Manuscript prepared for Nonlin. Processes Geophys.
with version 5.0 of the LaTeX class copernicus.cls.
Date: 5 September 2014

Correlations Between Climate Network and Relief Data

T. K. D. Peron[1], C. H. Comin[1], D. R. Amancio[2], L. da F. Costa[1], F. A. Rodrigues[2], and J. Kurths[3,4,5]

[1]São Carlos Institute of Physics, University of São Paulo, São Carlos, São Paulo, Brazil
[2]Institute of Mathematics and Computer Science, University of São Paulo, São Carlos, São Paulo, Brazil
[3]Potsdam Institute for Climate Impact Research (PIK), 14473 Potsdam, Germany
[4]Department of Physics, Humboldt University, 12489 Berlin, Germany
[5]Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB24 3UE, United Kingdom

Abstract. In the last few years, the scientific community has witnessed an ongoing trend of using ideas developed in the study of complex networks to analyze climate dynamics. This powerful combination, usually called climate networks, can be used to uncover non-trivial patterns of weather changes along the years. Here we investigate the temperature network of North America region and show that two network characteristics, namely degree and clustering, have markedly differences between the Eastern and Western regions. We show that such differences are a reflection of the presence of a large network community in the western side of the continent. Moreover, we provide evidence that this large community is a consequence of the peculiar characteristics of the western relief of North America.

1 Introduction

Complex networks are powerful tools for describing the structure and functioning of a wide range of natural, technological and social systems (da Fontoura Costa et al., 2011). Owing to the general framework that the network theory provides, a mathematical representation of such systems is straightforward, allowing not just the description of networked topologies, but also leading to a better comprehension of dynamical processes in systems whose elements are connected in a non-trivial fashion (Boccaletti et al., 2006). In the past few years, complex networks have also been applied in climate sciences, creating this way the new field of *climate networks* (Tsonis et al., 2006, 2008; Tsonis and Swanson, 2008; Donges et al., 2009a,b; Gozolchiani et al., 2008; Tsonis and Roebber, 2004; Yamasaki et al., 2008). According to this paradigm, climate networks are formed by nodes, corresponding to spatial grid points in a given global climate data. These nodes are connected by edges, which correspond to statistical similarities between times series of given climate variables (e.g., temperature, relative humidity, precipitation) associated to each node in the network. Although this field is relative new in the network research, several results have been reported showing that network measurements can indeed give new important insights into climate dynamics (Tsonis et al., 2006, 2008; Tsonis and Swanson, 2008; Donges et al., 2009a,b; Gozolchiani et al., 2008; Tsonis and Roebber, 2004; Yamasaki et al., 2008; Rheinwalt et al., 2012; Mheen et al., 2013; Runge et al., 2014). For instance, by using degree centrality measurements of climate networks, researchers were capable of identifying highly connected nodes, which turned out to be related with the North Atlantic Oscillation. These results revealed that climate networks can exhibit small-world properties due to long-range edges (called teleconnections) connecting highly distant nodes (Tsonis et al., 2006, 2008). Moreover, the analysis of the teleconnections unveiled by this framework has also shed light on the study of extreme climate events, such as the El Niño-Southern Oscillations (ENSO) (Tsonis and Swanson, 2008; Gozolchiani et al., 2008). More specifically, by constructing climate networks of the surface temperature field during El Niño and La Niña periods, it was found that ENSO has a strong impact on the stability of climate systems, which is manifested as the decrease of the temperature predictability during El-Niño years. It is worth noting that the application of concepts from complex network theory in climate sciences has brought new insights that could not be unveiled by using classical methods of climatology and statistics. Recently, by using cross-correlation and mutual information to construct climate networks and analyzing the betweenness centrality field (node centrality measurement based on shortest path lengths (Costa et al., 2007)), researches found wave-like structures that are related to surface ocean currents, detecting this way a backbone of significantly increased matter and energy flow in the global surface air temperature field (Donges et al., 2009a,b). Furthermore,

the authors also showed that these results cannot be achieved by using methods derived from multivariate analysis, such as principal component analysis (PCA) and singular spectrum analysis (SSA) (Donges et al., 2009a). In this work, we extend the analysis of climate networks investigating the influence of altitudes of the grid points on centrality measurements of the networks generated through similarities in temperature time series measured at the surface level. The main motivation for including the altitudes on the network model is the assumption that the flow of matter and energy can be affected by topographical barriers, leading to anomalies in the correlations between the time series of climate variables. Therefore, in order to uncover these phenomena and quantify the influence of the relief on the network correlations, for each node v we associate its geographical altitude h_v with centrality measurements of the climate network, such as betweenness and clustering coefficient.

We constructed climate networks allowing the existence of long-range connections. By detecting communities in the climate networks, we found clusters that correspond to groups of nodes embedded in geographical areas of similar relief properties. Moreover, it was also found that the correlation patterns between centrality measurements and relief properties vary according to the considered network community. Finally we point out a possible effect of time series interpolation generated by stations in the degree and clustering coefficient fields of the networks.

2 Materials and Methods

2.1 Dataset description

Throughout the analysis we used the following databases:

(i) Monthly land temperature records from the National Center for Environmental Prediction/National Center for Atmospheric Research NCEP/NCAR (Kistler et al., 2001; Fan and Van den Dool, 2008) obtained from January 1948 to January 2011. The dataset consists of a regular spatio-temporal grid with 0.5^o of latitude and longitude resolution. Each grid point i has a temperature time series $\boldsymbol{T}_i(t)$ associated, containing the time evolution of the monthly mean temperature. A visualization of stations employed in the analysis that originated the database is shown in Fig. 1 (data provided by the NOAA, 2013).

(ii) Relief dataset provided by National Geophysical Data Center (NGDC, 2009) and consisting of 1-arc minute regular gridded area measuring land topography and ocean bathymetry.

2.2 Complex networks measurements

In order to seek for relationships between the climate and relief, we use network measurements related to centrality and symmetry of connections. The most simple of them, referred to as node *degree*, is given by

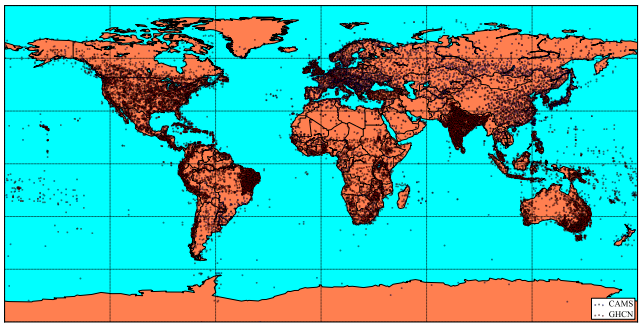

Fig. 1. Visualization of the stations used to interpolate the grid points in the temperature database.

$$k_i = \sum_{j=1}^{N} A_{ij}, \tag{1}$$

where $A_{ij} = 1$ if nodes i and j are connected and $A_{ij} = 0$ otherwise. The degree is a simple way to study the local importance of a node. Concerning climate networks, the degree can be used to quantify how many points of the studied region display a time series similar to a given point in the globe. In other words, nodes with large degrees are related to large regions of correlation.

The *clustering* coefficient of a node is the probability that two of its neighbors are also connected in the network, and is given by (da Fontoura Costa et al., 2011)

$$c_i = \frac{2\mathcal{T}(i)}{k_i(k_i - 1)}, \tag{2}$$

where $\mathcal{T}(i)$ is the number of triangles passing through i, or equivalently, the number of connections between neighbors of i. The clustering bears an interesting local information. If a given point of the globe is strongly correlated with two other points, the clustering quantifies how often these two points are also strongly correlated between themselves. The existence of regions taking low values of c_i suggests that the propagation of climate changes occurs in a streamlined fashion in those regions. Conversely, large clustering is related to a more diffusive propagation.

Another feature we study is *betweenness* centrality of a node. To define this measurement, consider the following notation. Let σ_{st} be the number of shortest paths from node s to node t (da Fontoura Costa et al., 2011). If $\sigma_{st}(i)$ is the number of such paths passing through node i, the betweenness centrality is given by (da Fontoura Costa et al., 2011)

$$b_i = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}. \tag{3}$$

It gives information about global relationships in climate dynamics. It is of great importance in quantifying if a node is

commonly used as a route for long-range correlations in the network (Donges et al., 2009a).

A node can be central but still not communicate well with the rest of the network. For instance, a node that is connected to another with large degree can be regarded as being central in the network, but it has a strong dependence on its highly connected neighbor. The *accessibility* measurement quantifies the number of nodes effectively accessed after h steps, where the node accessed in the next step is chosen randomly. Formally, the accessibility is computed as

$$a_i = \frac{1}{N_i^h} \exp\left(-\sum_j^N P_{ij}^h \log P_{ij}^h \right), \quad (4)$$

where P_{ij}^h is the probability that a random walk starting at node i arrives at node j in h steps, N_i^h the number of reachable nodes in h steps from node i and $\exp(\cdot)$ is the exponential function (see, e.g., Viana et al., 2012 for a detailed explanation of this measurement).

Real-world networks often display a modular structure, i.e., the presence of communities (Fortunato, 2010). The modular structure of a given network can be quantified by the measurement known as *modularity*, which is given by (Newman, 2003)

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m}, \right) \delta(C_i, C_j), \quad (5)$$

where $m = 1/2 \sum A_{ij}$ is the total number of edges, C_i is the community to which node i belongs and δ is the Kronecker delta. Once the partitioning of the nodes into communities is done, the modularity Q basically calculates the fraction of edges that connects nodes of the same community subtracting the fraction of these edges that we would expect to find in a random graph with the same degree sequence. Thus, eq. 5 provides a significance test of the obtained network partitioning, which will be used to validate our results in the next sections.

Since the modularity Q quantifies how good a given partition is, many methods intended to uncover communities in networks are based on the optimization of this measurement. Different strategies for the modularity optimization have been adopted in the literature such as simulated annealing (Reichardt and Bornholdt, 2006; Guimera et al., 2004), greedy algorithms (Newman, 2004; Clauset et al., 2004) and extremal optimization (Duch and Arenas, 2005). Although these algorithms provide accurate results, most of them a have great computational cost. For this reason, we adopt the method proposed in (Newman, 2006) to obtain the community structure of climate networks. This method consists in mapping the modularity optimization in terms of the spectrum of the so-called *modularity matrix* $\mathbf{B}$ defined as

$$\mathbf{B} = \mathbf{A} - \frac{\mathbf{k}\mathbf{k}^\mathbf{T}}{2m}, \quad (6)$$

where $\mathbf{A}$ is the adjacency matrix, m is as defined before in eq. 5 and $\mathbf{k} = [k_1, ..., k_N]^T$ the vector whose element k_i is the degree of the i-th node. The spectral optimization of the modularity Q has complexity of order $O(N^2 \log N)$, which turns out to be faster than, for instance, simulated annealing and extremal optimization approaches, besides providing more accurate results for large networks (Newman, 2006; Fortunato, 2010).

2.3 Climate networks

Because we are most interested in the topological characteristics of climate networks and its correlations with relief heights, we consider now only the connected subgraph whose nodes are located inside a continent. Note that we do not simply extract the subgraph over land discarding any edges which connects nodes on the ocean, rather we recalculate the threshold ϵ by taking into account only the nodes in the spatio-temporal grid which are over land.

Having the values of temperatures for each grid point in the dataset, a simple way to infer that two points have similar dynamical evolution is through the Pearson correlation coefficient between pairs of time series, which is given by

$$\rho_{ij} = \frac{\langle T_i T_j \rangle - \langle T_i \rangle \langle T_j \rangle}{\sqrt{(\langle T_i^2 \rangle - \langle T_i \rangle^2)(\langle T_j^2 \rangle - \langle T_j \rangle^2)}}, \quad (7)$$

where T_i is the time series associated to a point i in the spatio-temporal grid and $\langle X \rangle$ means the average of the variable X. Furthermore, we also remove the mean annual cycle in order to avoid seasonal effects in the time series. In this section, we describe the approach employed in our analysis.

We start with a fully connected network where each grid point is a node and two nodes are connected through an edge with an associated weight given by ρ_{ij}. The fully connected network can be studied by using weighted versions of the characteristics presented in section 2.2 (cf. Boccaletti et al., 2006 for a description of weighted measurements for graphs). Nevertheless, we are only interested in connections representing strong correlations. Hence, connections having a correlation smaller than a given threshold ϵ are discarded. This leads to a network defined by the adjacency matrix $\mathbf{A}$ whose elements are given by $A_{ij} = \Theta(\rho_{ij} - \epsilon) - \delta_{ij}$, where $\Theta(\cdot)$ is the Heaviside function. The threshold ϵ should be chosen in order to keep the network edges that correspond to strong correlation between time series, thus eliminating the non-relevant ones (Tsonis et al., 2006; Tsonis and Swanson, 2008; Tsonis et al., 2008; Gozolchiani et al., 2008; Donges et al., 2009a). Therefore, for all networks analysed in this approach, the threshold ϵ was chosen so that only 5% of the connections are kept in the network. Without the constraint of only first-neighbours connections, it is reasonable to expect a much richer pattern of connectivity with, e.g., presence of communities in the network, i.e., clusters of nodes that are more connected inside these groups than external nodes to

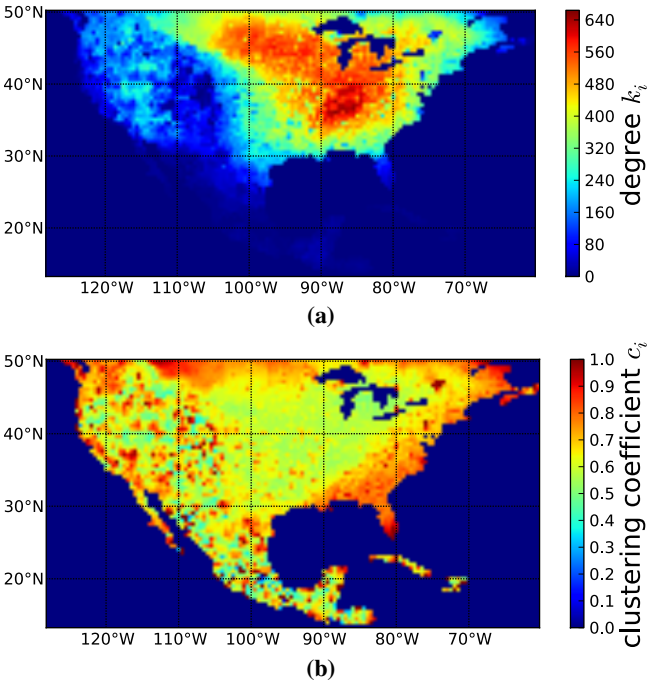

Fig. 2. (a) Degree k_i; and (b) *clustering* coefficient c_i obtained from the network of temperature correlations.

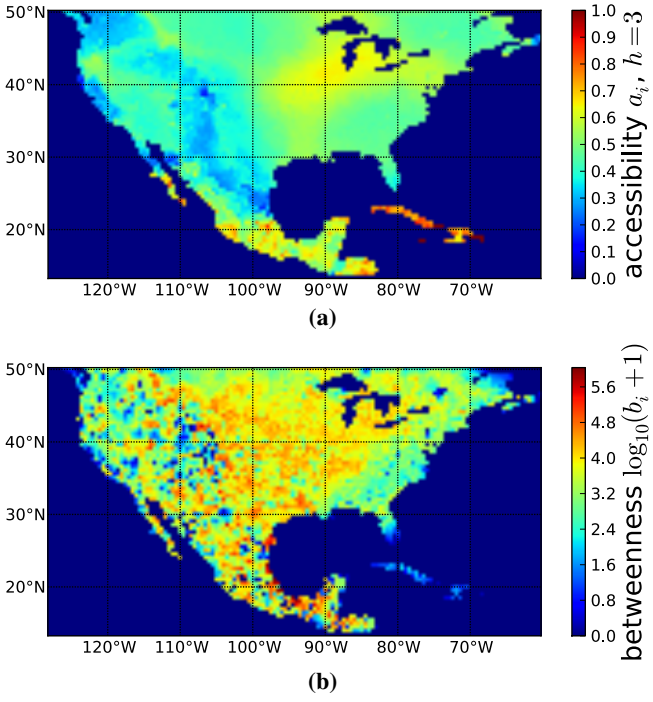

Fig. 3. (a) Betweenness centrality b_i; and (b) accessibility a_i for $h = 3$ steps obtained from the network of temperature correlations.

the cluster. In the context of climate networks, the grouping of nodes into communities was shown to be related to different climate patterns and to unveil different known climate zones (Tsonis et al., 2011).

3 Results

From reference (Fan and Van den Dool, 2008) we know that the land surface temperature database is constructed by interpolating recorded time series from stations spread over the globe. In order to avoid interpolation effects, it is useful to analyze the spatial distribution of the stations that generate this database. Using data from (NGDC, 2009), in Fig. 1 we show the stations location used to record the monthly average temperature time series. As we can see, except the northeast region of Brazil, South American is sparsely covered by stations, whereas North America and Europe are more densely covered. Therefore, in order to eliminate any doubts whether the observed patterns in the networks measurements are being affected by the interpolation or not, we turn our analysis to regions with high density of stations, namely, the North America region.

Applying the methodology described in Section 2.3, we obtain the climate networks and extract the centrality measurements for the region with the values of longitude θ and latitude ϕ ranging in the intervals $-128^o \leq \theta \leq -60^o$ and $30^o \leq \phi \leq 70^o$, respectively. Our results are shown in Fig. 2.

As we can see in Fig. 1, the region has stations approximately uniformly distributed. Therefore, we can discard the hypothesis that the area with high values for the degree in Fig. 2(a) is due to interpolation effects. It is also interesting to note that in Fig. 2(b) there are two distinct patterns in the clustering coefficient field. While the eastern region has an almost uniform distribution for c_i, the western region displays a more irregular distribution. The same pattern is also followed by the other centrality measurements. Figs. 3(a) and (b) shows the accessibility and betweenness centrality fields, respectively. Likewise, the patterns observed in the western and eastern regions differ significantly, especially for the accessibility. It is important to note that, according to Figs. 2(a) and 3(b), the regions taking low values of degree and accessibility overlap significantly. This pattern cannot be interpreted in a straightforward fashion, as the relevant correlation between degree and accessibility usually appears when the hierarchical definition of the degree is taken into account (Viana et al., 2012).

The topology of the climate network was further analyzed by identifying the natural topological communities. The communities arising from the application of the eigenvector strategy (see (Newman, 2006)) is shown in Fig. 4. A straightforward comparison of Figs. 2 and 4 reveals that the large community located at the western region corresponds to the nodes taking the lowest values of degree and accessibility

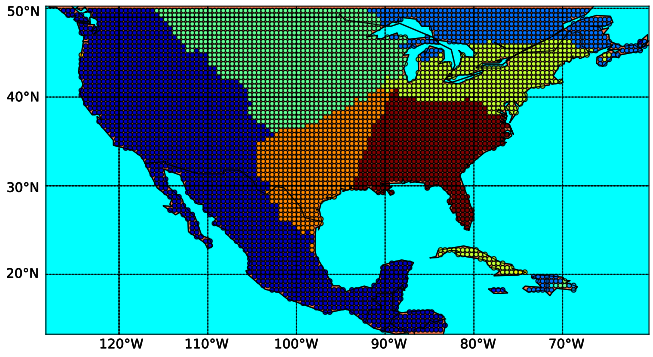

Fig. 4. Community structure for the network constructed with the grid points with θ and latitude ϕ in the intervals $-128^o \leq \theta \leq -60^o$ and $30^o \leq \phi \leq 70^o$ of the temperature database. Grid-points colored with the same color correspond to nodes belonging to the same network community.

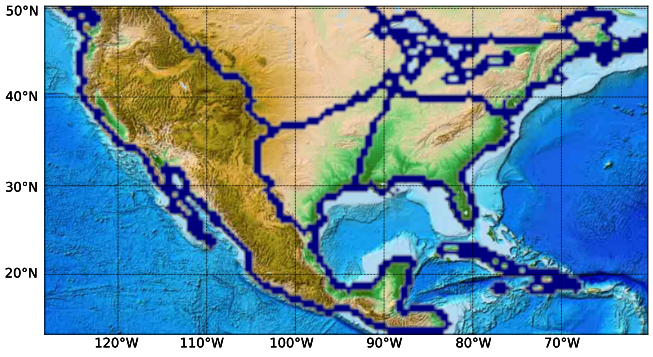

Fig. 5. Boundaries of the communities obtained from the climate networks. Note that the largest community coincides with a regular relief profile.

(see Figs. 2(a) and 3(a)). As for the clustering coefficient,it is irregularly distributed.

Fig. 5 displays the network communities and the relief structure. Remarkably, the variations in the largest community border in the west side of North America is followed by variations in the relief structure. Comparing Figs. 5 and 2, we notice that the contrast between the west and east region in the degree and clustering coefficient field is also observed in the relief structure. More specifically, the regions present very different patterns in the relief structure which is also revealed in the pattern of network measurements, suggesting that with our methodology we may be able to quantify the influence of the landscape in the climate network organization.

4 Conclusions

Despite being a recent field, climate networks have already been shown to provide valuable information about climate dynamics (Tsonis et al., 2006, 2008; Tsonis and Swanson, 2008; Donges et al., 2009a,b; Gozolchiani et al., 2008; Tsonis and Roebber, 2004; Yamasaki et al., 2008). In this study, we used the monthly land temperature records from NCEP/NCAR reanalysis to define correlations between stations, which are then transformed into network connections when they exceed a specified threshold. One important point raised during our investigation was the effect of the spatial distribution of stations on the resulting network. We found that data pertaining to the region in which $(-128^o, 30^o) \leq (\theta, \phi) \leq (-60^o, 70^o)$ should not suffer such effects, given its almost uniform distribution of stations. One important topic to be studied in the future is the specific effect of spatial heterogeneities in the sampled data on the formation of abnormal, but most likely predictable, structures in the network.

In this study, we showed that the North America, when modeled as a climate network, displays two regions with distinct topological properties. We have found that the eastern and western regions display striking differences of degree, accessibility and clustering coefficient, which may be explained by the presence of communities arising from the climate network. More specifically, the eastern side was found to be characterized by uniform values of centrality measurements. Conversely, the western side was mainly characterized by an heterogeneous distribution of measurements values. The relationship between climate and relief was analyzed in the relief dataset provided by NOAA jointly with the climate network data. Interestingly, we uncovered dynamics not detected by other traditional methods. The most important pattern arising from the analysis was the observation that the topological community of the climate network in the western region matched the region with peculiar relief structure, suggesting a strong influence of the relief on the climate dynamics.

Of paramount interest for future studies is to use other relevant climate variables (e.g., humidity, wind, pressure) to uncover additional relationships between relief and climate, using the ideas developed in the climate networks field, as well the boundary effects (Rheinwalt et al., 2012) of spatially embedded networks.

Acknowledgments

TKDP would like to acknowledge FAPESP (No. 2012/22160-7) and IRTG 1740. DRA is grateful for the financial support from FAPESP (grant number 2013/06717-4). FAR would like to acknowledge CNPq (No. 305940/2010-4), FAPESP (No. 2010/19440- 2) and IRTG 1740 for the financial support given to this research. LdaFC is grateful for the financial support from CNPq, FAPESP and CAPES. JK would like to acknowledge IRTG 1740 (DFG and FAPESP) for the sponsorship provided.

References

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.: Complex networks: Structure and dynamics, Physics reports, 424, 175–308, 2006.

Clauset, A., Newman, M. E., and Moore, C.: Finding community structure in very large networks, Physical review E, 70, 066 111, 2004.

Costa, L., Rodrigues, F., Travieso, G., and Boas, P.: Characterization of complex networks: A survey of measurements, Advances in Physics, 56, 167–242, 2007.

da Fontoura Costa, L., Oliveira Jr, O., Travieso, G., Rodrigues, F., Boas, P., Antiqueira, L., Viana, M., and Rocha, L.: Analyzing and modeling real-world phenomena with complex networks: a survey of applications, Advances in Physics, 60, 329–412, 2011.

Donges, J., Zou, Y., Marwan, N., and Kurths, J.: The backbone of the climate network, EPL (Europhysics Letters), 87, 48 007, 2009a.

Donges, J., Zou, Y., Marwan, N., and Kurths, J.: Complex networks in climate dynamics, The European Physical Journal-Special Topics, 174, 157–179, 2009b.

Duch, J. and Arenas, A.: Community detection in complex networks using extremal optimization, Physical review E, 72, 027 104, 2005.

Fan, Y. and Van den Dool, H.: A global monthly land surface air temperature analysis for 1948–present, Journal of Geophysical Research, 113, D01 103, 2008.

Fortunato, S.: Community detection in graphs, Physics Reports, 486, 75 – 174, 2010.

Gozolchiani, A., Yamasaki, K., Gazit, O., and Havlin, S.: Pattern of climate network blinking links follows El Nino events, EPL (Europhysics Letters), 83, 28 005, 2008.

Guimera, R., Sales-Pardo, M., and Amaral, L. A. N.: Modularity from fluctuations in random graphs and complex networks, Physical Review E, 70, 025 101, 2004.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., et al.: The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, Bulletin-American Meteorological Society, 82, 247–268, 2001.

Mheen, M., Dijkstra, H. A., Gozolchiani, A., Toom, M., Feng, Q., Kurths, J., and Hernandez-Garcia, E.: Interaction network based early warning indicators for the Atlantic MOC collapse, Geophysical Research Letters, 40, 2714–2719, 2013.

Newman, M. E.: Mixing patterns in networks, Physical Review E, 67, 026 126, 2003.

Newman, M. E.: Fast algorithm for detecting community structure in networks, Physical review E, 69, 066 133, 2004.

Newman, M. E. J.: Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E, 74, 036 104, 2006.

NGDC (National Geophysical Data Center): Relief dataset, available at; http://www.ngdc.noaa. gov/mgg/global/global.html (last access: 7 April 2014), 2009. 827, 830

NOAA (National Oceanic and Atmospheric Administration): Datasets and variables, available at: http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/ (last access: 7 April 2014), 2013. 826

Reichardt, J. and Bornholdt, S.: Statistical mechanics of community detection, Physical Review E, 74, 016 110, 2006.

Rheinwalt, A., Marwan, N., Kurths, J., Werner, P., and Gerstengarbe, F.-W.: Boundary effects in network measures of spatially embedded networks, EPL (Europhysics Letters), 100, 28 002, 2012.

Runge, J., Petoukhov, V., and Kurths, J.: Quantifying the Strength and Delay of Climatic Interactions: The Ambiguities of Cross Correlation and a Novel Measure Based on Graphical Models., Journal of Climate, 27, 2014.

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

Tsonis, A. and Swanson, K.: Topology and predictability of El Nino and La Nina networks, Physical Review Letters, 100, 228 502, 2008.

Tsonis, A., Swanson, K., and Roebber, P.: What do networks have to do with climate?, Bulletin of the American Meteorological Society, 87, 585–596, 2006.

Tsonis, A., Swanson, K., and Wang, G.: On the role of atmospheric teleconnections in climate, Journal of Climate, 21, 2990–3001, 2008.

Tsonis, A. A., Wang, G., Swanson, K. L., Rodrigues, F. A., and Costa, L. d. F.: Community structure and dynamics in climate networks, Climate dynamics, 37, 933–940, 2011.

Viana, M. P., Batista, J. a. L. B., and Costa, L. d. F.: Effective number of accessed nodes in complex networks, Phys. Rev. E, 85, 036 105, 2012.

Yamasaki, K., Gozolchiani, A., and Havlin, S.: Climate networks around the globe are significantly affected by El Nino, Physical Review Letters, 100, 228 501, 2008.