



Regional and inter-regional effects in evolving climate networks

J. Hlinka¹, D. Hartman¹, N. Jajcay¹, M. Vejmelka¹, R. Donner², N. Marwan², J. Kurths², and M. Paluš¹

¹Department of Nonlinear Dynamics and Complex Systems, Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodarenskou vezi 2, 182 07 Prague 8, Czech Republic

²Potsdam Institute for Climate Impact Research (PIK), 14473 Potsdam, Germany

Correspondence to: J. Hlinka (hlinka@cs.cas.cz)

Received: 17 December 2013 – Revised: 3 February 2014 – Accepted: 6 February 2014 – Published: 1 April 2014

Abstract. Complicated systems composed of many interacting subsystems are frequently studied as complex networks. In the simplest approach, a given real-world system is represented by an undirected graph composed of nodes standing for the subsystems and non-oriented unweighted edges for interactions present among the nodes; the characteristic properties of the graph are subsequently studied and related to the system's behaviour. More detailed graph models may include edge weights, orientations or multiple types of links; potential time-dependency of edges is conveniently captured in so-called evolving networks. Recently, it has been shown that an evolving climate network can be used to disentangle different types of El Niño episodes described in the literature. The time evolution of several graph characteristics has been compared with the intervals of El Niño and La Niña episodes. In this study we identify the sources of the evolving network characteristics by considering a reduced-dimensionality description of the climate system using network nodes given by rotated principal component analysis. The time evolution of structures in local intra-component networks is studied and compared to evolving inter-component connectivity.

1 Introduction

Complex networks (Newman, 2003; Boccaletti et al., 2006) represent a relatively young scientific discipline that has already influenced many research fields ranging from technological areas such as the Internet (Faloutsos et al., 1999), the World Wide Web (Albert et al., 1999), power grids or transportation networks (Guimera et al., 2005; Rosvall et al., 2005), through socially oriented research topics such as social networks (Wasserman and Faust, 1994), scientific collaboration networks (Newman, 2001) or financial markets

(Mantegna, 1999), to networks dealing with complex natural systems. The latter comprise systems like protein–protein interaction networks (Jeong et al., 2001), brain networks (Bullmore and Sporns, 2009) or climate networks (Tsonis and Roebber, 2004). Most of these systems are composed of many interacting subsystems whose coupling is hard to describe using any homogeneous or regular structural model. In these cases, the complex network paradigm can help with understanding the topology of the underlying connectivity structure (Arenas et al., 2008). Complex networks provide an analysis tool to uncover patterns and their influence on the behaviour of the studied system. The analysis is usually based on computing-specific characteristics of the networks under study.

In many complex systems there are connections between subsystems which are represented by physically existing or well-defined coupling such as connections in the Internet network, routes in transportation networks or documented cooperation in scientific collaboration networks. On the other hand, there is a relatively large class of systems, e.g. the brain or the Earth climate, where such structural connectivity is either not well-defined or not effectively measurable, while instrumental measurements given in the form of multivariate time series of key variables are available. In these cases, the connectivity may be inferred from statistical associations of pairs of time series. This means that the strength of the connection between two subsystems is represented as the mutual statistical relationship between two time series recorded from respective subsystems. This “functional connectivity” (a term coined originally in a neuroscientific context, see Friston et al., 1993 for details) may be identified using various measures of statistical dependence. The networks constructed from multivariate time series are known as functional networks or alternatively as interaction networks

(Bialonski et al., 2010). We will further focus on this class of systems and consider the climate system captured by multivariate time series of a meteorological variable.

The choice of an appropriate statistical dependence measure, such as Pearson's (linear) correlation or mutual information, may depend on the data character. In the context of complex nonlinear systems, the use of nonlinear functional connectivity methods seems advantageous due to sensitivity to specific non-linear dependences, and has also been advocated for climate network analysis (Donges et al., 2009a, b). However, recently proposed quantification approaches presented by Hlinka et al. (2011) and Hartman et al. (2011) suggest that although statistically significant, the nonlinear contribution to connectivity is for many practical purposes negligible, including the analysis of temperature data from climate reanalysis data sets (Hlinka et al., 2013a). For other issues related to the choice of an appropriate dependence measure see e.g. Paluš et al. (2011) and Hlinka et al. (2013b).

Since the edges in interaction networks are inferred from measures of statistical association, time series reflecting a long time interval are convenient in order to obtain robust estimates of an association measure. This is an optimal approach for stationary processes; in the case of non-stationary processes it provides some type of average connectivity estimate at best, but potentially also biased estimates (such as spurious correlations due to common long-term trends). Importantly, even the global structural information can be, and for sufficiently long time intervals usually is, time-dependent in a relevant manner, i.e. the non-stationarity may be of direct interest for the researcher. This fact forces network scientists to use generalisations that embody time evolution for complex networks. There are several approaches, of which the most prominent are *temporal networks* and *evolving networks*. For the temporal networks, see for example Holme and Saramäki (2012); every connection has defined intervals of its existence that altogether constitute a specific time-weighted network that is analysed using the designated approach. On the other hand, evolving networks are less complicated, since they merely involve a number of static networks using sliding windows in time and usual analyses from a complex network perspective that are parameterised with time. The result of this analysis is a time evolution of network characteristics that are often compared to the temporal evolution of specific phenomena. In the field of brain networks, Kuhnert et al. (2010) studied the evolution of network characteristics using several days of continuing EEG recording, while Bialonski and Lehnertz (2013) traced the evolution of a particular network property during an epileptic seizure.

In the field of climate networks the method of evolving networks has recently been used to analyse the temporal variability of surface air temperature correlation structures, providing insights into the global response of the climate system to events such as volcanic eruptions or different phases of the El Niño Southern Oscillations (Radebach et al., 2013). The authors analysed the evolving climate

network for 62 years of surface air temperature anomalies using the gridded whole-Earth NCEP/NCAR reanalysis data set (Kalnay et al., 1996). The original NCEP/NCAR data on an angularly regular grid were transformed into a quasi-isotropic icosahedral grid (Heikes and Randall, 1995) that ensures homogeneity in the number of geographic neighbours and nearest-neighbour grid point distances using interpolation. The quasi-isotropic icosahedral grid represents a more suitable spatial sampling of the globe for complex network analysis.

It is a question, however, how relevant for the characterisation of the dynamics of the climate system the spatial sampling is of either the original angularly regular grid or that of the quasi-isotropic icosahedral grid. Inspired by the approaches used in research on brain dynamics by Shirer et al. (2012), one may search for a functional parcellation of the globe in order to find regions of coherent climate variability, which might serve as natural nodes of the climate network. Modern data analysis provides a range of other methods for dimensionality reduction by definition of regions with relatively homogeneous time series, of which several have been applied recently in or developed for analysis of climate time series in the context of graph-theoretical analysis (see e.g. Steinhäuser et al., 2012; Tsonis et al., 2011; Fountalis et al., 2013).

Recently, Vejmelka et al. (2014) proposed a principled approach for estimation of the count of dynamically relevant components in a climate time series data set and their identification (a basic description of the method is available already in Hlinka et al. (2013b), where it was used along with grid-based dimensionality reduction to assess the effects of non-linearity for causal network estimates). We use this method in conjunction with varimax-rotated principal component analysis (Vejmelka and Paluš, 2010; Groth and Ghil, 2011) as an intermediate step in network analysis, as a method for concise representation of the data compared to the (climatologically largely irrelevant) spatial sampling used in the original gridded data set.

Using thus identified key (spatially representable) components of the dynamics, in the current paper we pose the questions whether, how and to what extent the temporal evolution of the grid-based climate network is already reflected in the dynamics of the higher level inter-component network and vice versa. Then we study the role of local (within-component) and distant (between-component) links in the global network evolution; in particular, we try to identify the sources of the most prominent extrema in the evolution of the time-dependent grid-based network characteristics.

2 Evolving network characteristics

2.1 Data

Following the study of Radebach et al. (2013) we use surface air temperature (SAT) data from the NCEP/NCAR reanalysis (Kalnay et al., 1996) on an angularly regular $2.5^\circ \times 2.5^\circ$ grid. The time interval covered by our data spans 62 years from 1948 to 2009. For the purpose of this study both daily and monthly data are used. The high temporal resolution (daily) data are used for the network evolution analysis using a sliding temporal window, while full-duration monthly data are for computational reasons used for identification of the key components/regions for the considered dimensionality reduction.

In order to minimise the bias introduced by the periodic changes in the solar input we first produce the *anomaly time series* by removing the long-term mean annual cycle from all considered local time series. Depending on the type of the data this long-term mean is either computed for each day or month of the year.

While the anomalization procedure is a standard step in most climate data analyses, for the dimensionality reduction, following the results of Hlinka et al. (2013a), we also remove the seasonality in variance from the time series at each location. This preprocessing step removes the differences in local temperature variability in different periods of the year. To apply this procedure for monthly data, each anomaly time series for a given month is divided by the standard deviation of anomalies for the given month. The final step of the preprocessing of monthly data is detrending removing slow trends from the data. Following Hlinka et al. (2013a) we consider only linear trends.

2.2 Defining regions of interests

In order to construct a climate network one needs to define a set of regions whose connections establish the network edges based on the strengths of statistical associations between the time series of a meteorological variable characterising each region of interest. For this purpose several strategies can be applied. The simplest way is to use the original data grid with time series directly adopted from the data source with appropriate preprocessing steps. While we have an angularly regular grid and use measures of statistical associations to define connections between regions, namely the correlation coefficient (see below), there can be a bias of the correlation dependent on the distance of the grid points from the poles and thus from each other due to mechanisms including inherent smoothness of the implicit interpolation scheme of the reanalysis as well as typical spatial scales of climate variability.

There are several ways to treat this potential source of bias. One standard procedure is so-called *cosine reweighting*. This process just scales the time series of each grid point by the

square root of the cosine of the latitude. Note also that the poles obtain zero weight in this scheme. As another alternative, we may mention the use of network statistics correcting for different areas represented by the grid points, especially “node-splitting invariant” network measures (Heitzig et al., 2012).

Another approach to account for different areas represented by each grid point on the angularly regular grid is the remapping of the original grid to a more suitable one. Following the data analysis settings used by Radebach et al. (2013), we apply an interpolation scheme based on the quasi-isotropic icosahedral grid (Heikes and Randall, 1995). This transformation starts with the initial projection of the vertices of an icosahedron to a sphere (putting two of them at the poles) and follows with a subsequent grid refinement by means of interpolation to obtain the final grid points’ placements. For details see the documentation of the SCRIP package used for this purpose (Jones, 1997). The resulting grid ensures that the area each point represents is approximately the same in all cases and the number of neighbours is the same almost everywhere.

In contrast to the use of a relatively fine geographically defined grid, we explore the use of data-driven dimensionality reduction. Of the many methods available, here we define the regions of interests via a rotated version of *principal component analysis* (PCA). (Rotated) PCA has long been used in climatology; however, typically only a few strongest components were taken into account and interpreted. Recently, rotated PCA has been proposed to be used to provide a set of all statistically significant climatic modes in the decomposition of a predefined scalar field (Vejmelka et al., 2014). This method is used here to detect the spatial distribution and representative time series of a set of key components of SAT variability.

For some purposes, including parcellation of the globe into subregions, the components need to be transformed into clusters. To this end, we used a simple maximum loading criterion, i.e. for each point of the original grid we determine the component with the highest component loading at the point, and assign this point to the corresponding cluster.

2.3 Construction of static network

Having defined the nodes of the climate network, previously called regions of interest, we have to determine edges to make the definition of the network complete. For this purpose measures of statistical association are used. According to Hlinka et al. (2013a) the correlation coefficient is satisfactory for monthly or daily surface air temperature data when using appropriate preprocessing. The pairwise correlation coefficients $r(x_i, x_j)$ of the time series $x_i(t)$, $i \in \{1, \dots, n\}$ (n is the number of nodes) define the $n \times n$ weighted connectivity matrix $\mathbf{W}_{i,j} = r(x_i, x_j)$. This matrix in fact represents a complete weighted graph, i.e. a graph where each edge exists and has a weight attached. Since for these networks the resulting

characteristics can be hard to interpret, it may be more convenient to transform it to an unweighted network given by a matrix $\mathbf{A}$ (simply called a graph in the following).

The simplest approach is to apply a threshold W^* by defining $A_{i,j} = 1$, $W_{i,j} > W^*$ and $A_{i,j} = 0$ otherwise. The resulting matrix $\mathbf{A}$ is called the *adjacency matrix*. When applying thresholding to several graphs it is sometimes convenient to determine the threshold based on a fixed predefined density of the graphs, where the density of the graph is defined as the actual number of edges divided by the maximum possible number of edges in a graph of a given size, i.e. for a graph G with m edges we have a density of $\rho = m / \binom{n}{2}$.

2.4 Network characteristics

For a graph $G = (V, E)$ with a node set $V = \{v_1, v_2, \dots, v_n\}$ of size n and $|E| = m$ edges several characteristics can be computed that could be used to analyse the underlying system (Boccaletti et al., 2006; Newman, 2003). One particularly useful characteristic defined for a node v_i is the *node degree* k_i defined as

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

One interesting aspect related to the degree is that it is one of the characteristics that can reveal a typical behaviour of the corresponding network which subsequently leads to classifying it as a complex network – namely according to the presence or absence of a power law-shaped degree distribution (see Barabási and Albert, 1999).

In addition, we also concentrate on characteristics that determine another typical complex network property, small-worldness. This property can be defined using two characteristics (Watts and Strogatz, 1998; Hlinka et al., 2012). The first one is the *clustering coefficient* defined as

$$C = \frac{1}{n} \sum_{i=1}^n C_i, \quad C_i = \frac{\sum_{j,k=1}^n A_{i,j} A_{i,k} A_{j,k}}{k_i (k_i - 1)}. \quad (2)$$

The local version of this characteristic, C_i , measures the average density of triangles centered at vertex v_i , which means that it shows the density of edges between all neighbours of vertex v_i . The last mentioned value can also be interpreted as the probability that two randomly chosen neighbours of v_i are mutually connected.

A slightly different property is measured by the network *transitivity* defined as

$$T = \frac{\sum_{i,j,k=1}^n A_{i,j} A_{i,k} A_{j,k}}{\sum_{i,j,k=1, j \neq k}^n A_{i,j} A_{i,k}}. \quad (3)$$

We can see that in this case the probability of neighbour interconnection is computed as the ratio between the numbers of triangles and connected triples of vertices in the graph, which imposes a different behaviour of this characteristic than revealed by C .

For any pair of vertices v_i and v_j in a graph G we can furthermore compute their distance $d_{i,j}$ as the length of the shortest path between v_i and v_j . Using this distance definition we can compute the *average path length* as

$$L = \frac{1}{n(n-1)} \sum_{i,j=1}^n d_{i,j}. \quad (4)$$

This characteristic can be influenced by the connectedness of the underlying graph. A graph is called *connected* if for all pairs of vertices there exists a path connecting them and unconnected otherwise. For unconnected graphs there exist maximal subgraphs that are connected themselves. These subgraphs are called *connected components*. Due to the structure of complex networks, for an unconnected network there usually exists one so-called *giant component* and possibly several smaller ones.

Considering the average path length, there are two basic approaches for its computation for unconnected graphs. Either we can compute the average from all connected components excluding unconnected pairs, or set the distance of unconnected pairs to the number of vertices and keep averaging over the whole vertex set. The latter option is used in this paper in line with the approach of Radebach et al. (2013). Another alternative would be to use a different yet related measure, the global efficiency, instead of L . The latter is defined using the harmonic mean of the path lengths, so that it is reasonable to set $d_{i,j} = \infty$ for vertices from disconnected network components.

2.5 Generalisation to evolving network

In natural complex systems changes to their structure in time can occur. Assuming an approximate step-wise stationarity, static networks can be constructed with a time-sliding window approach. This means that for a time interval t_{start} to t_{end} spanned by our data we define for a window size w and offset δ a series of intervals from $t_i - w$ to t_i for $i = 0, 1, 2, \dots, \ell$, where $t_0 = t_{\text{start}} + w$ and $t_{i+1} = t_i + \delta$ for $i = 0, 1, 2, \dots, \ell - 1$. Sliding windows analysis is used for daily SAT data for the entire time interval of 62 years from 1948 to 2009 using windows with $w = 365$ days and offset $\delta = 30$ days.

Considering the initial window displacement and 365 days per year, 62 years correspond to 742 time points. Note that time series of consecutive windows are partially overlapping and thus dependent.

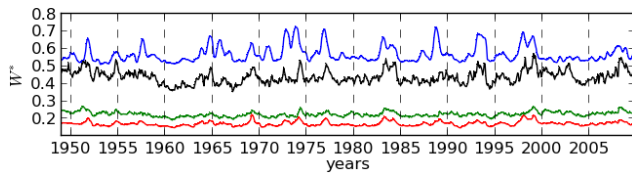


Fig. 1. Comparison of the evolution of the threshold W^* for the RPCA graph (68 components) and grid-based graph (with 2562 nodes) obtained for fixed densities. Blue line: 2562 nodes, $\rho = 0.005$, black line: 68 components, $\rho = 0.005$, red line: 2562 nodes, $\rho = 0.142$, green line: 68 components, $\rho = 0.142$.

3 Results

3.1 Comparison of grid-based and component-based network evolution

As a first step, we have constructed the evolving network according to the settings of Radebach et al. (2013), using an equidistant grid of 2562 points and density 0.005.

The evolution of the graph-theoretical characteristics reproduces well that reported in Radebach et al. (2013) (see Fig. 1).

To assess the extent to which the features of the evolution of the main graph-theoretical characteristics are already reproduced using the much smaller network of 68 components obtained from rotated PCA, we plot these along in the same Fig. 1.

Visual inspection suggests that most of the significant peaks within the grid-based time series are reproduced in the component network, albeit to a variable degree. There is a clear qualitative correspondence between the two time series, with most of the pronounced peaks shared between grid-based and component-based graph evolution.

This relationship is conserved, although varying in strength, when varying the densities of the two networks (see Fig. 2). Notably, the correlation increases with increasing density of the grid network (corresponding to a higher proportion of long-range connections, see Fig. 4 in Radebach et al., 2013). Also, the correlation is stronger for a lower density of the component network, suggesting that the evolution of only the core backbone of the inter-regional network is most relevant for explanation of the evolution of the global grid-based network.

This correspondence between the evolution of grid and component graphs suggests that the phenomena driving the evolution of the global grid graph structure have a substantial relation to effects that involve the interaction between modes of climate dynamics (to the extent that these are represented by the components). However, it does not inform us about the role of intra-regional connectivity.

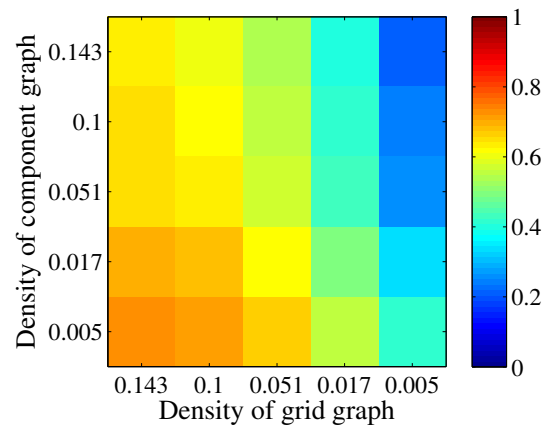


Fig. 2. Correlation coefficients between the time evolution of the applied thresholds for varying density of the RPCA graph (68 components) and grid-based graph with 2562 nodes.

3.2 Disentangling intra-regional and inter-regional contributions to graph evolution

To assess the relative role of local and global interactions, we utilise the rotated PCA for a crisp parcelation of the globe. In particular, each grid node is assigned uniquely to a component that has the highest loadings at the node, as described in Sect. 2.2.

We use these parcels for splitting the whole graph into multiple subgraphs, each defining either interactions within a given parcel, or between two selected parcels (forming a bipartite graph).

For each of these subgraphs, the evolution of its graph-theoretical characteristics is computed.

However, using a density of 0.005 as for the original full graph may lead to graphs with very few or even no links in the case of smaller subgraphs. Therefore we use a heuristic discounting method, changing the density as a function of graph size n so that the average degree k scales approximately as $k \sim \alpha \log n$, where $\alpha \sim 1.6$ is a constant that was fitted to the requirement that the density for the original network with $\log n = 2562$ is 0.005.

The role of the dynamics within a specific parcel (or among two parcels) in the evolution of the global dynamics can be then inspected by comparison of the time evolution. The regions (region pairs) that show intra-regional (inter-regional) connectivity variations with time similar to the global threshold evolution are further referred to as *dominant* regions (region pairs). We use this simplified term for ease of presentation, without speculating on the nature of the mechanism linking the (inter-)regional phenomena to the global graph evolution at this point.

For multiple regions, the evolution of the threshold needed for obtaining a pre-defined density of connection *within a region* shows marked similarity to the threshold evolution in the global network, for the most pronounced examples (see

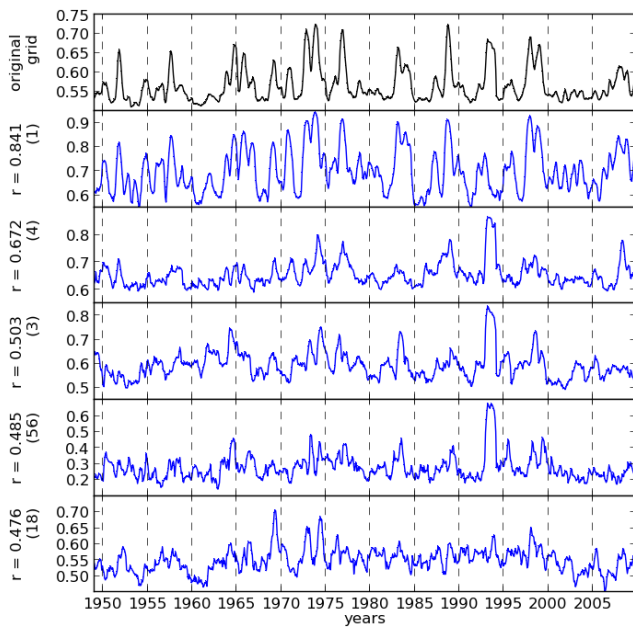


Fig. 3. The temporal evolution of threshold W^* for selecting the intra-regional components. Top panel: the temporal evolution of the global 2562-node network threshold. Bottom panel: the temporal evolution of threshold W^* for the selected five regions with the most similar evolution to the global graph evolution. For each component, its identification number and the correlation of the respective time series with the global 2562-node network threshold evolution is shown.

Fig. 3). The similarity is strongest for component 1, corresponding to a region in the eastern tropical Pacific, giving a Pearson correlation coefficient of 0.77 with respect to the global threshold evolution time series.

The other regions showing high correlation between the evolution of intra-regional connectivity and the global evolution include mainly tropical regions (see Fig. 4). Interestingly, among these tropical regions are the regions corresponding to components 3 and 4 that show a distinct peak in the year 1993 in their evolution, which can be related to the Mount Pinatubo volcanic eruption (see Radebach et al., 2013 for a discussion of the reflection of Pinatubo event in the global climate temperature network). The correlations of the connectivity dynamics from extra-tropical regions are considerably lower. The regions dominated by the respective components are visualised in Fig. 5.

As a next step, we explore the evolution of the inter-regional connectivity in a similar way. For each pair of components, we compute the threshold evolution and compare that to the global threshold evolution. The results are given in Figs. 6 and 7. The dominant inter-regional links show correlations up to 0.82 with the global threshold time series and typically correspond to interactions of the ENSO region with some other tropical or adjacent subtropical region.

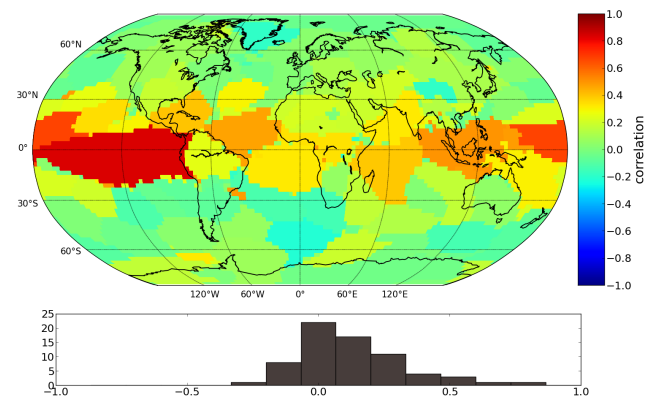


Fig. 4. Regions that show high relevance for the global graph evolution. Top panel: correlation of the intra-regional threshold evolution with the global 2562-node network threshold evolution. Bottom panel: histogram thereof.

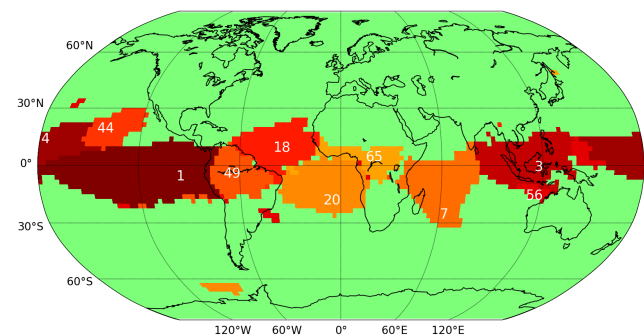


Fig. 5. Regions that show high relevance for the global graph evolution, due to characteristic evolution of either the intra-regional or inter-regional connectivity. Top 10 regions involved highlighted along with the respective component numbers for visual inspection of relevant time series in text and other figures.

To summarise the above-reported results, both intra-regional and inter-regional interactions are related to the global network evolution. The eastern to central tropical Pacific plays a central role in both intra-regional and inter-regional dominant interactions; however, connectivity of other tropical regions also plays an important role in the global connectivity changes. In particular, the global graph evolution is to a large extent reflected in the evolution of the interactions between the ENSO region and other tropical regions.

3.3 Splitting the network

To elucidate the relative role of the ENSO region in the eastern to central tropical Pacific (or more widely the whole tropics) in the global graph evolution, one may split the global grid-based graph into two parts: the ENSO (or tropics) and the extra-ENSO (or extra-tropics) areas. For each of the two regions, the set of graph-theoretical properties can be

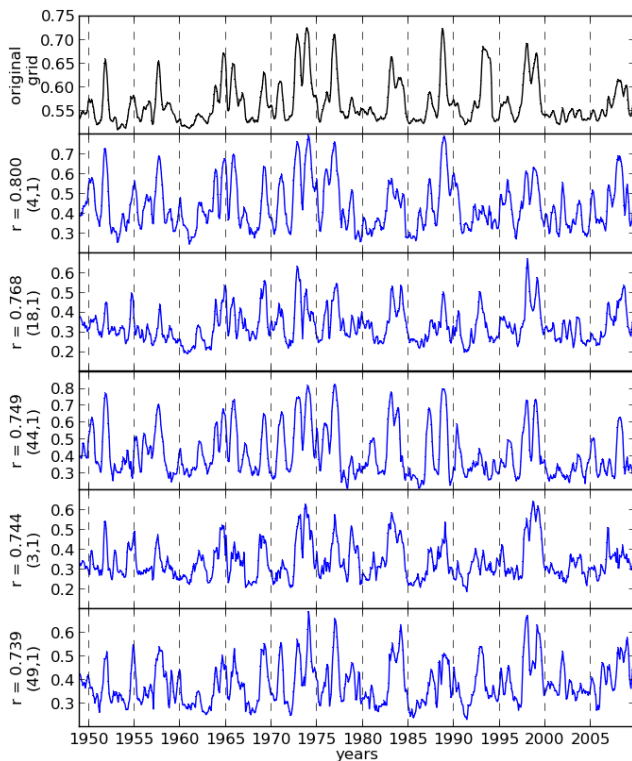


Fig. 6. The temporal evolution of the inter-regional components. Visualisation as in Fig. 3.

calculated and compared to the graph-theoretical properties of the full graph.

The evolution of connectivity within the ENSO region to a large extent copies that of the global graph evolution ($r = 0.77$) (see Fig. 8). However, the same is true, even to a higher extent, of the evolution of the extra-ENSO part of the network ($r = 0.93$).

Compared to the evolution of the ENSO network, the evolution of the whole tropical network captures even better the global graph evolution. On the other hand, the remaining part of the network shows a much weaker resemblance to the global network evolution. A comparison of the threshold evolutions for the full, tropical and extra-tropical graphs is shown in Fig. 9. Note that the full graph evolution is almost perfectly correlated with the tropical graph evolution ($r = 0.97$), while the correlation with the extra-tropical graph is much weaker ($r = 0.52$).

For the evolution of the other graph-theoretical properties, see Figs. 10, 11, 12 and 13.

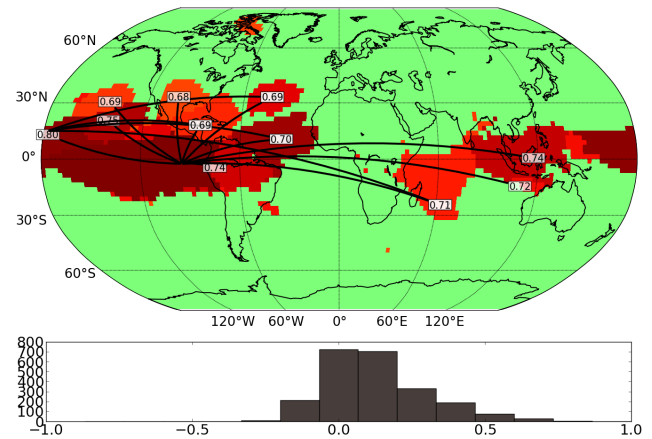


Fig. 7. Pairs of regions that show high relevance for the global graph evolution. Top panel: for fifteen region pairs with strongest correlation of inter-regional connectivity evolution with the global 2562-node network threshold evolution, a link is shown and the regions highlighted (darker red corresponds to the highest correlation the region is involved in, also given by the number written within the region).

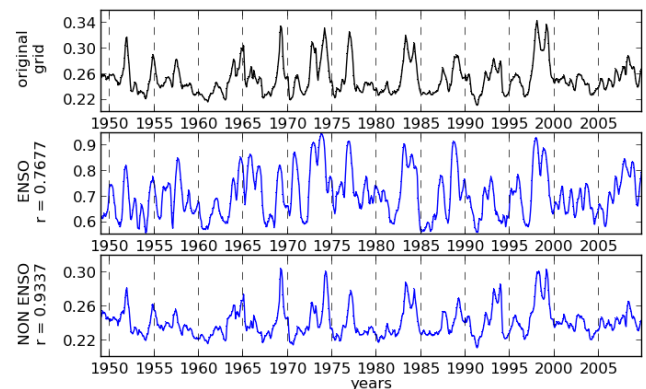


Fig. 8. The temporal evolution of the ENSO and extra-ENSO subgraph.

4 Discussion

4.1 Role of intra-regional and inter-regional connectivity in global graph evolution

Our analysis demonstrated that the global graph evolution reflects to a comparable extent the evolution of intra-regional and inter-regional connectivity. The contribution of the inter-regional connectivity evolution is clear from the comparison with the evolution of the component-based graph, which was able to explain up to 50 % of the global evolution variability (see Fig. 2).

The evolution of the component graph was most similar to the evolution of the full grid-based graph for low component-graph density and high grid-graph density. We suggest that particularly for higher densities, the global grid graph is more

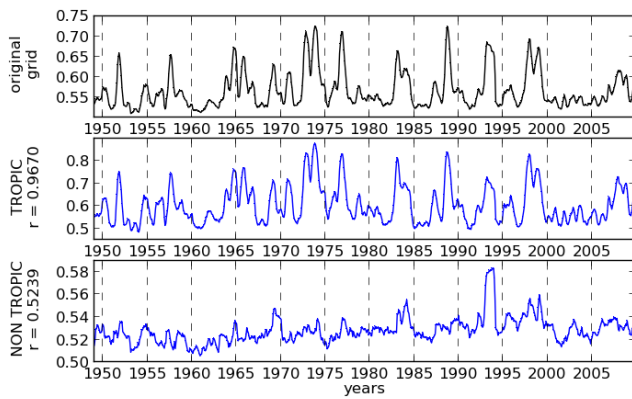


Fig. 9. The temporal evolution of the tropical and extra-tropical subgraph.

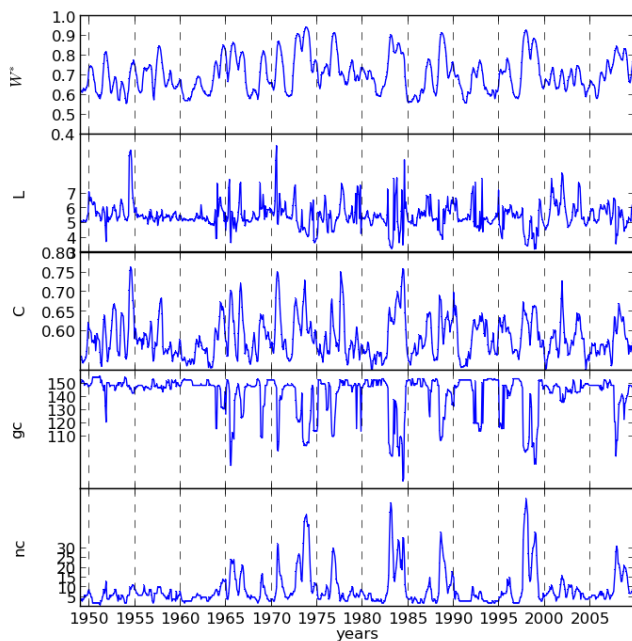


Fig. 10. Evolution of the graph-theoretical properties of the ENSO subgraph: threshold W^* , characteristic path length L , average clustering coefficient C , size of the giant component gc and number of connected components nc .

strongly determined by long-range connections and therefore more related to the component graph. For instance, the volcanic eruption of Mount Pinatubo in 1991, which had relatively localised effects and is not strongly reflected in the component-graph network, but which is clearly visible in the low-density grid-based graph as a peak in the years 1992–1993 (Radebach et al., 2013), but not so much for the higher densities (when the El Niño effect is much more dominant) (see Fig. 1).

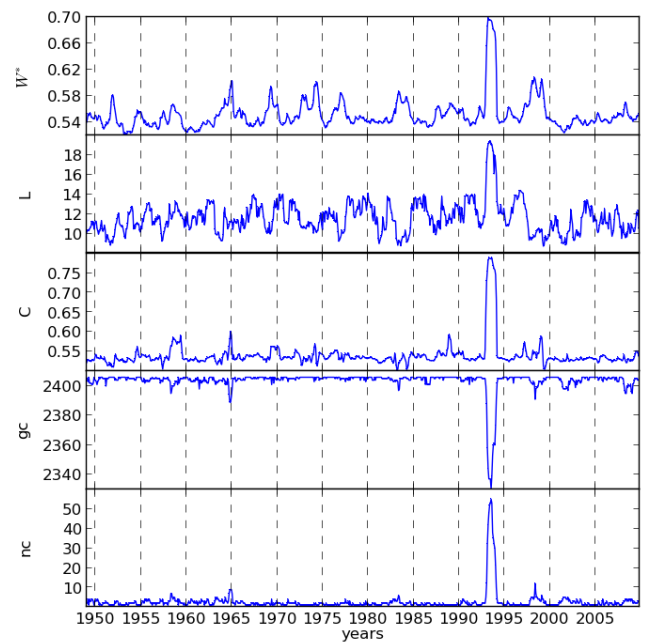


Fig. 11. Evolution of the graph-theoretical properties of the extra-ENSO subgraph.

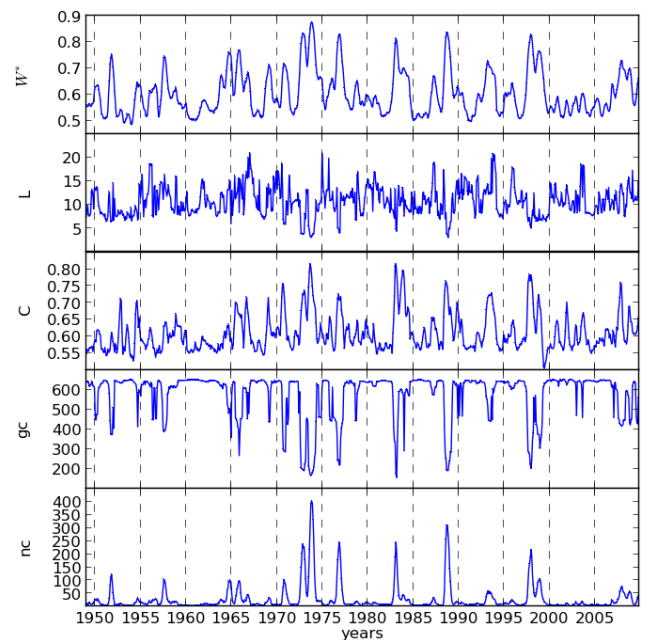


Fig. 12. Evolution of the graph-theoretical properties of the tropics subgraph.

4.2 Alternative grid size

Given the extent to which the global graph evolution reflects the dynamics of long-range links, it should not be crucially affected by spatial sub-sampling of the grid. To further explore this hypothesis, we have repeated the analysis for

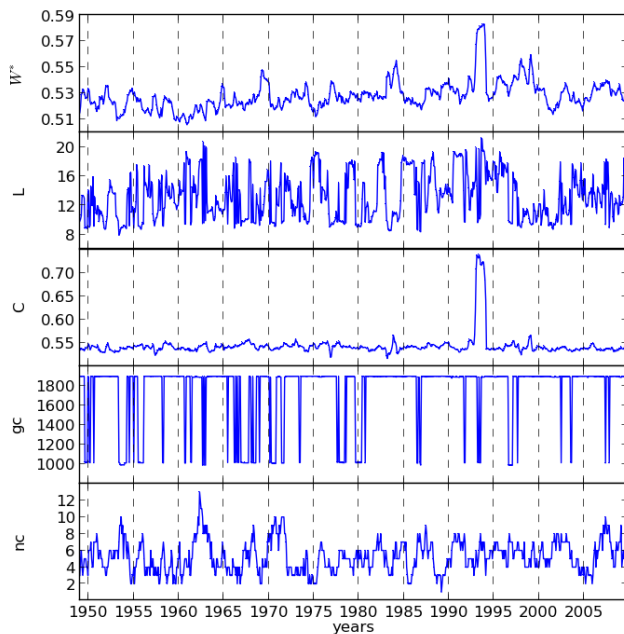


Fig. 13. Evolution of the graph-theoretical properties of the extra-tropics subgraph.

sparser spatial grids, with 642, 162 and 42 grid points instead of the original 2562. Note that particularly the last choice corresponds to a quite severe coarse-graining of the globe. The similarity of the threshold evolutions across a range of predefined densities is shown in Fig. 14. Note that the correlation is relatively high across a wide range of densities, although it decreases with increasing coarse-graining.

Figure 14 also demonstrates that the component graph with 68 nodes shows qualitatively different behaviour than even the sparse grids, in that it shows highest correlation when thresholded at very low densities, irrespective of the density used for the full 2562 grid. This suggests that the dimensionality reduction produced by the RPCA is qualitatively different from pure coarse-graining of the network; and the component network probably also reflects other climate phenomena than a coarsened grid-based network does.

4.3 Alternative thresholding for subgraphs

In the analysis reported above, the intra-regional and inter-regional networks have been thresholded to fixed pre-defined densities. However, the edges present in these graphs within certain time windows do not necessarily correspond to the edges present in the same window in the global graph thresholded at density 0.005. This motivates the use of an alternative thresholding strategy, where for each time window, the threshold of the global graph is used for all subgraphs. In this case, the time-dependent variable of interest for each region or region pair is the density (or number) of edges. Notably, as the total number of edges is fixed across time points, the

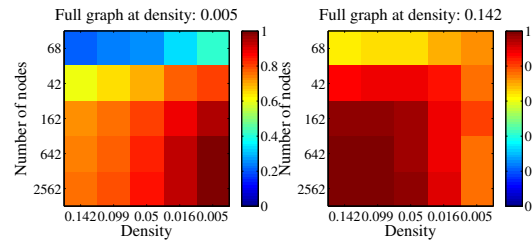


Fig. 14. Similarity of evolutions for varying grid size and density of the graphs compared to the original graph with 2562 nodes. Left panels: original graph density 0.005. Right panels: original graph density 0.142.

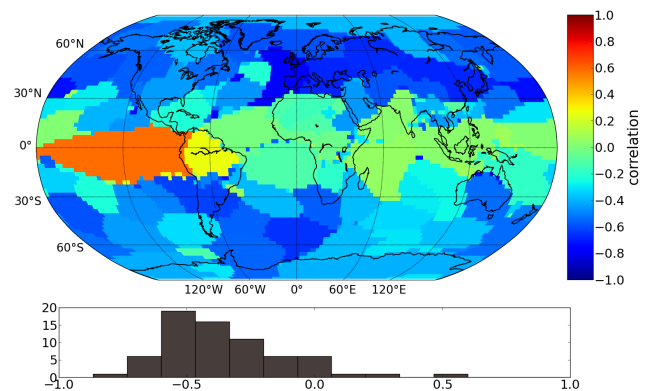


Fig. 15. Regions that show high relevance for the global graph evolution, when the intra-regional graph threshold at each time point is set equal to the global 2562-node grid graph with density 0.005. Visualisation as in Fig. 4.

densities of the subgraphs are mutually inevitably dependent, i.e. any increase in connectivity in some intra-regional or inter-regional graph must be compensated for by a decrease in density somewhere else.

The results of the corresponding analysis are shown in Figs. 15 and 16. Note that the prominent role of the ENSO region, and more generally, the whole tropical strip is largely unchanged; however, the correlations with the global evolution are shifted downwards, with many negative correlations, particularly for extra-tropical regions. This can be understood as a consequence of the intrinsic dependence due to this alternative thresholding strategy.

4.4 Non-locality of RPCA components

The RPCA network is based on correlations between the component time series. These time series represent a weighted average of all time series observed across the globe (with the weights corresponding to the component loadings), and therefore do not represent phenomena strictly localised in a particular region. To address this issue, one can use the RPCA-based crisp parcelation to define a new representative time series for each component as the average of all time

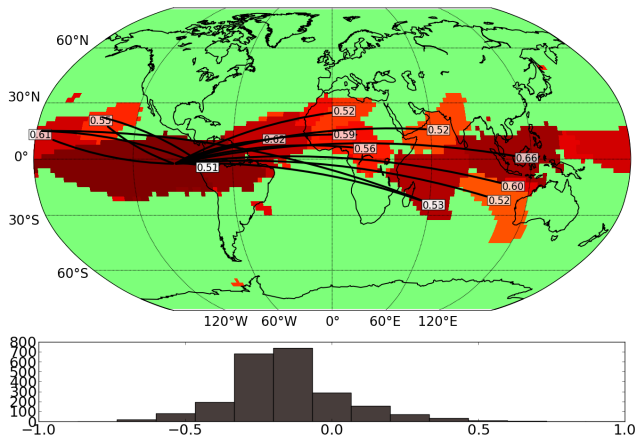


Fig. 16. Pairs of regions that show high relevance for the global graph evolution, when the inter-regional graph threshold at each time point is set equal to the global 2562-node grid graph with density 0.005. Visualisation as in Fig. 7.

series within the respective parcel. Time series defined in this way intuitively more clearly represent a given area due to no overlap among the parcels – this approach corresponds to defining climate indexes based on a spatial average. Importantly, the graph theoretical analysis based on such 68 time series qualitatively reproduces the results with component time series, although the component-graph evolution is more similar to that of sparse grids (results not shown).

4.5 Alternative region definition schemes

For definition of the regions in the current work, we have used VARIMAX-rotated PCA, endowed with a dimensionality-selection scheme based on statistical analysis of the eigenvalue spectra, a method used previously in Hlinka et al. (2013b). For more details about methods, we refer the reader to Vejmelka et al. (2014). Modern data analysis provides a range of other methods for dimensionality reduction by definition of regions with relatively homogeneous time series, of which several have been recently applied in or developed for analysis of climate time series in the context of graph-theoretical analysis (see e.g. Steinhäuser et al., 2012; Tsonis et al., 2011; Fountalis et al., 2013). Unlike Radebach et al. (2013), the authors do not study the detailed temporal evolution of the networks; however, their definition of parcels may in principle be used in an analysis analogous to the one presented here. A direct comparison of the parcelations is difficult to interpret meaningfully due to different data sets and assumptions behind the parcelation methods; however, somewhat promising is the fact that the number of separate parcels are relatively similar across the schemes. A comparison or optimisation of parcelation approaches would be a valuable avenue for further work in multi-scale climate network analysis; however, selection of optimality criteria and

dependence on parameter choice and data type or preprocessing may make it a very strenuous task.

5 Conclusions

The detailed analysis of the evolution of connectivity in the surface air temperature network has shown that the temporal changes of both localised and inter-regional connectivity are reflected in the global graph evolution. Apart from the important role of the ENSO region that was suggested earlier based on evidence including the relation to the El Niño and La Niña phases (Radebach et al., 2013; Tsonis and Swanson, 2008; Yamasaki et al., 2008), our detailed analysis provided evidence for the additional role of other tropical regions. These are relevant both through variations of their intra-regional connectivity and variations in the inter-regional connectivity, particularly between these regions and the ENSO region. The dominant role of the tropical region was further confirmed by evaluation of the evolution of the tropical and extra-tropical subnetworks.

Although the overall qualitative character of the global graph evolution was conserved across a range of graph densities used, the height of specific peaks and likely the relative importance of regional and intra-regional events depends on the graph density choice.

Generally, to reproduce the dynamics of the spatially well-sampled network at a given density, a similar density of the coarse-grained network is best used. However, the obtained RPCA-based climate networks reproduce even relatively dense grid-based networks better when thresholded very conservatively, i.e. to low density.

This suggests that the character of networks constructed using the RPCA dimension reduction scheme differs markedly from networks obtained by a simple spatial down-sampling. This phenomenon together with detailed analysis of network signatures of climate phenomena at specific scales are subjects of further research.

Acknowledgements. This study was supported by the Czech Science Foundation, project no. P103/11/J068, and by DFG grant no. KU34-1.

Edited by: M. Small

Reviewed by: S. Sultornsanee and an anonymous referee

References

- Albert, R., Jeong, H., and Barabási, A. L.: Internet – Diameter of the World Wide Web, *Nature*, 401, 130–131, 1999.
- Arenas, A., Diaz-Guilera, A., Kurths, J., Moreno, Y., and Zhou, C.: Synchronization in complex networks, *Phys. Reports*, 469, 93–153, 2008.
- Barabási, A.-L. and Albert, R.: Emergence of scaling in random network, *Science*, 286, 509–512, 1999.

- Bialonski, S. and Lehnertz, K.: Assortative mixing in functional brain networks during epileptic seizures, *Chaos*, 23, 033139, doi:10.1063/1.4821915, 2013.
- Bialonski, S., Horstmann, M.-T., and Lehnertz, K.: From brain to earth and climate systems: Small-world interaction networks or not?, *Chaos*, 20, 013134, doi:10.1063/1.3360561, 2010.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U.: Complex networks: Structure and dynamics, *Phys. Reports*, 424, 175–308, 2006.
- Bullmore, E. and Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems, *Nat. Rev. Neurosci.*, 10, 186–198, 2009.
- 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, 2009a.
- Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: Complex networks in climate dynamics, *Eur. Phys. J.*, 174, 157–179, 2009b.
- Faloutsos, M., Faloutsos, P., and Faloutsos, C.: On power-law relationships of the internet topology, *Comput. Commun. Rev.*, 29, 251–262, 1999.
- Fountalis, I., Bracco, A., and Dovrolis, C.: Spatio-temporal network analysis for studying climate patterns, *Clim. Dynam.*, 42, 879–899, doi:10.1007/s00382-013-1729-5, 2013.
- Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. J.: Functional Connectivity – The Principal-Component Analysis Of Large (PET) Data Sets, *J. Cereb. Blood Flow Metabol.*, 13, 5–14, 1993.
- Groth, A. and Ghil, M.: Multivariate singular spectrum analysis and the road to phase synchronization. *Phys. Rev. E*, 84, 036206, doi:10.1103/PhysRevE.84.036206, 2011.
- Guimera, R., Mossa, S., Turtschi, A., and Amaral, L. A. N.: The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles, *P. Natl. Acad. Sci. USA*, 102, 7794–7799, 2005.
- Hartman, D., Hlinka, J., Palus, 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.
- Heikes, R. and Randall, D. A.: Numerical Integration of the Shallow-Water Equations on a Twisted Icosahedral Grid, Part I: Basic Design and Results of Tests, *Mon. Weather Rev.*, 123, 1862–1880, 1995.
- Heitzig, J., Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: Node-weighted measures for complex networks with spatially embedded, sampled, or differently sized nodes, *Eur. Phys. J.*, 85, 1–22, doi:10.1140/epjb/e2011-20678-7, 2012.
- Hlinka, J., Paluš, M., Vejmelka, M., Mantini, D., and Corbetta, M.: Functional connectivity in resting-state fMRI: Is linear correlation sufficient?, *Neuroimage*, 54, 2218–2225, 2011.
- Hlinka, J., Hartman, D., and Paluš, M.: Small-world topology of functional connectivity in randomly connected dynamical systems, *Chaos*, 22, 033107, doi:10.1063/1.4732541, 2012.
- Hlinka, J., Hartman, D., Vejmelka, M., Novotná, D., and Paluš, M.: Non-linear dependence and teleconnections in climate data: sources, relevance, nonstationarity, *Clim. Dynam.*, doi:10.1007/s00382-013-1780-2, in press, 2013a.
- Hlinka, J., Hartman, D., Vejmelka, M., Runge, J., Marwan, N., Kurths, J., and Paluš, M.: Reliability of Inference of Directed Climate Networks Using Conditional Mutual Information, *Entropy*, 15, 2023–2045, 2013b.
- Holme, P. and Saramäki, J.: Temporal networks, *Physics Reports*, 519, 97–125, 2012.
- Jeong, H., Mason, S. P., Barabasi, A. L., and Oltvai, Z. N.: Lethality and centrality in protein networks, *Nature*, 411, 41–42, 2001.
- Jones, P. W.: A user's guide for SCRIP: A spherical coordinate remapping and interpolation package, Los Alamos National Laboratory, Los Alamos, USA, 1997.
- 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. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, *B. Am. Meteorol. Soc.*, 77, 437–471, 1996.
- Kuhnert, M.-T., Elger, C. E., and Lehnertz, K.: Long-term variability of global statistical properties of epileptic brain networks, *Chaos*, 20, 043126, doi:10.1063/1.3504998, 2010.
- Mantegna, R. N.: Hierarchical structure in financial markets, *Eur. Phys. J. B*, 11, 193–197, 1999.
- Newman, M. E. J.: The structure of scientific collaboration networks, *P. Natl. Acad. Sci. USA*, 98, 404–409, 2001.
- Newman, M. E. J.: The structure and function of complex networks, *SIAM Rev.*, 45, 167–256, 2003.
- Paluš, M., Hartman, D., Hlinka, J., and Vejmelka, M.: Discerning connectivity from dynamics in climate networks, *Nonlin. Processes Geophys.*, 18, 751–763, doi:10.5194/npg-18-751-2011, 2011.
- Radebach, A., Donner, R. V., Runge, J., Donges, J. F., and Kurths, J.: Disentangling different types of El Niño episodes by evolving climate network analysis, *Phys. Rev. E*, 88, 052807, doi:10.1103/PhysRevE.88.052807, 2013.
- Rosvall, M., Trusina, A., Minnhagen, P., and Sneppen, K.: Networks and cities: An information perspective, *Phys. Rev. Lett.*, 94, 028701, doi:10.1103/PhysRevLett.94.028701, 2005.
- Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., and Greicius, M. D.: Decoding Subject-Driven Cognitive States with Whole-Brain Connectivity Patterns, *Cerebral Cortex*, 22, 158–165, 2012.
- Steinhaeuser, K., Ganguly, A. R., and Chawla, N. V.: Multivariate and multiscale dependence in the global climate system revealed through complex networks, *Clim. Dynam.*, 39, 889–895, 2012.
- Tsonis, A. A. and Roebber, P. J.: The architecture of the climate network, *Physica A*, 333, 497–504, 2004.
- Tsonis, A. A. and Swanson, K. L.: Topology and predictability of El Niño and La Niña networks, *Phys. Rev. Lett.*, 100, 228502, doi:10.1103/PhysRevLett.100.228502, 2008.
- Tsonis, A. A., Wang, G., Swanson, K. L., Rodrigues, F. A., and Costa, L. D. F.: Community structure and dynamics in climate networks, *Clim. Dynam.*, 37, 933–940, 2011.
- Vejmelka, M. and Paluš, M.: Partitioning Networks into Clusters and Residuals with Average Association, *Chaos*, 20, 033103, doi:10.1063/1.3460360, 2010.
- Vejmelka, M., Pokorná, L., Hlinka, J., Hartman, D., Jajcay, N., and Paluš, M.: Non-random correlation structures and dimensionality reduction of multivariate climate data, *Clim. Dynam.*, in review, 2014.

- Wasserman, S. and Faust, K.: Social Network Analysis, Cambridge University Press, 1994.
- Watts, D. J. and Strogatz, S. H.: Collective dynamics of “small-world” networks, *Nature*, 393, 440–442, 1998.
- Yamasaki, K., Gozolchiani, A., and Havlin, S.: Climate networks around the globe are significantly affected by El Nino, *Phys. Rev. Lett.*, 100, 228501, doi:10.1103/PhysRevLett.100.228501, 2008.