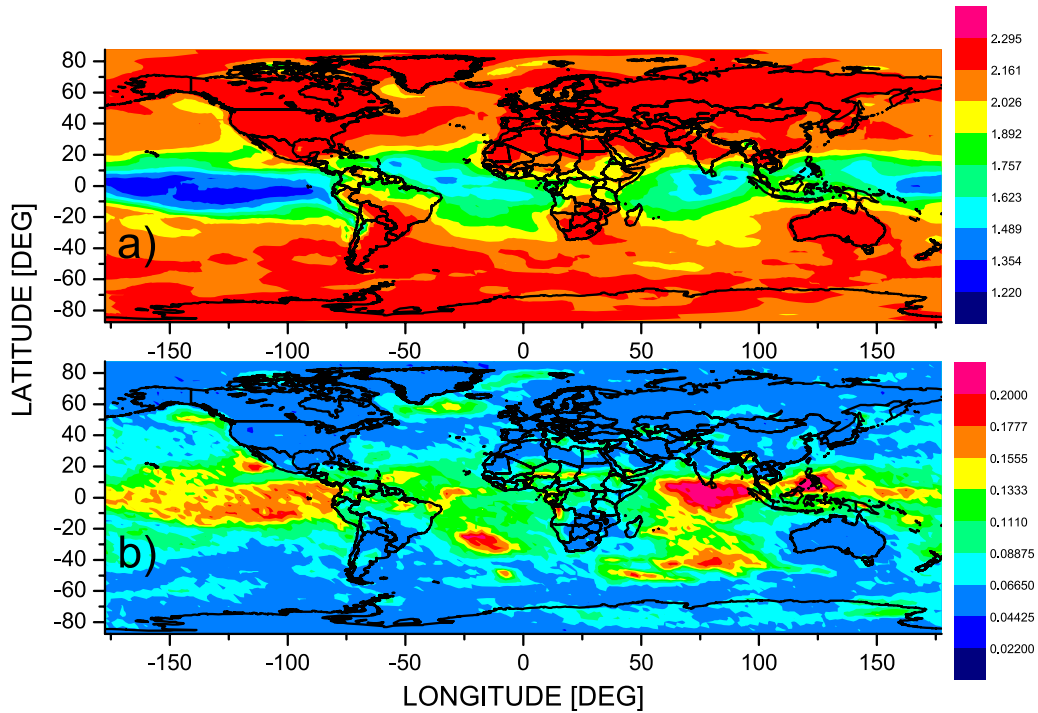


Fig. 5. SATA signal complexity measured by the Gaussian process entropy rate h_G (a). Mean absolute correlations of SATA with their independent univariate FT surrogate data (b).

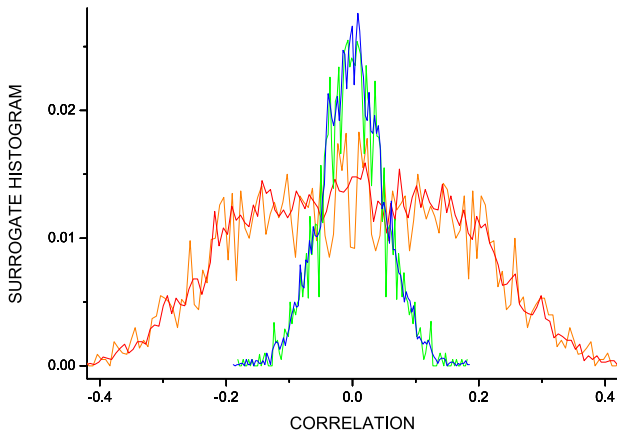


Fig. 6. Histograms of cross-correlations of independent FT (green and orange lines) and amplitude-adjusted FT (red and blue lines) surrogate data for SATA of a pair of nodes from the low entropy rate area (orange and red lines, nodes with the latitude 0° , longitude 90° W and 10° S, 120° W) and a pair from the high entropy rate area (green and blue lines, nodes 60° N, 25° E and 60° N, 75° E).

networks for the ENSO– and ENSO+ episodes and present related AWC geographical distributions, however, no generally accepted methodology of network comparison has been established yet. We can either use the approach of Donges et al. (2009a) and require $\varrho = 0.005$ for both the conditions,

which leads to thresholds $c_T = 0.74052$ and $c_T = 0.81235$ for the ENSO– and ENSO+ episodes, respectively (Fig. 7). Or, using the approach of Tsonis and Swanson (2008) we put $c_T = 0.5$ in both the cases which leads to $\varrho = 0.0166$ and $\varrho = 0.0229$ for the ENSO– and ENSO+ episodes, respectively (Fig. 8). Consequently, we can see network-construction-related differences between Figs. 7 and 8, however, in a relative sense they are similar and the main message is the same: During the ENSO– episodes the strong AWC is confined to the tropical Pacific region, while during the ENSO+ episodes the strong AWC regions extend to all tropical, mainly oceanic areas. In this conclusion we reproduce the result of Tsonis and Swanson (2008), although our definition of ENSO± periods is not exactly the same.

Yet this result is not the answer whether AWC of the climate network obtained by thresholding the absolute correlations reflects the unbiased connectivity of the climate network. Unfortunately we cannot reliably estimate the entropy rates for the number of short segments of various lengths related to the ENSO± episodes, however, in the same way as we can compute the absolute cross-correlations between the nodes, we can compute for each node the mean absolute cross-correlations of the SATA series with its independent FT surrogates. For this purpose, we construct the surrogate data using the whole analysed segment and consider only cross-correlations between the observational and surrogate values falling into the respective ENSO± episodes. These measures

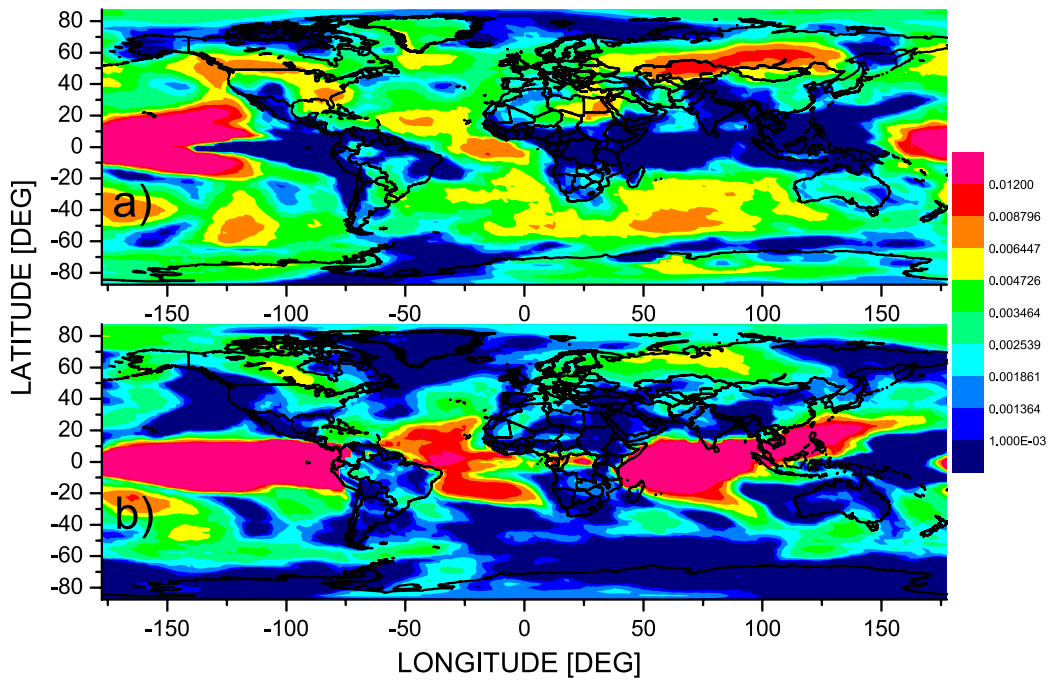


Fig. 7. Area weighted connectivity for SATA climate networks with density $\rho = 0.005$ obtained by uniform thresholding of the absolute correlations for negative (a) and positive (b) ENSO index.

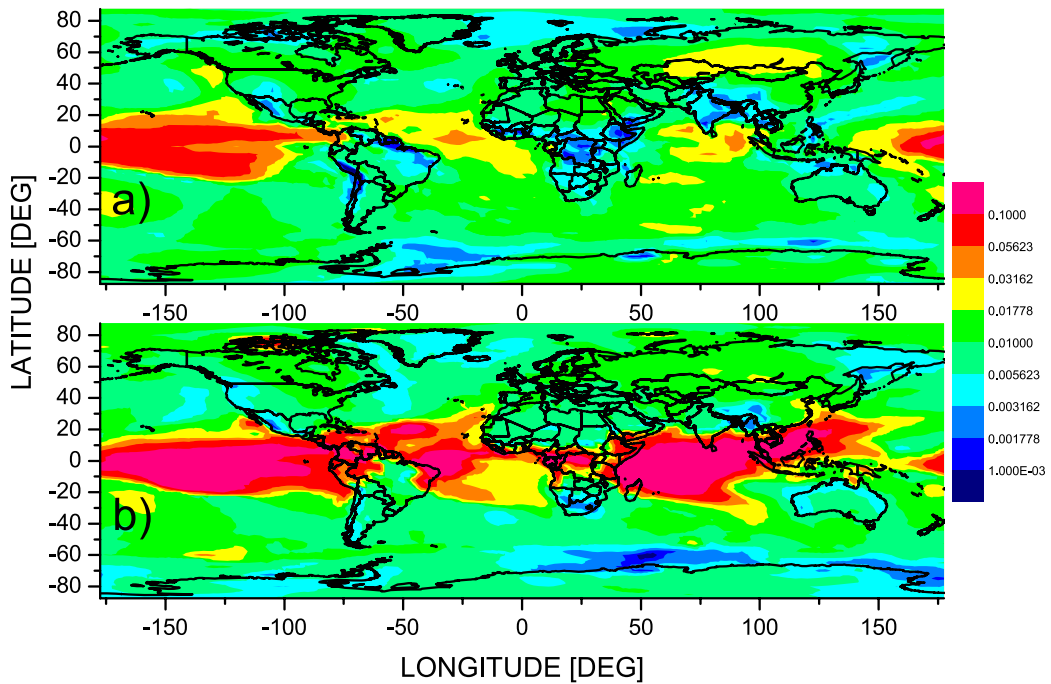


Fig. 8. Area weighted connectivity for SATA climate networks obtained by uniform thresholding of the absolute correlations using the threshold $c_T = 0.5$ for negative (a) and positive (b) ENSO index.

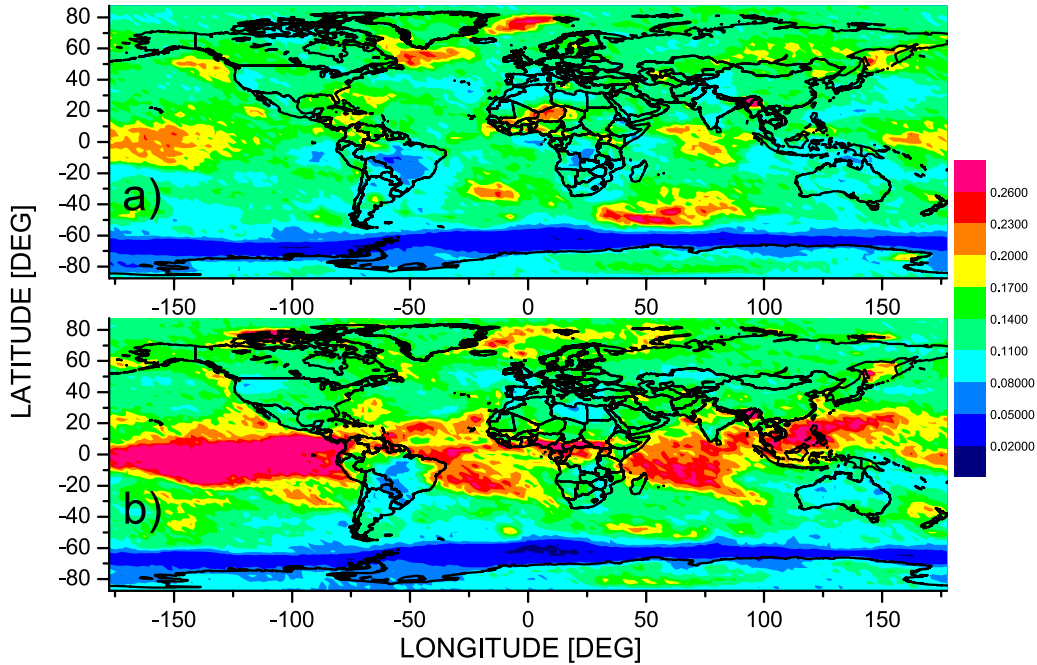


Fig. 9. Mean absolute correlations of SATA with their independent univariate FT surrogate data for negative (a) and positive (b) ENSO index.

of the dynamics-related bias of the absolute correlations for both the ENSO- and ENSO+ conditions are mapped in Fig. 9. Although the patterns in Figs. 7, 8 and 9 are not exactly the same, qualitatively they are very similar. This comparison supports the explanation of the tropical AWC patterns and their dependence on the ENSO mainly by the dynamics of SATA in the tropics, rather than by connectivity in the climate network.

3.4 Connectivity adjusted to dynamics

Let us adjust the values of absolute cross-correlations relatively to the dynamics of particular pairs of SATA time series. For each pair of nodes i, j let us construct one hundred pairs of realisations of independent FFT surrogates, preserving the entropy rate h_G in each respective node, and compute their cross-correlations which have reasonably normal distribution (Fig. 6). We compute the mean absolute cross-correlation $\bar{C}_{i,j}[S]$ of the independent surrogate pairs, as well as their standard deviation $\sigma_{i,j}[S]$. Then, for each absolute cross-correlation $C_{i,j}$ we compute its Z-score as

$$Z_{i,j} = \frac{C_{i,j} - \bar{C}_{i,j}[S]}{\sigma_{i,j}[S]}. \quad (11)$$

Let us construct the binary network adjacency matrix with $\varrho = 0.005$ by thresholding $Z_{i,j}$ instead of $C_{i,j}$. The geographical distribution of AWC for such a network is presented in Fig. 4b. We can see that the patterns of maximum

connectivity are quite different from the AWC map obtained by thresholding $C_{i,j}$ (Fig. 4a). The upward bias in connectivity of the tropics is compensated and the most connected areas are located in the land areas of the Northern Hemisphere (NH). Let us remind that in both cases in Fig. 4 we preserve 0.5% of all possible links. In the upper map we have preserved the links with the largest absolute correlations, while in the lower map we have preserved the links with the most significant correlations. There is no universal critical level for significance – each pair of nodes has its own significance test accounting for the dynamics of each SATA time series by constructing FFT surrogate data for each SATA series individually. Requiring $\varrho = 0.005$ leads to the threshold $Z_T = 18.187$ which is large enough even for accounting for the high multiplicity of statistical tests.

4 North Atlantic Oscillation and solar activity in climate network connectivity

Looking again to the map in Fig. 4b we can see that after accounting for the connectivity bias due to dynamics of SATA time series, the most connected areas moved from tropics, from the influence of the El Niño Southern Oscillation, to NH areas which are known to be affected by the North Atlantic Oscillation (cf. Fig. 4b and Plate 1a in Marshall et al., 2001).

The North Atlantic Oscillation (NAO) is a dominant pattern of atmospheric circulation variability in the extratropical Northern Hemisphere and is a major factor influencing meteorological variables including temperature, precipitation, occurrence of storms, wind strength and direction in the Atlantic sector and surrounding continents. On the global scale, NAO has a climate significance that rivals the Pacific ENSO (Marshall et al., 2001). It is natural to ask what is the influence of NAO on the climate network. In full analogy to the definition of the ENSO– and ENSO+ periods, we define NAO– and NAO+ periods using the NAO index. Since NAO reflects synchronous variations of the pressure gradient between the Icelandic Low and Azores High on timescales from daily to multidecadal, the NAO index is defined as the normalised pressure difference between the Azores and Iceland. The AWC geographic distributions, computed for the networks obtained by thresholding the Z -scores, for the negative and positive NAO periods, are depicted in Fig. 10a and Fig. 10b, respectively. In these and other similar computations, the surrogate data are used in order to obtain average reference levels. Therefore they are constructed using the whole analysed segment, while the NAO– and NAO+ division applies only for the original temperature data.

In Fig. 10 we can see that the connectivity in the NH NAO areas is enhanced during the NAO+ periods. Hurrell and Dickson (2005) help us to understand our finding: “in the so-called positive phase, higher than normal surface pressures south of 55° N combine with a broad region of anomalously low pressure throughout the Arctic and subarctic. Consequently, this phase of the oscillation (NAO+) is associated with stronger-than-average westerly winds across the middle latitudes of the Atlantic onto Europe, with anomalous southerly flow over the eastern United States and anomalous northerly flow across western Greenland, the Canadian Arctic, and the Mediterranean. The easterly trade winds over the subtropical North Atlantic are also enhanced during the positive phase of the oscillation. During the negative phase (NAO–), both the Icelandic low- and Azores high-pressure centres are weaker-than-normal, so both the middle latitude westerlies and the subtropical trade winds are also weak”. To put it more simply, the stronger transport of air masses during NAO+ periods enhances the climate network connectivity in the areas affected by NAO.

It is also interesting to study the geographical distribution of the betweenness centrality (BC, Fig. 11) computed under the same conditions as AWC in Fig. 10. We can see that the areas most central in the sense of information transfer, as defined by BC, are very close to the most connected areas according to AWC.

If we map AWC for NAO± periods, obtained from the networks based on thresholding the absolute correlations without significance testing, we obtain both maps (not presented) similar to the map in Fig. 4a. The NAO influence is masked by the upward biased connectivity of the tropical areas. Figure 12 displays betweenness centrality for this case. Similar

patterns like in Fig. 12 have been observed by Donges et al. (2009b). In this approach the most central areas (referred to as the backbone of the climate network by Donges et al., 2009b) are different from the most connected (tropical) areas. It is a question what is the role of the connectivity bias in computing BC and whether the patterns in Fig. 12 reflect a physical phenomenon. In any case, constructing the network without accounting for the dynamic bias in the connectivity measure, neither BC nor AWC are able to uncover the NAO influence in the Northern Hemisphere.

Solar radiation is the main energy source for the atmospheric envelope of the Earth, thus the Sun has an obvious effect on climate. Nevertheless, there is a long-standing controversy about the role of the solar variability in climate change (see reviews by Bard and Frank (2006); Gray et al. (2010) and references therein). It is natural to ask if we can detect any influence of variable solar activity on the climate network topology. Interestingly, Kodera (2002) has found that there is a solar influence on the extent of the NAO phenomenon: During the periods of high solar activity the NAO has a quite more extended spatial structure than during low solar activity.

The F10.7 index is a measure of the solar radio flux per unit frequency at a wavelength of 10.7 cm, near the peak of the observed solar radio emission. It represents a measure of diffuse, nonradiative heating of the coronal plasma trapped by magnetic fields over active regions, and is an excellent indicator of overall solar activity levels. Therefore we use the F10.7 index in the same way as the ENSO and NAO indices, i.e. we define F10.7– and F10.7+ periods for low and high solar activity. The related geographical distribution of AWC of the climate networks obtained by thresholding the Z -scores is illustrated in Fig. 13. In agreement with more extended NAO during solar maxima (Kodera, 2002) we can see stronger connectivity during the high solar activity periods (Fig. 13b). In both the NAO+ (Fig. 10b) and F10.7+ (Fig. 13b) periods, the connectivity in the areas influenced by NAO is enhanced, however, the patterns are not identical.

5 Conclusion

We have demonstrated that dynamics of time series (e.g. reflected by serial correlations) has a marked influence on dependence measures such as the absolute cross-correlations or mutual information. In particular, processes with lower entropy rates (i.e. processes with higher regularity, or dynamical memory) tend to have their non-negative dependence measures biased upward. We observe this phenomenon in the case of surface air temperature anomalies from tropical areas, which leads to falsely increased connectivity in the part of the climate network related to the tropical, especially tropical Pacific areas. After correcting for this bias, the most connected areas of the climate networks move from the tropical areas influenced by the El Niño Southern Oscillation to

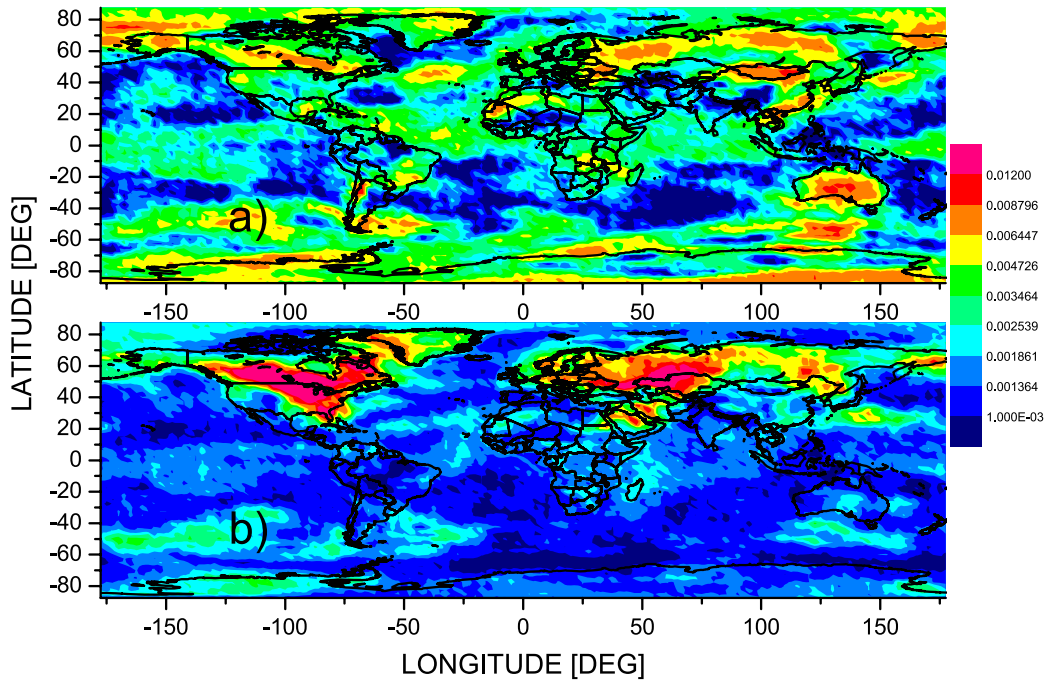


Fig. 10. Area weighted connectivity for SATA climate networks with density $\rho = 0.005$ obtained by thresholding of the absolute correlation Z-scores related to independent univariate FT surrogate data for negative (a) and positive (b) NAO index.

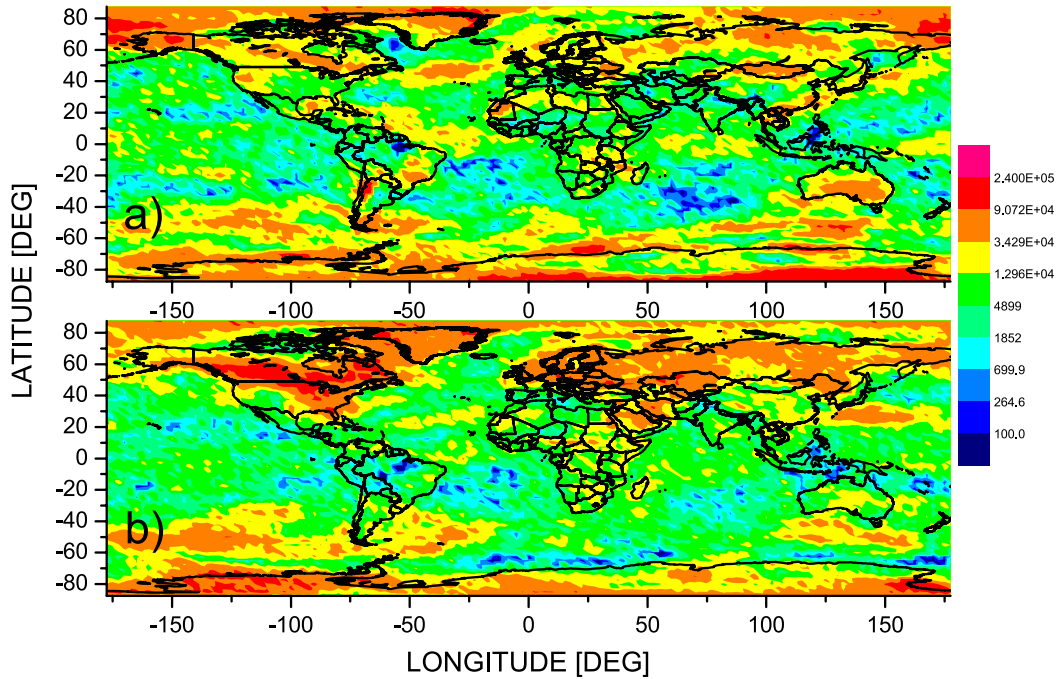


Fig. 11. Betweenness centrality for SATA climate networks with density $\rho = 0.005$ obtained by thresholding of the absolute correlations Z-scores related to independent univariate FT surrogate data for negative (a) and positive (b) NAO index.

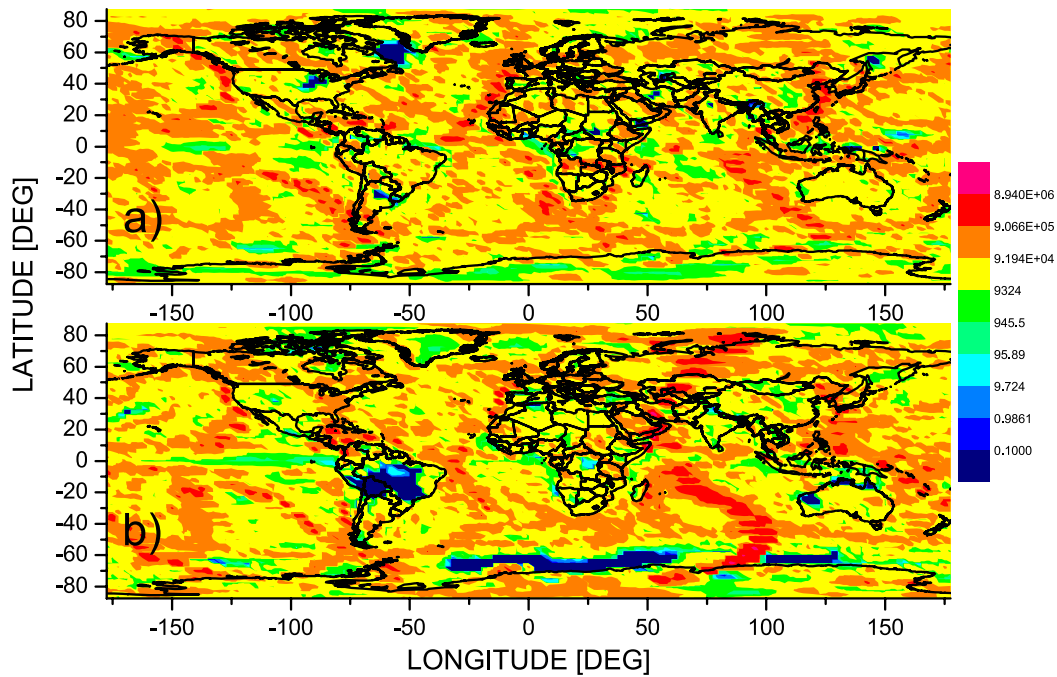


Fig. 12. Betweenness centrality for SATA climate networks with density $\rho = 0.005$ obtained by uniform thresholding of the absolute correlations for negative (a) and positive (b) NAO index.

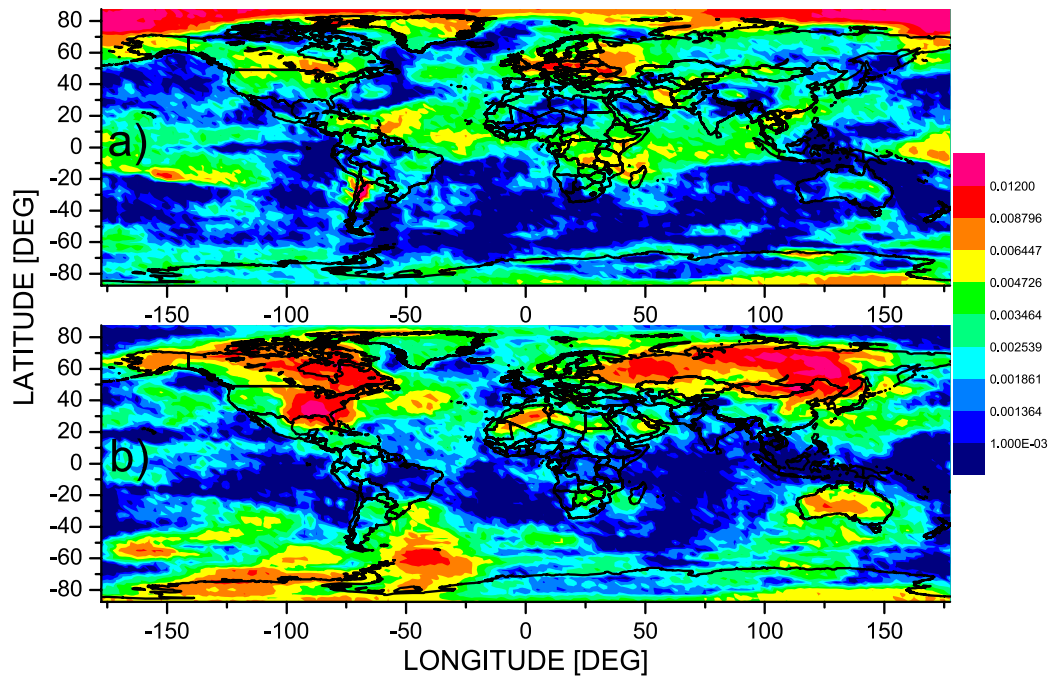


Fig. 13. Area weighted connectivity for SATA climate networks with density $\rho = 0.005$ obtained by thresholding of the absolute correlations Z-scores related to independent univariate FT surrogate data for low (a) and high (b) solar activity as measured by the 10.7 cm solar radio flux.

the Northern Hemisphere areas dominated by the North Atlantic Oscillation. Using this corrected connectivity measure we are able to observe the influence of NAO and solar variability on the connectivity in the climate networks.

The correction of the bias is done by replacing the absolute correlations by their Z-scores based on independent Fourier transform surrogate data, realisations of processes preserving the original spectra of the studied process and their entropy rates in the Gaussian approximation. This approach corrects for the bias due to high regularity of analysed time series, such as SATA in the tropics. We do not claim, however, that the proposed approach to measuring connectivity removes all possible biases. It should also be investigated whether this compensation is generally applicable and is not disproportionate under certain circumstances, decreasing the actual role of El Niño influence on the climate network connectivity. Possible distortions of SATA distributions from Gaussianity do not play an important role when the FT surrogate data are used for the null hypothesis of independence – see the negligible effect of amplitude adjustment in the AAFT surrogates in Fig. 6. The cross-correlations of the independent FT surrogates have reasonably normal distribution (Fig. 6), and the Z-score is a statistical quantity, suitable for the thresholding used to construct binary networks. However, its applicability for weighted networks should be carefully assessed. Therefore the choice of a proper connectivity measure for constructing climate networks is still a question of intensive research. Further research is also necessary to distinguish direct from indirect connections (Nawrath et al., 2010) as well as to assess possible causal influences (Hlavackova-Schindler et al., 2007; Paluš, 2007; Vejmelka and Paluš, 2008). Such tasks, however, may be computationally intractable, since the large networks consisting of $\approx 10^4$ nodes based on correlations of $\approx 10^2$ samples lead to singular, noninvertible matrices. Decreasing of dimensionality by searching for communities in the climate networks (Tsonis et al., 2011) or applying some clustering methods (Vejmelka and Paluš, 2010) is a desirable step for both the preprocessing of the multivariate atmospheric data as well as for a better understanding of evolving spatio-temporal structures in the climate networks.

Acknowledgements. This study was supported by the Czech Science Foundation project No. P103/11/J068.

Edited by: R. Donner

Reviewed by: two anonymous referees

References

- Bard, E. and Frank, M.: Climate change and solar variability: What's new under the sun?, *Earth Planet. Sci. Lett.*, 248, 1–14, doi:10.1016/j.epsl.2006.06.016, 2006.
- Barreiro, M., Marti, A. C., and Masoller, C.: Inferring long memory processes in the climate network via ordinal pattern analysis, *Chaos*, 21, 013101, doi:10.1063/1.3545273, 2011.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U.: Complex networks: Structure and dynamics, *Phys. Rep.*, 424, 175–308, doi:10.1016/j.physrep.2005.10.009, 2006.
- Cover, T. and Thomas, J.: *Elements of information theory*, Wiley Online Library, 1991.
- Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: Complex networks in climate dynamics, *Eur. Phys. J.-Spec. Top.*, 174, 157–179, doi:10.1140/epjst/e2009-01098-2, 2009a.
- Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: The backbone of the climate network, *Europhys. Lett.*, 87, 48007, doi:10.1209/0295-5075/87/48007, 2009b.
- Donges, J. F., Schultz, H. C. H., Marwan, N., Zou, Y., and Kurths, J.: Investigating the topology of interacting networks, *Eur. Phys. J. B*, doi:10.1140/epjb/e2011-10795-8, in press, 2011.
- Gozolchiani, A., Yamasaki, K., Gazit, O., and Havlin, S.: Pattern of climate network blinking links follows El Niño events, *Europhys. Lett.*, 83, 28005, doi:10.1209/0295-5075/83/28005, 2008.
- Gray, L. J., Beer, J., Geller, M., Haigh, J. D., Lockwood, M., Matthes, K., Cubasch, U., Fleitmann, D., Harrison, G., Hood, L., Luterbacher, J., Meehl, G. A., Shindell, D., van Geel, B., and White, W.: Solar influences on climate, *Rev. Geophys.*, 48, RG4001, doi:10.1029/2009RG000282, 2010.
- Hartman, D., Hlinka, J., Paluš, M., Mantini, D., and Corbetta, M.: The role of nonlinearity in computing graph-theoretical properties of resting-state functional magnetic resonance imaging brain networks, *Chaos*, 21, 013119, doi:10.1063/1.3553181, 2011.
- Hlavackova-Schindler, K., Paluš, M., Vejmelka, M., and Bhat-tacharya, J.: Causality detection based on information-theoretic approaches in time series analysis, *Phys. Rep.*, 441, 1–46, doi:10.1016/j.physrep.2006.12.004, 2007.
- Hurrell, J. and Dickson, R.: Climate variability over the North Atlantic, in: *Marine ecosystems and climate variation: the North Atlantic: a comparative perspective*, edited by: Stenseth, N., Ottersen, G., and Hurrell, J., Oxford University Press, USA, 15–31, 2005.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Met. Soc.*, 77, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.
- Kodera, K.: Solar cycle modulation of the North Atlantic Oscillation: Implication in the spatial structure of the NAO, *Geophys. Res. Lett.*, 29, 1218, doi:10.1029/2001GL014557, 2002.
- Malik, N., Bookhagen, B., Marwan, N., and Kurths, J.: Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks, *Clim. Dynam.*, doi:10.1007/s00382-011-1156-4, in press, 2011.
- Marshall, J., Kushnir, Y., Battisti, D., Chang, P., Czaja, A., Dickson, R., Hurrell, J., McCartney, M., Saravanan, R., and Visbeck, M.: North Atlantic climate variability: phenomena, impacts and mechanisms, *Int. J. Clim.*, 21, 1863–1898, doi:10.1002/joc.693, 2001.
- Nawrath, J., Romano, M. C., Thiel, M., Kiss, I. Z., Wickramasinghe, M., Timmer, J., Kurths, J., and Schelter, B.: Distinguishing direct from indirect interactions in oscillatory networks with multiple time scales, *Phys. Rev. Lett.*, 104, 038701, doi:10.1103/PhysRevLett.104.038701, 2010.

- Paluš, M.: Testing for nonlinearity using redundancies – quantitative and qualitative aspects, *Physica D*, 80, 186–205, 1995.
- Paluš, M.: Coarse-grained entropy rates for characterization of complex time series, *Physica D*, 93, 64–77, 1996.
- Paluš, M.: On entropy rates of dynamical systems and Gaussian processes, *Phys. Lett. A*, 227, 301–308, 1997.
- Paluš, M.: From nonlinearity to causality: statistical testing and inference of physical mechanisms underlying complex dynamics, *Contemp. Phys.*, 48, 307–348, doi:10.1080/00107510801959206, 2007.
- Paluš, M. and Novotna, D.: Testing for nonlinearity in weather records, *Phys. Lett. A*, 193, 67–74, 1994.
- Paluš, M., Albrecht, V., and Dvorak, I.: Information theoretic test for nonlinearity in time series, *Phys. Lett. A*, 175, 203–209, 1993.
- Schreiber, T. and Schmitz, A.: Surrogate time series, *Physica D*, 142, 346–382, doi:10.1016/S0167-2789(00)00043-9, 2000.
- Steinhaeuser, K., Ganguly, A., and Chawla, N.: Multivariate and multiscale dependence in the global climate system revealed through complex networks, *Clim. Dynam.*, doi:10.1007/s00382-011-1135-9, in press, 2011.
- Tsonis, A. and Roebber, P.: The architecture of the climate network, *Physica A*, 333, 497–504, doi:10.1016/j.physa.2003.10.045, 2004.
- Tsonis, A. and Swanson, K.: Topology and predictability of El Niño and La Niña networks, *Phys. Rev. Lett.*, 100, 228502, doi:10.1103/PhysRevLett.100.228502, 2008.
- Tsonis, A., Swanson, K., and Roebber, P.: What do networks have to do with climate?, *Bull. Amer. Meteorol. Soc.*, 87, 585–596, doi:10.1175/BAMS-87-5-585, 2006.
- Tsonis, A., Swanson, K., and Wang, G.: On the role of atmospheric teleconnections in climate, *J. Climate*, 21, 2990–3001, doi:10.1175/2007JCLI1907.1, 2008.
- Tsonis, A., Wang, G., Swanson, K., Rodrigues, F., and Costa, L.: Community structure and dynamics in climate networks, *Clim. Dynam.*, 37, 933–940, doi:10.1007/s00382-010-0874-3, 2011.
- Vejmelka, M. and Paluš, M.: Inferring the directionality of coupling with conditional mutual information, *Phys. Rev. E*, 77, 026214, doi:10.1103/PhysRevE.77.026214, 2008.
- Vejmelka, M. and Paluš, M.: Partitioning networks into clusters and residuals with average association, *Chaos*, 20, 033103, doi:10.1063/1.3460360, 2010.
- von Bloh, W., Romano, M. C., and Thiel, M.: Long-term predictability of mean daily temperature data, *Nonlin. Processes Geophys.*, 12, 471–479, doi:10.5194/npg-12-471-2005, 2005.
- Yamasaki, K., Gozolchiani, A., and Havlin, S.: Climate networks around the globe are significantly affected by El Niño, *Phys. Rev. Lett.*, 100, 228501, doi:10.1103/PhysRevLett.100.228501, 2008.
- Yamasaki, K., Gozolchiani, A., and Havlin, S.: Climate networks based on phase synchronization analysis track El-Niño, *Prog. Theor. Phys. Suppl.*, 179, 178–188, doi:10.1143/PTPS.179.178, 2009.