Submitted to *Entropy*. Pages 1 - 20.

OPEN ACCESS

Article

Reliability of inference of directed climate networks using conditional mutual information

Jaroslav Hlinka [1,⋆], David Hartman [1], Martin Vejmelka [1], Jakob Runge [2,3], Norbert Marwan [2], Jürgen Kurths [2,3,4] and Milan Paluš [1]

[1] Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodarenskou vezi 2, 182 07, Prague 8, Czech Republic

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

[3] Department of Physics, Humboldt University, 12489 Berlin, Germany

[4] Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB24 3UE, United Kingdom

* Author to whom correspondence should be addressed; hlinka@cs.cas.cz, (+420) 266053808

Version January 29, 2013 submitted to Entropy. *Typeset by LaTeX using class file mdpi.cls*

Abstract: Across geosciences, many investigated phenomena relate to specific complex systems consisting of intricately intertwined interacting subsystems. Such dynamical complex systems can be represented by a directed graph, where each link denotes an existence of a causal relation, or information exchange between the nodes. For geophysical systems such as global climate, these relations are commonly not known theoretically but estimated from recorded data using causality analysis methods. These include bivariate nonlinear methods based on information theory and their linear counterpart. A trade-off between the valuable sensitivity of nonlinear methods to more general interactions and potentially higher numerical reliability of linear method may affect inference regarding structure and variability of climate networks. We investigate the reliability of directed climate networks detected by selected methods and parameter settings, using stationarized model of dimensionality-reduced surface air temperature data from reanalysis of 60-year global climate records. Overall, all studied bivariate causality methods provided reproducible estimates of climate causality networks; with linear approximation showing higher reliability than the investigated nonlinear methods. On the example dataset, optimizing the investigated nonlinear methods with respect to reliability increased similarity of the detected networks to their linear counterparts, supporting the particular hypothesis of surface air temperature climate reanalysis data near-linearity.

Keywords: causality; climate; nonlinearity, transfer entropy, network, stability

1. Introduction

Across geosciences, many investigated phenomena relate to specific complex systems consisting of intricately intertwined interacting subsystems. These can be suitably represented as networks, an approach that is gaining increasing attention in complex systems community [1,2]. The meaning of the existence of a link between nodes of a network depends on the area of application, but in many cases it is related to some form of information exchange between the nodes.

This approach has already been adopted for the analysis of various phenomena in the global climate system [3–7]. Typically, a graph is constructed by considering two locations linked by a connection, if there is an instantaneous dependence between the localized values of a variable of interest.

This dependence can be conveniently quantified by mutual information - an entropy-based general measure of statistical dependence that takes into account nonlinear contributions to the coupling. In practice, for reasons of theoretical and numerical simplicity, linear Pearson's correlation coefficient might be sufficient, although potentially negliging the nonlinear contributions to interactions. In particular, while initial works by Donges et al. stressed the role of mutual information in detecting important features of global climate networks [8,9], more detailed recent work has shown that the differences between correlation and mutual information graphs are mostly (but not necessarily completely) spurious, such as due to natural and instrumental (related to data collection) nonstationarities of the data [10].

However, these methods do not allow to assess the directionality of the links and of the underlying information flow. This motivates the use of more sophisticated measures, known also as causality analysis methods.

The family of causality methods include linear approaches such as the Granger causality analysis [11] as well as more general nonlinear methods. A prominent representative of nonlinear causality assessment method is the conditional mutual information [12] known also as transfer entropy [13].

Arguably, the nonlinear methods, due to their model-free nature, have the theoretical advantage of being sensitive to forms of interactions that linear methods may detect only partially or not at all. On the other side, this advantage might be more than outweighed by a potentially lower precision. Depending on specific circumstances, this may adversely affect the reliability of detection of network patterns.

Apart from uncertainty about the general network pattern, reliability is important when the interest is in detecting changes in time, with the need to distinguish them from random variance of the estimates among different sections of time series under investigation - a task that is relevant in many areas of geoscience including climate research.

In other words, before analyzing a complex dynamical system using network theory, a key initial question is that of the reliability of the network construction, and of its dependence on the causality method choice and settings.

We study this question for a selection of standard causality methods, using a timely application in the study of climate network and its variability. In particular, surface air temperature data from the NCEP/NCAR reanalysis dataset [14,15] are used. The original data contain more than 10,000 time series - a relatively dense grid covering the whole globe. For efficient computation and visualization of the results, it is convenient to reduce the dimensionality of the data. We use principal component analysis and select only components that have significantly high explained variance compared to corresponding spatially independent but temporally dependent (i.e. 'colored') random noise.

As the causality network construction reliability may crucially depend on the specific choice of the causality estimator, we quantitatively assess the effect of choice of different causality measures and their parametrization.

We assess the network construction reliability by quantifying the similarity of causality matrices reconstructed from independent realization of a stationary model of data. These realizations are either independently generated, or they represent individual non-overlapping temporal windows in a single stationary realization. Optimal parameter choice of the applied nonlinear methods is detected, and the reliability of networks constructed using linear and nonlinear methods compared.

The latter method, i.e. comparing networks reconstructed from temporal windows, allows to assess the network variability on real data and compare it with variability on the stationary model.

2. Data and Methods

2.1. Causality assessment methods

2.1.1. Granger causality analysis

A prominent method for assessing causality is so-called Granger causality analysis, named after Sir Clive Granger, who proposed this approach to time series analysis in a classical paper [11]. However, the basic idea can be traced back to Wiener [16], who proposed that if the prediction of one time series can be improved by incorporating the knowledge of a second time series, then the latter can be said to have a causal influence on the former. This idea was formalized by Granger in the context of linear regression models. In the following, we outline the methods of assessment of Granger causality, following the description given in [17] and [18,19].

Consider two stochastic processes X_t and Y_t and assume they are jointly stationary. Let further the autoregressive representations of each process be:

$$X_t = \sum_{j=1}^{\infty} a_{1j} X_{t-j} + \epsilon_{1t}, \qquad \mathrm{var}(\epsilon_{1t}) = \Sigma_1, \tag{1}$$

$$Y_t = \sum_{j=1}^{\infty} d_{1j} Y_{t-j} + \eta_{1t}, \qquad \mathrm{var}(\eta_{1t}) = \Gamma_1, \tag{2}$$

and the joint autoregressive representation be:

$$X_t = \sum_{j=1}^{\infty} a_{2j} X_{t-j} + \sum_{j=1}^{\infty} b_{2j} Y_{t-j} + \epsilon_{2t}, \tag{3}$$

$$Y_t = \sum_{j=1}^{\infty} c_{2j} X_{t-j} + \sum_{j=1}^{\infty} d_{2j} Y_{t-j} + \eta_{2t}, \tag{4}$$

where the covariance matrix of the noise terms is:

$$\Sigma = \mathrm{Cov}\begin{pmatrix} \epsilon_{2t} \\ \eta_{2t} \end{pmatrix} = \begin{pmatrix} \Sigma_2 & \Lambda_2 \\ \Lambda_2 & \Gamma_2 \end{pmatrix}. \tag{5}$$

The causal influence from Y to X is then quantified based on the decrease in the residual model variance when we include the past of Y in the model of X, i.e. when we move from the independent model given by Equation (1) to the joint model given by Equation (3):

$$F_{Y\to X} = \ln \frac{\Sigma_1}{\Sigma_2}. \tag{6}$$

Similarly, the causal influence from X to Y is defined as:

$$F_{X\to Y} = \ln \frac{\Gamma_1}{\Gamma_2}. \tag{7}$$

Clearly, the causal influence defined in this way is always nonnegative.

The original introduction of the concept of statistical inference of causality [11] includes a third (potentially highly multivariate) process Z, representing all the other intervening process that should be controlled for in assessing the causality between X and Y. The bivariate (or 'pairwise') implementation of the estimator thus constitutes a computational cimplification of the original process, for the sake of numerical stability as well as comparability to the bivariate transfer entropy (conditional mutual information) approach introduced later. See section 4 for further discussion of related issues.

2.2. Estimation of GC

Practical estimation of the Granger causality involves fitting the full and depleted models described above to experimental data. While the theoretical framework outlined above is formulated in terms of infinite sums, the fitting procedure requires selection of the model order p for the models. For our report, we have selected $p = 1$ to allow direct comparability of the Granger causality analysis to the nonlinear methods considered later. This choice is the most common choice for Granger causality in literature and amounts to looking for links with lag 1 time unit.

2.3. Transfer entropy

To provide a framework for discussion of the related issues, it is useful to consider that for a general bivariate stochastic process the Granger causality concept, can be captured in information-theoretic terms. In particular, we can define that X causes Y if the knowledge of past of X decreases the uncertainty about Y (above what the knowledge of past of Y and potentially all other relevant confounding variables already informs). This simple concept is captured in the definition of *transfer entropy* (TE, [13]). TE as can be defined in terms of *conditional mutual information* as shown below, following closely [12].

For two discrete random variables X, Y with sets of values Ξ and Υ and probability distribution functions (PDFs) $p(x), p(y)$ and joint PDF $p(x, y)$, the Shannon entropy $H(X)$ is defined as

$$H(X) = -\sum_{x\in\Xi} p(x) \log p(x), \tag{8}$$

and the joint entropy $H(X, Y)$ of X and Y as

$$H(X, Y) = -\sum_{x\in\Xi}\sum_{y\in\Upsilon} p(x, y) \log p(x, y). \tag{9}$$

The conditional entropy $H(X|Y)$ of X given Y is

$$H(X|Y) = -\sum_{x\in\Xi}\sum_{y\in\Upsilon} p(x, y) \log p(x|y). \tag{10}$$

The amount of common information contained in the variables X and Y is quantified by the mutual information $I(X; Y)$ defined as

$$I(X; Y) = H(X) + H(Y) - H(X, Y). \tag{11}$$

The conditional mutual information $I(X; Y|Z)$ of the variables X, Y given the variable Z is given as

$$I(X; Y|Z) = H(X|Z) + H(Y|Z) - H(X, Y|Z). \tag{12}$$

Entropy and mutual information are measured in bits if the base of the logarithms in their definitions is 2. It is straightforward to extend these definitions to more variables, and to continuous rather than discrete variables.

Transfer entropy from process X_t to process Y_t then corresponds to the conditional mutual information between X_t and Y_{t+1} conditional on Y_t:

$$T_{X\to Y} = I(X_t, Y_{t+1}|Y_t). \tag{13}$$

While the definition of these information-theoretic functionals describing dependence structure between variables is very general and elegant, the practical estimation faces challenges related to the problem of efficient estimation of the PDF of the studied variables from samples of finite size. For the further considerations, it is important to bear in mind the distinction between the true quantities of the underlying stochastic process, and their finite-sample estimators.

2.4. Potential causes of observed difference

Interestingly, it can be shown that for linear Gaussian processes, transfer entropy is equivalent to linear Granger causality, up to a multiplicative factor [20]:

$$\mathcal{T}_{X\to Y} = \frac{1}{2}\mathcal{F}_{X\to Y}. \tag{14}$$

However, in practice, the estimates of transfer entropy and linear Granger causality may differ. There are principally two main reasons for this divergence between the results. Firstly, when the underlying process is not linear Gaussian, the true transfer entropy may differ from the true linear Granger causality corresponding to the linear approximation of the process. A second reason for divergence between sample estimates of transfer entropy and linear Granger causality, valid even for linear Gaussian processes, is the difference in the properties of the estimators of these two quantities, in particular bias and variance of the estimates.

2.5. TE estimation

There are many algorithms for estimation of information-theoretical functionals, that can be adapted to compute transfer entropy estimates. Two basic classes of nonparametric methods for the estimation of conditional mutual information are the *binning methods* and the *metric methods*. The former discretize the space in to regions usually called bins or boxes - a robust example is the equiquantal method, based on discretization of studied variables into Q equiquantal bins (EQQ, [21]). In the latter methods, the probability distribution function estimation depends on distances between the samples computed using some metric. An example of a metric method is the k-nearest neighbor (kNN, [12]) algorithm. For more detail on methods of estimation of conditional mutual information and their comparison see [12].

Note that both these algorithms require setting an additional parameter. While some heuristic suggestions have been published in the literature, the suitable values of the parameters may depend on specific aspects of the application including the character of the time series. For the purpose of this study, we use a range of parameter values and subsequently select the parameter values providing the most stable results for further comparison with linear methods, see below.

2.6. Data

2.6.1. Dataset

Data from the NCEP/NCAR reanalysis dataset [14] have been used. In particular, we utilize the time series $x_i(t)$ of the daily and monthly mean surface air temperature from January 1948 to December 2007 ($T_d = 21900$ and $T_m = 720$ time points respectively), sampled at latitudes λ_i and longitude ϕ_i forming a regular grid with a step of $\Delta\lambda = \Delta\phi = 2.5°$. The points located at the globe poles have been removed, giving a total of $N = 10224$ spatial sampling points.

2.6.2. Preprocessing

To minimize the bias introduced by periodic changes in the solar input, the mean annual cycle is removed from the data to produce so-called anomaly time series. The data were further standardized so that the time series at each grid point has unit variance. The time series are then scaled by the cosine of the latitude to account for grid points closer to the poles representing smaller areas and being closer together (thus biasing the correlation with respect to grid points farther apart). The poles are thus omitted entirely by effectively removing data for latitude ± 90.

2.6.3. Computing the components

The covariance matrix of the scaled time series obtained by preprocessing is computed. Note that this covariance matrix is equal to the correlation matrix, where each correlation is scaled by the inverse of the product of the cosines of the latitudes of the time series entering the correlation.

Next, the eigendecomposition of the covariance matrix is computed. The eigenvectors corresponding to genuine components are extracted (estimation of the number of components is explained in the next paragraph). The eigenvectors are then rotated using the VARIMAX method [22].

The rotated eigenvectors are the resulting components. Each component is represented by a scalar field of intensities over the globe, and by a corresponding representative time series. See Figure 6 for a parcellation of the globe by the components. For each location, the color corresponding to the component with maximal intensity is used - due to good spatial localization and smoothness of the components this leads to parcellation of the globe into generally contiguous regions.

2.6.4. Estimating the dimensionality of the data

To reduce the dimensionality, only a subset of the components is selected for further analysis. The main idea rests in determining significant components by comparing the eigenvalues computed from the original data to eigenvalues computed from a control dataset corresponding to the null hypothesis of uncoupled time series with the same temporal structure as the original data. To accomplish this, the time series in the control datasets are generated as realizations of autoregressive (AR) models fit to each time series independently. The dimension of the AR process is estimated for each time series separately using the Bayesian Information Criterion [23].

This model is used to generate 10000 realizations in the control dataset. The eigendecomposition if each realization is computed and aggregated so a distribution of each eigenvalue (1st, 2nd, ...) is available under the null hypothesis.

Finding the significant eigenvalues then reduces to a multiple comparison problem which we resolved using the False Discovery Rate (FDR) technique [24] which has led to the identification of 67 genuine components.

For computational reasons, the decomposition was carried out on the monthly data and the component spatial distributions were used to extract daily time series from correspondingly preprocessed daily data (anomalization, standardization, cosine transform). The method thus yielded full-resolution component localization on the 10224-point grid while also providing a high-resolution time series associated with each component.

2.7. Network construction

Formally, in the graph-theoretical approach a network is represented by an graph $G = (V, E)$, where V is the set of nodes of G, $n = \#V$ is the number of nodes and $E \subset V^2$ is the set of the edges (or links) of G. In weighted graphs, each edge connecting nodes i and j can be assigned a weight $a_{i,j}$ representing the strength of the link. Thus, the causality matrix $\mathcal{T}$ having as its entries the pairwise causalities $\mathcal{T}_{i,j} = \mathcal{T}_{X_i \to X_j}$ can be understood as a weighted graph with variable strength of links. Commonly, the graph is transformed into an unweighted matrix by suitable thresholding, keeping only links with weights higher than some threshold (and setting their weights to 1), while removing all the weaker links (setting the weights to zero).

There are three principal strategies to choose the threshold - either a fixed value based on expert judgment of what constitutes a strong link, or adaptively to enforce a required density of the graph (relative number of links with respect to the maximum number possible, i.e. in a full graph of given size). The third option is to use statistical testing to detect statistically significant links.

In the current paper, we start with the original unthresholded graphs, but provide also example results for thresholded graphs using the above named approaches.

2.8. Reliability assessment

In line with the terminology of psychometrics or classical test theory, by reliability we mean the overall consistency of a measure (consistency here not meant in the statistical sense of asymptotic behavior). In the context of network construction, we considered a method reliable if the networks constructed by its means from different samples of the same dynamical process would be similar to each other. Note that this does not necessarily imply validity or accuracy of the method - under some circumstances, a method could consistently arrive at wrong results. In a way, reliability/consistency can be considered a first step to validity. In practical terms, even if the validity was undoubted, reliability can give the researcher an estimate on the confidence he/she can have in reproducibility of the results.

To assess similarity of two matrices, many methods are available, including (entry-wise) Pearson's linear correlation coefficient. Inspection of the causality matrices suggests heavily non-normal distribution of the values with many outliers. Therefore, the correlation of ranks, using Spearman's correlation coefficient, may be more suitable.

Apart from reliability of the full weighted causality graphs, we also study the unweighted graphs derived by thresholding. Based on inspection of the causality matrices, a density of 0.01 (keeping 1 percent of strongest links) was chosen for the analysis.

To assess the similarity of two binary matrices, we use the Jaccard similarity coefficient. This is the relative number of links that are shared by the matrices with respect to the total number of links that appear at least in one of the matrices. Such a ratio is a natural measure of matrix overlap, ranging from 0 for matrices with no common links to 1 for identical matrices.

2.8.1. Model

A convenient method of assessing the reliability of a method on time series is to compare the results obtained on different sections (temporal windows) of that time series. However, dissimilarity among the results can be theoretically attributed both to lack of reliability of the method as well as to hypothetical true changes in the underlying systems over time (nonstationarity).

Therefore, to isolate the effect of method properties, we test the methods on a realistic, but stationary model of the data.

To provide such a stationary model of the potentially non-stationary data, a so-called surrogate data was constructed.

Technically, the surrogate data are conveniently constructed as multivariate Fourier transform (FT) surrogates [25,26]; i.e. obtained by computing the Fourier transform of the series, keeping unchanged the magnitudes of the Fourier coefficients (the amplitude spectrum), but adding the same random number to the phases of coefficients of the same frequency bin; the inverse FT into the time domain is then performed.

The surrogate data represent a realization of a linear stationary process conserving the linear structure (covariance and autocovariance) of the original data, and hence also the linear component of causality. Note that any nonlinear component of causality should be removed, and the nonlinear methods should therefore converge to the linear ones (as discussed in section 2.1).

After testing the reliability on the stationary linear model, we assess the stability of the methods also on the real data. The variability here should reflect a mixture of method in-reliability and true climate changes. Note that also the nonlinear methods may potentially diverge from the linear, as there may be strictly nonlinear component of the causalities in real data.

2.8.2. Implementation details

For estimation, both the stationary model and real data time series were split into 6 windows (one for each decade, i.e. with approximately 3650 time points). For each of the windows, causality matrix has been computed with several causality methods.

In particular, we have used pairwise Granger causality as a representative linear method, and conditional mutual information (transfer entropy) computed by two standard algorithms, using a range of critical parameter values. The first is an algorithm based on discretization of studied variables into Q equiquantal bins (EQQ, [21], $Q \in \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}$) and the second is a k-nearest neighbor algorithm (kNN, [12], $k \in \{2, 4, 8, 16, 32, 64, 128, 256, 512\}$).

Each of these algorithms provides a matrix of causality estimates among the 67 climate components within the respective decade. We further assess the similarity of these matrices both across time and methods; first in stationary data (where temporal variability is attributable to method instability only) and subsequently in real data. Apart from direct visualization, the similarity of constructed causality matrices is quantified by the Spearman's rank correlation coefficient of off-diagonal entries. The reliability is then estimated as the average Spearman's rank correlation coefficient across all $(6 * 5)/2 = 15$ pairs of temporal windows.

To inspect the robustness of the results, the analysis was repeated with several possible alterations to the paradigm. Firstly, the similarity among the thresholded rather than unweighted graphs was assessed by the means of the Jaccard similarity coefficient instead of Spearman's rank correlation coefficient. Secondly, we repeated the analysis using linear multivariate AR(1) process for generation of the stationary model, instead of Fourier surrogates. Thirdly, the analysis was repeated on subsampled data (by averaging each 6 days to give one data point). This way, the same methods should provide causality on a longer time scale. To keep the same (and sufficiently high) number of time points, the subsampled data were not split into windows, but 6 realizations were generated from a fitted multivariate AR(1) process.

Finally, to assess the role the different reliability of the methods might have for statistical inference, for selected methods we have tested the number of links that it was able to statistically distinguish compared to an empirical distribution corresponding to independent linear processes. This null hypothesis was realized by computing the causalities on a set of N=19 univariate Fourier surrogates. Under the hypothesis of no dependence between the processes, the probability of the data causality value for a given pair of variables being the highest from the total 20 values available is $p = 0.05$, providing a convenient nonparametric test of causality.

3. Results

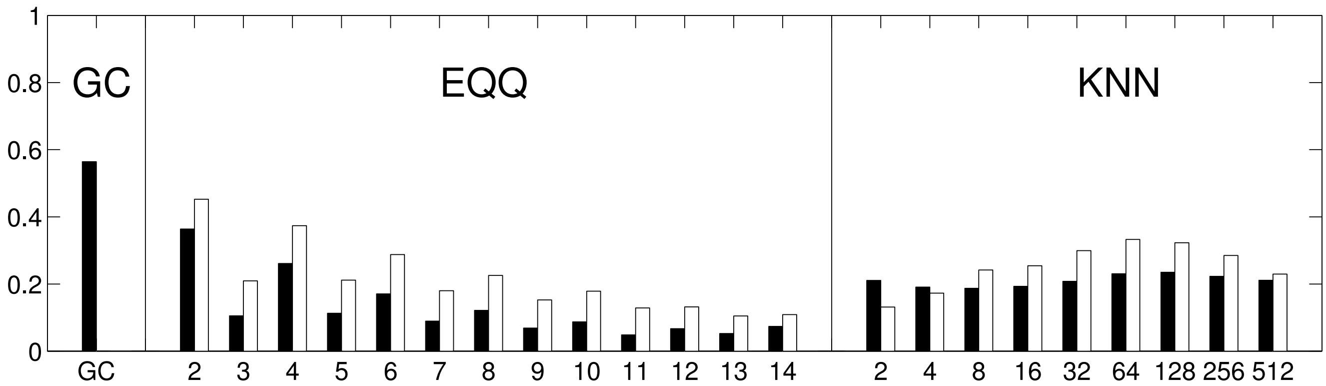
3.1. Weighted causality networks

The reliability of weighted causality networks computed from a decade of stationaty model data is shown in Figure 1 (for all methods and parameter values), along with the average similarity of the nonlinear network estimate by each method with the one obtained for the linear Granger causality method. The linear Granger causality shows the highest reliability, with the average Spearman's rank coefficient ~ 0.6. The equiquantal binning method provided most reliable network estimates for $Q = 2$ ($\bar{r} \sim 0.36$), with reliability generally decreasing for increasing Q. The k-nearest neighbors algorithm provided even less reliable network estimates, with only weak dependence on the values of the k-parameter and optimum reliability of $\bar{r} \sim 0.33$ for $k = 64$.

The causality networks constructed by each nonlinear method have been compared to the causality network obtained using the linear Granger causality analysis, see white bars Figure 1. In general, the nonlinear causality networks have shown higher similarity to linear estimates than to nonlinear estimates for different section of the stationary model time series. Interestingly, the parameter settings that optimized the reliability also provided the (almost) closest results to the linear methods. We have also observed generally lower reliability of the EQQ method for odd Q-values, an effect that will be investigated in detail elsewhere.

Figure 2 shows the results of an analogous analysis on original data rather than the stationary model. Note that here the computed causality network similarities reflect a combination of (lack of) reliability of the methods and real variability in the dynamical properties of the time series across time (i.e. true changes in the causality pattern). The results are both qualitatively and quantitatively similar to those

Figure 1. Reliability of causality network detection using different causality estimators, and the similarity to linear causality network estimates: Fourier surrogates model. For each estimator, six causality networks are estimated, one for each decade-long section of model stationary data (a Fourier surrogate realization of the original data). Black: the height of the bar corresponds to the average Spearman's correlation across all 15 pairs of decades. White: the height of the bar corresponds to the average Spearman's correlation of nonlinear causality network and linear causality network across 6 decades.



shown in Figure 1, suggesting that the true variability of the causal networks on this time scale is likely rather small compared to the coarseness of the causality assessment methods.

The results for other settings are shown in Figure 3 (use of multivariate AR(1) as the stationary model) and Figure 4 (6-day averages), generally confirming the main observations. However, some differences were observable, for instance in the 6-days-averaged data, the reliability dependence of the kNN method on the k-parameter was more pronounced and peak for a higher value of $k = 256$. The increase of reliability of the EQQ method for high Q was found to be spurious and is discussed in section 4.

3.2. Unweighted causality networks

For unweighted causality networks, the after thresholding to keep 1 percent of the strongest links, the network similarity was assessed by the Jaccard correlation coefficient. The results are plotted analogously as in the previous figures, see 5.

3.3. Components and resulting networks

To visualize the climatic networks, we first provide an overview of the localization of the networks in Figure 6 showing a parcellation of the globe by the components.

As an example of causality networks detected, we provide the networks detected by the linear Granger causality (Figure 7) and the EQQ with $Q = 2$ (Figure 8) for the decade 1948-1957.

4. Discussion

Figure 2. Variability of causality network detection using different causality estimators, and the similarity to linear causality network estimates: original data. For each estimator, six causality networks are estimated, one for each decade of the data. Black: the height of the bar corresponds to the average Spearman's correlation across all 15 pairs of decades. White: the height of the bar corresponds to the average Spearman's correlation of nonlinear causality network and linear causality network across 6 decades.

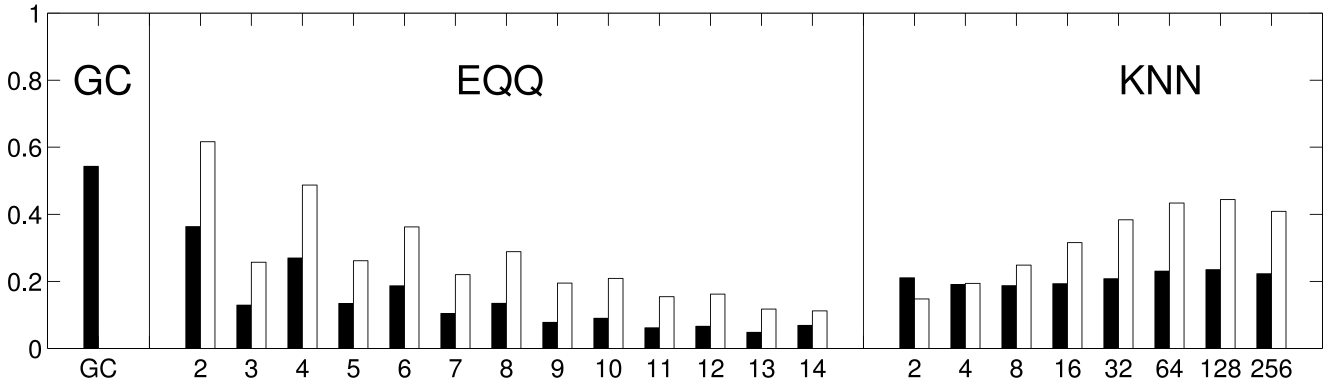


Figure 3. Reliability of causality network detection using different causality estimators, and the similarity to linear causality network estimates, for stationary model constructed as multivariate AR(1) surrogate of the original data. For each estimator, six causality networks are estimated, one for each decade of modeled stationary data. Black: the height of the bar corresponds to the average Spearman's correlation across all 15 pairs of decades. White: the height of the bar corresponds to the average Spearman's correlation of nonlinear causality network and linear causality network across 6 decades.

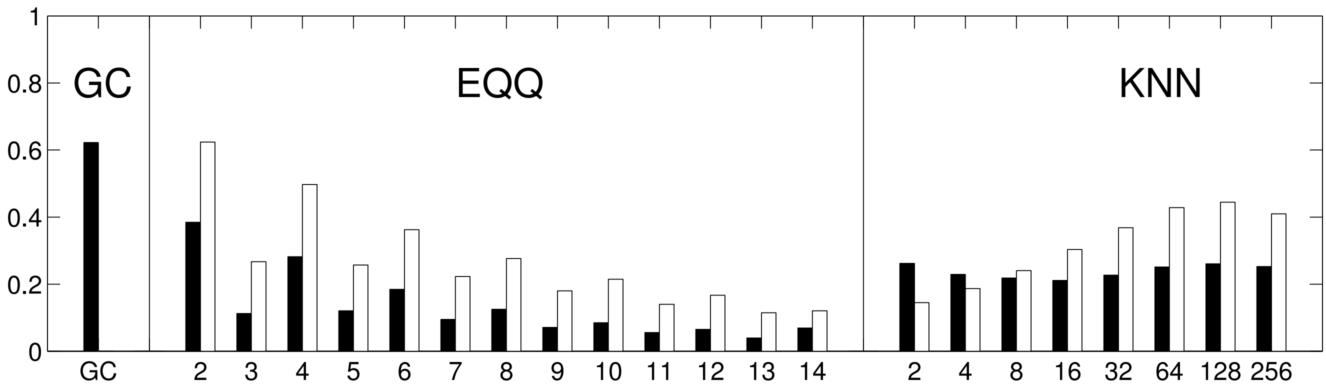


Figure 4. Reliability of causality network detection using different causality estimators, and the similarity to linear causality network estimates, for stationary model constructed as multivariate AR(1) surrogate of the original data. For each estimator, six causality networks are estimated, each for a separate realization of the multivariate AR(1) process fitted to the original data. Black: the height of the bar corresponds to the average Spearman's correlation across all 15 pairs of decades. White: the height of the bar corresponds to the average Spearman's correlation of nonlinear causality network and linear causality network across 6 decades.

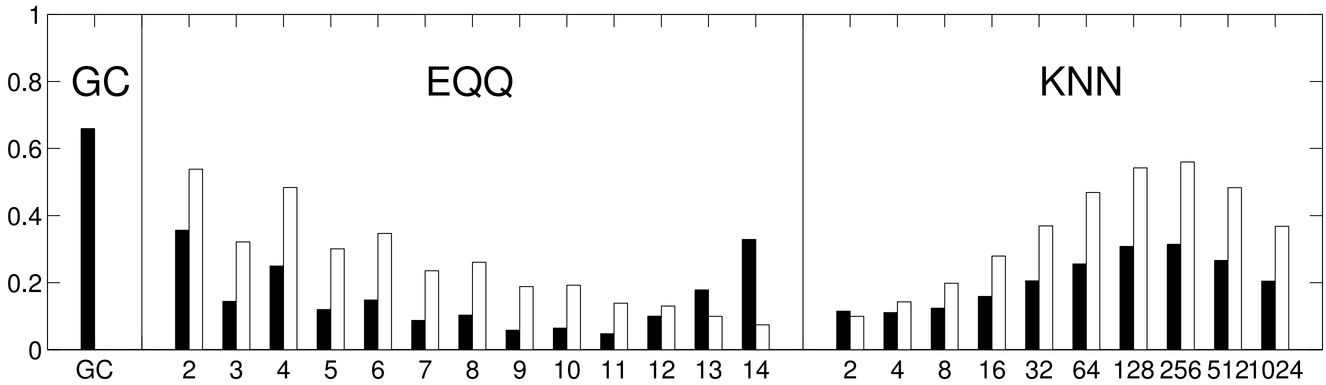


Figure 5. Reliability of causality network detection using different causality estimators, and the similarity to linear causality network estimates. For each estimator, six causality networks are estimated, one for each decade of modeled stationary data. Black: the height of the bar corresponds to the average Jaccard similarity coefficient across all 15 pairs of decades. White: the height of the bar corresponds to the average Jaccard similarity coefficient of nonlinear causality network and linear causality network across 6 decades.

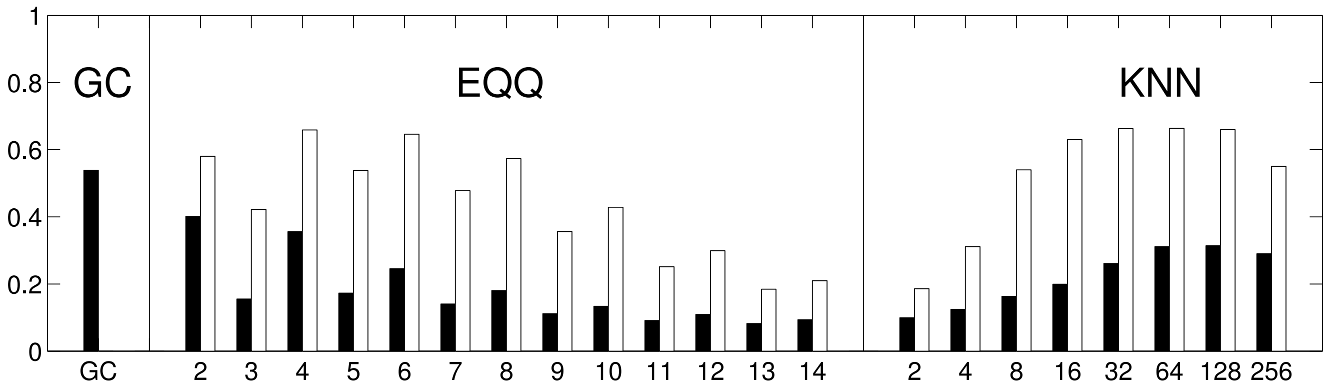


Figure 6. Location of areas dominated by specific components of the climate surface air temperature data VARIMAX-rotated PCA decomposition. For each location, the color corresponding to the component with maximal intensity it used. White dots represent approximate centers of mass of the components, used in subsequent figures for visualization of the nodes of the networks.

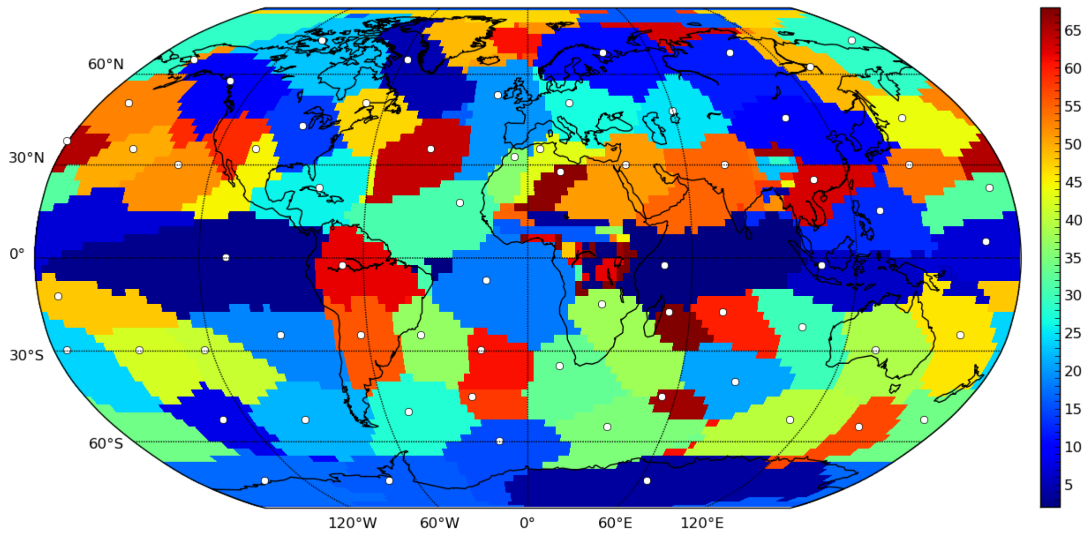


Figure 7. Example of detected causality network, detected by the linear Granger causality for the decade 1948-1957. Links with $\mathcal{T}_{X\rightarrow Y} \geq 0.02$ shown.

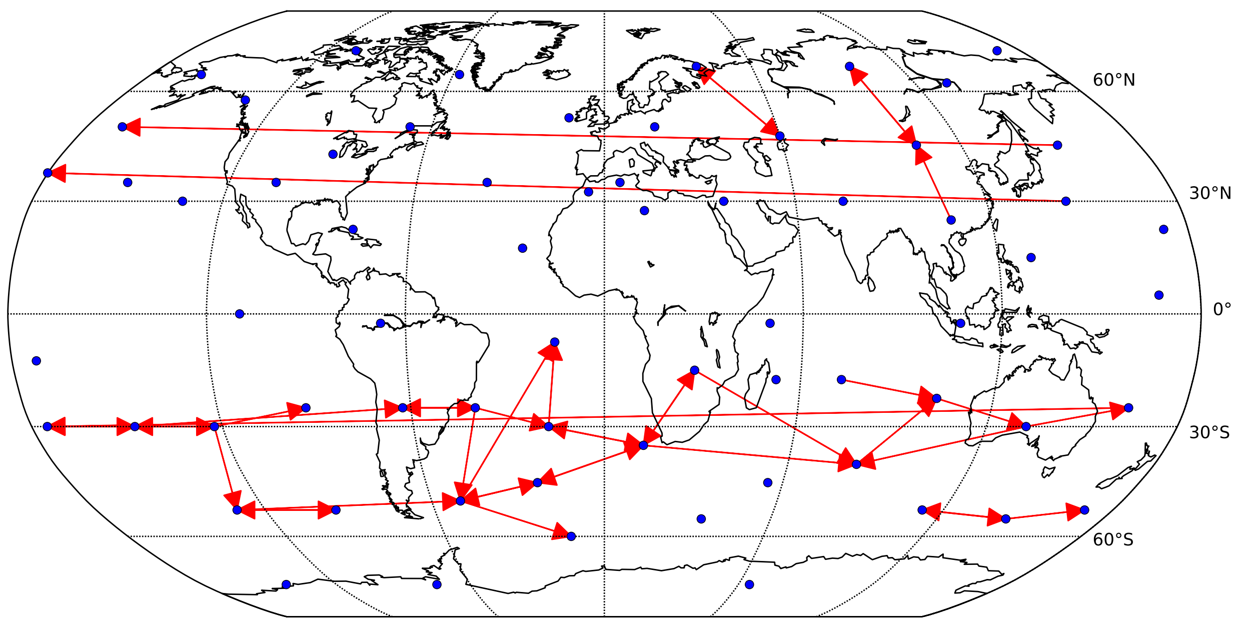
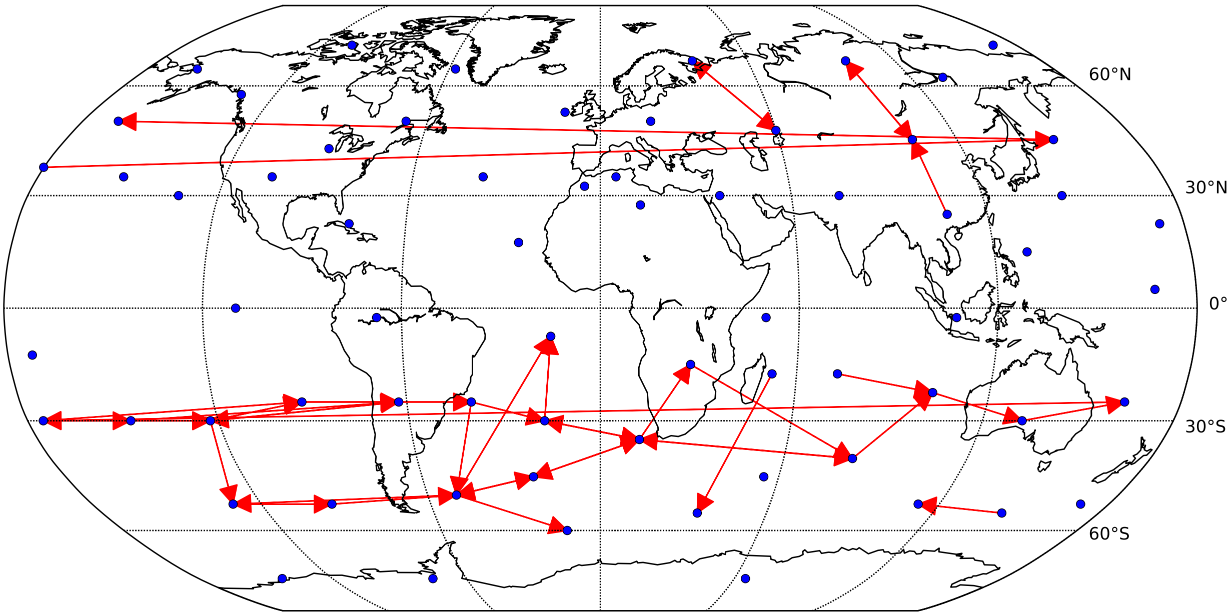


Figure 8. Example of detected causality network, detected by the equiquantal conditional mutual information method with $Q = 2$, for the decade 1948-1957. Density fixed to correspond to density of the network shown in Figure 7.



The series of examinations provided evidence that both nonlinear and linear methods may be used to construct directed climate networks in a reliable way under a range of settings, with the basic linear Granger causality outperforming the studied nonlinear methods.

For the sake of tractability, we have limited the investigation in several ways. On the side of nonlinear causality methods, we focused on the prominent family of methods based on estimation of conditional mutual information (transfer entropy). Two algorithms were used that represent key approaches of estimation of conditional mutual information and have been extensively used and proven efficient on real-world data. Alternative approaches also exist including the use of recurrence plots [27].

The particular pairwise version of linear Granger causality was chosen for its theoretical equivalence to the transfer entropy (under the assumption of linearity), as this provides a fair comparison.

However, the use of the strictly pairwise causality estimators suffers from inherent limitations. To give an example, a system consisting of three processes X, Y, Z, where Z drives both X and Y, but with different temporal lags, may erroneously show causal influence between X and Y even if these were not directly coupled. To deal with such situations, the concepts can be generalized to allow to take into account the variance explained by third variable(s).

Similarly, the assumption of a single possible lag (1 time step of 1 or 6 days respectively in our investigation) may in real context not be suitable, although at least the relative reliability of different methods may not be strongly affected by this within reasonable range of parameters.

In general, estimation of these generalized causality patterns from relatively short time series is technically challenging, particularly in the context of nonlinear, information-theory based causality measures, due to the exponentially increasing dimension of probability distributions to be estimated.

However, recent work has provided promising approaches to tackle this curse of dimensionality by decomposing TE into low-dimensional contributions [28]. For theoretical and numerical considerations on how a causal coupling strength can be defined in the multivariate context, see [29].

For completeness we mention that apart from the time-domain treatment of causality, the whole problem can also be reformulated in the spectral domain, leading to frequency-resolved causality indices such as partial directed coherence (PDC, [30]) or Directed Transfer Function (DTC, [31]).

The study also shows some key properties of the conditional mutual information estimates. Note that for instance for the 6-days networks (Figure 4), the EQQ method reliability increases again for $Q \geq 11$, however, the similarity to the linear estimate further decreases. A direct check of the causality networks shows that they tend to a trivial column-wise structure, with the intensities for each column highly correlated to the autoregressive coefficient of the given time series. This corresponds to a manifestation of a dominant autocorrelation-dependent bias in the EQQ estimator for too high Q values (note that $Q = 14$ corresponds to less than one time point per average in a 4D bin, an unsuitable sampling of the space for effective probability distribution function approximation).

Reconstruction of networks directly from gridded climatic field data is a challenging and perhaps not always the best approach, for reasons including efficient computation and visualization of the results. Instead, we apply here a decomposition of the data in order to get the most important components, i.e., by using a varimax-rotated principle component analysis. This provides a useful dimension-reduction of the studied problem. In particular, the gridded data does not reflect the real climate subsystems, which may be better approximated by the decomposition modes. However, as the decomposition is an implicit (weighted) coarse-graining, the detected difference between the linear and nonlinear methods may be different from that in the original time series data. In particular, the spatial averaging may increase the reliability of both approaches by suppressing noise, but also suppress any highly spatially localized heterogeneous patterns of both nonlinear and linear character. This might be reflected in the specific results of the paper, in particular the obtained quantitative reliability estimates.

In general, the differences in reliability may have important consequences for the detectability of causal links as well as of their changes. Of course, theoretically this is not necessary as in a particularly nonlinear system links may exist that would have negligible linear causality component, but their nonlinear causality fingerprint would be strong enough to be detected. Recent works have proposed approaches for explicit quantification of the nonlinear contribution of equal-time dependence [32] and applied them to neuroscientific [32] as well as climate data [10]. However, the generalization of the approach to higher dimension information-theoretic functionals is not straightforward and is subject of ongoing work. Thus, we give here at least an illustrative example of the practical result of trade-off between generality of transfer entropy and higher reliability of linear Granger causality: using the original 67 components data (divided into 6 decadal sections, see section 2), the basic statistical test at the 5 percent significant level (described in the 2 section) marked on average 1592 links as statistically significant (out of 4422 possible), while the EQQ method showed in general a lower number of significant links, dependent on the parameter value in a way similar to the reliability estimate, with maximum of 983 significant links for $Q = 2$ and minimum of 287 significant links for $Q = 13$. Note that given the significance level of the test, on average $221 \approx 4422 * 0.05$ significant links would be expected to appear by chance in a collection of completely unrelated processes.

5. Conclusions

Meaningful interpretation of climate networks and their observed temporal variability requires knowledge and minimization of the methodological limitations of the methods of their construction. In the presented work, we discussed the problem of reliability of network construction from time series of finite length, quantitatively assessing the reliability for a selection of standard bivariate causality methods. These included two major algorithms for estimating transfer entropy with a wide range of parameter choices, as well as the linear Granger causality analysis, which can be understood as linear approximation of transfer entropy. Overall, causality methods provided reproducible estimates of climate causality networks, with the linear approximations outperforming in reliability the studied nonlinear methods. Interestingly, optimizing the nonlinear methods with respect to reliability has led to improved similarity of the detected networks to those discovered by linear methods, in line with the hypothesis of near-linearity of the investigated climate reanalysis data, in particular the surface air temperature time series.

The latter hypothesis regarding the surface air temperature has been supported by the study in [10] which extended the older results of [33] who tested for possible nonlinearity in the dynamics of the station (Prague-Klementinum) SAT time series and found that the dependence between the SAT time series $x(t)$ and its lagged twin $x(t+\tau)$ was well-explained by a linear stochastic process. This result about a linear character of the temporal evolution of SAT time series, as well as the results of this study about causal relations between the principal components obtained from the reanalysis SAT time series cannot be understood as arguments for a linear character of atmospheric dynamics *per se*. Rather, these results characterize properties of measurement or reanalysis data at a particularly coarse level of resolution, when the data reflecting a spatially and temporally averaged mixture of dynamical processes on a wide range of spatial and temporal scales are considered. For instance, a closer look on the dynamics on specific temporal scales in temperature and other meteorological data has led to identification of oscillatory phenomena with nonlinear behavior, exhibiting phase synchronization [34–38]. Also the leading modes of atmospheric variability exhibit nonlinear behavior [39,40] and can influence important seasonal circulation phenomena in a nonlinear way [41].

Further work is needed to assess the usability and advantages of more sophisticated, recently proposed causality estimation methods. The current work also provides an important step towards reliable characterization of climate networks and detection of potential changes over time.

Acknowledgements

This study is supported by the Czech Science Foundation, Project No. P103/11/J068 and by the DFG grant No. KU34-1.

References

1. Newman, M.E.J. The structure and function of complex networks. *SIAM Review* **2003**, *45*, 167–256.
2. Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D.U. Complex networks: Structure and dynamics. *Physics Reports* **2006**, *424*, 175–308.

3. Tsonis, A.; Roebber, P. The architecture of the climate network. *Physica A* **2004**, *333*, 497–504.
4. Tsonis, A.A.; Swanson, K.L.; Roebber, P.J. What do networks have to do with climate? *Bulletin of the American Meteorological Society* **2006**, *87*, 585+.
5. Yamasaki, K.; Gozolchiani, A.; Havlin, S. Climate networks around the globe are significantly affected by El Nino. *Physical Review Letters* **2008**, *100*.
6. Malik, N.; Bookhagen, B.; Marwan, N.; Kurths, J. Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks. *Climate Dynamics* **2012**, *39*, 971–987.
7. Steinhaeuser, K.; Ganguly, A.R.; Chawla, N.V. Multivariate and multiscale dependence in the global climate system revealed through complex networks. *Climate Dynamics* **2012**, *39*, 889–895.
8. Donges, J.F.; Zou, Y.; Marwan, N.; Kurths, J. The backbone of the climate network. *EPL* **2009**, *87*, 48007.
9. Donges, J.F.; Zou, Y.; Marwan, N.; Kurths, J. Complex networks in climate dynamics. *European Physical Journal* **2009**, *174*, 157–179.
10. Hlinka, J.; Hartman, D.; Vejmelka, M.; Novotna, D.; Paluš, M. Non-linear dependence and teleconnections in climate data: sources, relevance, nonstationarity. *submitted*.
11. Granger, C.W. Investigating causal relations by econometric model and cross-spectral methods. *Econometrica* **1969**, *37*, 414–&.
12. Vejmelka, M.; Palus, M. Inferring the directionality of coupling with conditional mutual information. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)* **2008**, *77*.
13. Schreiber, T. Measuring information transfer. *Physical Review Letters* **2000**, *85*, 461–464.
14. Kistler, R.; Kalnay, E.; Collins, W.; Saha, S.; White, G.; Woollen, J.; Chelliah, M.; Ebisuzaki, W.; Kanamitsu, M.; Kousky, V.; van den Dool, H.; Jenne, R.; Fiorino, M. The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society* **2001**, *82*, 247–267.
15. 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.; Joseph, D. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* **1996**, *77*, 437–471.
16. Wiener, N., Modern Mathermatics for Engineers; McGraw-Hill, New York, 1956; chapter The theory of prediction, pp. 165 – 190.
17. Ding, M.; Chen, Y.; Bressler, S.L., Granger Causality: Basic Theory and Application to Neuroscience. In *Handbook of Time Series Analysis*; Wiley-VCH Verlag GmbH & Co. KGaA, 2006; pp. 437–460.
18. Geweke, J.F. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association* **1982**, *77*, 77.
19. Geweke, J.F. Measures of Conditional Linear Dependence and Feedback Between Time Series. *Journal of the American Statistical Association* **1984**, *79*, 907–915.
20. Barnett, L.; Barrett, A.B.; Seth, A.K. Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables. *Physical Review Letters* **2009**, *103*.

21. Palus, M.; Albrecht, V.; Dvorak, I. Information theoretic test for nonlinearity in time series. *Physics Letters A* **1993**, *175*, 203–209.
22. Kaiser, H. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **1958**, *23*, 187–200.
23. Schwarz, G. Estimating the dimension of a model. *The Annals of Statistics* **1978**, *5*, 461–464.
24. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal Of The Royal Statistical Society Series B-methodological* **1995**, *57*, 289–300.
25. Prichard, D.; Theiler, J. Generating surrogate data for time series with several simultaneously measured variables. *Physical Review Letters* **1994**, *73*, 951.
26. Palus, M. Detecting phase synchronization in noisy systems. *Physics Letters A* **1997**, *235*, 341–351.
27. Zou, Y.; Romano, M.C.; Thiel, M.; Marwan, N.; Kurths, J. Inferring indirect coupling by means of recurrences. *International Journal Of Bifurcation And Chaos* **2011**, *21*, 1099–1111.
28. Runge, J.; Heitzig, J.; Petoukhov, V.; Kurths, J. Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy. *Physical Review Letters* **2012**, *108*.
29. Runge, J.; Heitzig, J.; Marwan, N.; Kurths, J. Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy. *Physical Review E* **2012**, *86*.
30. Baccala, L.A.; Sameshima, K. Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics* **2001**, *84*, 463–474.
31. Kaminski, M.; Ding, M.; Truccolo, W.A.; Bressler, S.L. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol Cybern* **2001**, *85*, 145–157.
32. Hlinka, J.; Palus, M.; Vejmelka, M.; Mantini, D.; Corbetta, M. Functional connectivity in resting-state fMRI: Is linear correlation sufficient? *NeuroImage* **2011**, *54*, 2218–2225.
33. Palus, M.; Novotna, D. Testing for nonlinearity in weather records. *Physics Letters A* **1994**, *193*, 67–74.
34. Palus, M.; Novotna, D. Enhanced Monte Carlo Singular System Analysis and detection of period 7.8 years oscillatory modes in the monthly NAO index and temperature records. *Nonlinear Processes in Geophysics* **2004**, *11*, 721–729.
35. Palus, M.; Novotna, D. Quasi-biennial oscillations extracted from the monthly NAO index and temperature records are phase-synchronized. *Nonlinear Processes in Geophysics* **2006**, *13*, 287–296.
36. Palus, M.; Novotna, D. Phase-coherent oscillatory modes in solar and geomagnetic activity and climate variability. *Journal of Atmospheric and Solar-terrestrial Physics* **2009**, *71*, 923–930.
37. Palus, M.; Hartman, D.; Hlinka, J.; Vejmelka, M. Discerning connectivity from dynamics in climate networks. *Nonlinear Processes in Geophysics* **2011**, *18*, 751–763.
38. Feliks, Y.; Ghil, M.; Robertson, A.W. Oscillatory Climate Modes in the Eastern Mediterranean and Their Synchronization with the North Atlantic Oscillation. *Journal of Climate* **2010**, *23*, 4060–4079.

39. Boucharel, J.; Dewitte, B.; du Penhoat, Y.; Garel, B.; Yeh, S.W.; Kug, J.S. ENSO nonlinearity in a warming climate. *Climate Dynamics* **2011**, *37*, 2045–2065.
40. Osprey, S.M.; Ambaum, M.H.P. Evidence for the chaotic origin of Northern Annular Mode variability. *Geophysical Research Letters* **2011**, *38*.
41. Mokhov, I.I.; Smirnov, D.A.; Nakonechny, P.I.; Kozlenko, S.S.; Seleznev, E.P.; Kurths, J. Alternating mutual influence of El-Nino/Southern Oscillation and Indian monsoon. *Geophysical Research Letters* **2011**, *38*.