Smooth Information Flow in Temperature Climate Network Reflects Mass Transport

Jaroslav Hlinka,[1, a)] Nikola Jajcay,[1, 2, b)] David Hartman,[1, c)] and Milan Paluš[1, d)]
[1)] *Department of Nonlinear Dynamics and Complex Systems, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, Czech Republic*
[2)] *Department of Atmospheric Physics, Faculty of Mathematics and Physics, Charles University in Prague, Czech Republic*

(Dated: 6 January 2017)

A directed climate network is constructed by Granger causality analysis of air temperature time series from a regular grid covering the whole Earth. Using winner-takes-all network thresholding approach, a structure of a smooth information flow is revealed, hidden to previous studies. The relevance of this observation is confirmed by comparison with the air mass transfer defined by the wind field. Their close relation illustrates that although the information transferred due to the causal influence is not a physical quantity, the information transfer is tied to the transfer of mass and energy.

PACS numbers: 89.75.Fb, 02.50.Sk, 05.45.Tp
Keywords: directed network; causal network; Granger causality; climate network; information flow; temperature network

Recently the complex network approach to analysing dynamical systems reached also the climate science and became one of the tools for uncovering dependence structures and teleconnections in the atmospheric data. Usually, for the sake of clarity and interpretability, symmetric statistical measures such as correlation or mutual information are taken into account. However, the drawback is that these measures lack the notion of directionality. In this paper we investigate the possibility of inferring the causal climate network with the explicit direction of causal influence, taking the conservative approach of linear Granger causality applied to gridded temperature data. Following this avenue, in conjunction with a novel winner-takes-all thresholding scheme, yields an easy-to-interpret causal network with smooth information flow structure. To assess the significance of this observation, we designed random graph models for climate networks and quantitatively compared the temperature causal network with prevailing wind direction.

I. INTRODUCTION

The network approach to characterization of complex systems is ever growing in applications, covering diverse fields ranging from social networks via neuroscience to technological networks[1,2]. Recently, the complexity of the climate system and the high dimensionality of the relevant data has lead to increased utilization of this approach in climate science[3–6].

Probably the most commonly taken avenue to graph-theoretical characterization of complex dynamical systems is the construction of an interdependence graph. In the case of global climate, local temperature or pressure time series are commonly utilized to represent the evolution of the local states. It is widely accepted that their statistical dependence structure reveals important phenomena, among all the existence, spatial patterns and dynamics of the major modes of climate variability[7] and distant atmospheric teleconnections[8,9].

[a)] Electronic mail: hlinka@cs.cas.cz
[b)] Electronic mail: jajcay@cs.cas.cz
[c)] Electronic mail: hartman@cs.cas.cz
[d)] Electronic mail: mp@cs.cas.cz

The network is constructed using a pairwise estimate of statistical dependence between regional temperature time series, and potentially thresholding the resulting matrix to obtain a binary graph representation for further analysis. Detailed approaches differ between studies; for discussion of the effect of using (non)linear dependence measures see Hlinka et al.[10].

The majority of the climate network studies have been limited to the use of symmetric dependence measures such as correlation or mutual information, which lack an explicit notion of direction and thus fall short of capturing the dynamics, as well as directionality of causal interaction or information flow. Some authors[11,12], however, have made inferences of the latter using time-lagged cross-correlations. Most recently, a modified partial correlation method was applied for estimation of direct links and teleconnection paths[13].

A principled approach to the construction of directed networks is based on the ideas of Granger causality[14]. Here, a source process is considered causal, if its inclusion in a predictive model decreases the uncertainty about the target process. Tirabassi et al.[15] applied bivariate Granger causality in order to construct directed bipartite networks describing the air–sea interaction in the region of the South Atlantic Convergence Zone. In a different context, Deza et al.[16] used bivariate conditional mutual information (a nonlinear approach related to Granger causality) to construct a directionality index and map the direction of climate interactions in daily and monthly surface air temperature reanalysis data. The authors of Deza et al.[16] restricted the analysis to interaction maps for only a few selected nodes, likely due to computational demands of their method.

Recently also directed climate networks based on multivariate causality approaches have been introduced[17–19]. In these studies, the notion of conditional independence is used in order to uncover *direct* links, i.e., to remove directed links resulting probably due to indirect influences. Also, estimation of causal influence across a range of temporal lags is included in these more general approaches. However, due to the curse of dimensionality (albeit alleviated by various algorithms for preselection of variables included in the condition)[20], the practical application of the multivariate approaches is seriously limited with respect to the tractable size of the network. Therefore these approaches require an initial convenient reduction of dimensionality[19] and/or a selection of a region of interest[18].

In this study we document how one can tackle the challenge of large causal network discovery by combining a simple yet efficient bivariate linear causality approach with appropriately sampled data and proper random graph models. Although theoretically only approximate and potentially blind to indirect links, bivariate linear causality is supremely numerically robust, computationally efficient and reliable even in the case of large networks – as recently shown in a direct comparison with nonlinear and multivariate approaches[21].

By applying this approach, we obtain a directed graph of influences within the global temperature field. We reveal a global smooth/laminar information flow structure in the climate temperature network, that is not directly observable in correlation networks, and could not be properly observed in previous causality studies due to sampling limitations of less conservative causality estimates.

Thus this classical method proves suitable exactly for the original, spatiotemporally densely sampled data, for which more sophisticated methods become computationally infeasible or unreliable. The global information flow structure is particularly clearly captured when a nontraditional winner-takes-all thresholding is applied in the network construction; this remedies both the spatial heterogeneity of the causal strengths as well as the iconic issue with indirect links in bivariate causality networks.

To provide quantitative evidence concerning the observed climate network properties, we design random graph models for climate networks that account for some of the spatiotemporal constraints. Finally, to gain insight into the meaning of relatively abstract climate network structures, we advocate quantitative comparison with other measurable physical quantities. As a pioneering example, we prove a plausible relation between the temperature causality graph and the wind speed and direction field.

II. DATA AND METHODS

A. Temperature data

For estimating the Granger causality flow in temperature, we employed spatio-temporal fields from the NCEP/NCAR reanalysis[22]. We used daily air temperature at 1000hPa level, spanning the period from 1/1/1948 up to 12/31/2012; in particular we worked with air temperature anomalies obtained by subtracting the average yearly cycle. The data (originally sampled at 2.5° × 2.5° angularly regular Gaussian grid) were remapped to an equidistant geodesic grid of 2562 grid points[23] in order to suppress the effect of unequal distances between grid points of pairs located in different latitudes. The poles were omitted.

B. Construction of the directed graph

The estimation of causal interactions in the temperature data was carried out using Granger causality analysis[14]. This defines a stochastic process X_t as causing Y_t if it is possible to better predict Y_t using past of X_t (and potentially other all available information in processes Z_t) than if information from past of X_t had not been used. The causal influence from X to Y is then quantified based on the decrease in the model residual variance when we include the past of X in the model of Y; in particular by the logarithm of the variance ratio.

In the presented analysis we have utilized a basic (but numerically robust[21]) approach to Granger causality analysis, working with bivariate linear autoregressive models of order 1. The time unit was chosen to be 1 day.

Mathematically, in order to estimate the Granger causality between two grid points i and j, we estimate the residuals of the AR(1) model from univariate models $X_i(t)$ and $X_j(t)$ and then from a bivariate model $X_{i,j}(t)$ as

$$X(t) = w + A_1 X(t-1) + \epsilon(t), \tag{1}$$

where $X(t) \in \mathbb{R}^m$ is a state vector (in our case the vector of temperature anomalies), $w \in \mathbb{R}^m$ is an intercept vector, $\epsilon(t) \in \mathbb{R}^m$ is the vector of residuals and finally $A_1 \in \mathbb{R}^{m\times m}$ is a coefficient matrix of the AR(1) model, while $m = 1$ for univariate model and $m = 2$ for bivariate model.

In the next step we compute the variances of the four residual vectors from the AR(1) model, namely the univariate residual vectors for grid points i and j separately (ϵ_i and ϵ_j), and then the residual vectors from the bivariate AR(1) model for both of them ($\epsilon_{i,ij}$ and $\epsilon_{j,ij}$). Finally, we estimate the Granger causality as

$$\mathrm{GC_{i\to j}} = \log\left(\frac{\mathrm{var}\,(\epsilon_j)}{\mathrm{var}\,(\epsilon_{j,ij})}\right) \tag{2}$$

$$\mathrm{GC_{j\to i}} = \log\left(\frac{\mathrm{var}\,(\epsilon_i)}{\mathrm{var}\,(\epsilon_{i,ij})}\right). \tag{3}$$

III. RESULTS

A. Directional graph

For the first overview of the results, we plot the strongest 2562 causalities in Fig. 1 (self-causalities between the same grid points - the diagonal elements of the causality matrix - are set to zero). The strongest causalities are located mainly in the northern hemisphere extra-tropical region (roughly above 30°N) above the continents. We believe that this effect has relation with the fact that above the continental areas the variance in temperature time series (as well as the variance of residuals of AR(1) model of SATA) is higher than

over the oceans. The temperature variability of the oceans has slower dynamics and longer persistence, so that the short range changes investigated here (lag of 1 day) can be relatively small in comparison with long range changes. Furthermore, the temperature field above continents is less homogeneous and have less memory, hence more prone to show significant information transfer. In general, the causal influence has an eastward direction with little exceptions which seem to point equator-ward. This causal influence structure resembles a laminar flow of liquid in the sense that there are no sudden changes of direction and no substantial sinks or sources.

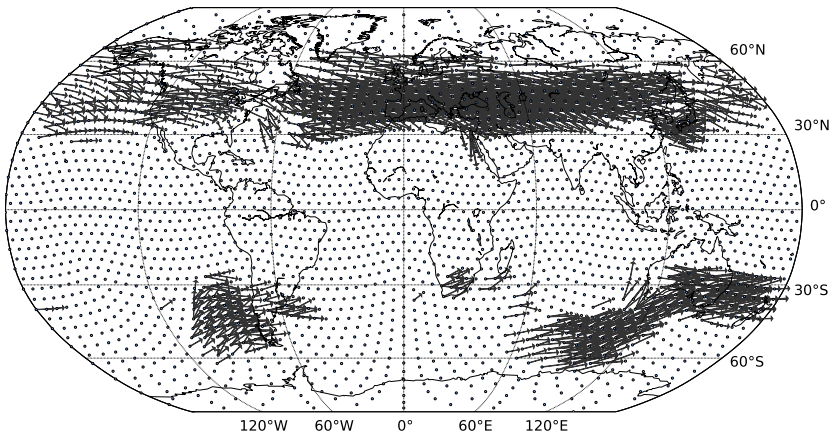

FIG. 1. (color online). Granger causality estimate for temperature field (at 1000hPa height level). Showing 2562 overall strongest causalities for the period 1/1/1948 - 12/31/2012.

Fig. 1 shows a prominent non-homogeneity in the density of the directed graph over the globe, revealed by the 2562 strongest links. While thresholding a graph based on a fixed threshold is a common approach, in cases such as this one it complicates both visual and quantitative inspection of the graph representation of the studied complex system. Note that similar results can be expected from alternative thresholding approaches such as based on statistical significance (due to more or less monotonic relation between significance and coupling strength within a network, up to issues such as those related to violation of data sample independence). Moreover, the heuristic choice of significance threshold is further complicated by problems related to multiple testing corrections in large networks and related computational difficulties.

For this reason we introduce an alternative representative graph construction (winner-takes-all): for each grid point i we add a single outgoing edge (i, j) corresponding to the node j that receives the strongest causal influence from the node i. Therefore, we force the out-degree of each grid point to be 1. The complete directed network from the temperature data is shown in Fig. 2. In agreement with the first approach (Fig. 1), the global network shows strongest causal influences in the extra-tropical northern hemisphere. However, even the substantially weaker interactions found in the extra-tropical southern hemisphere can now be inspected. The general direction of causal influence on both hemispheres is eastward, with a quite ordered, laminar flow-like pattern. The causal influences in the equatorial belt have no general direction, are weaker when compared to the extra-tropical influences, and many of the nodes even exhibit convergence (in-degree of these nodes is larger than 1). In other words, the causal 'pathways' in the equatorial area are less ordered.

B. Smooth flow properties

The qualitative visual similarity of the temperature causality graph to a laminar flow is an interesting observation in the climate network analysis that is not apparent when the graph is constructed based on symmetric measures of statistical dependence such as Pearson

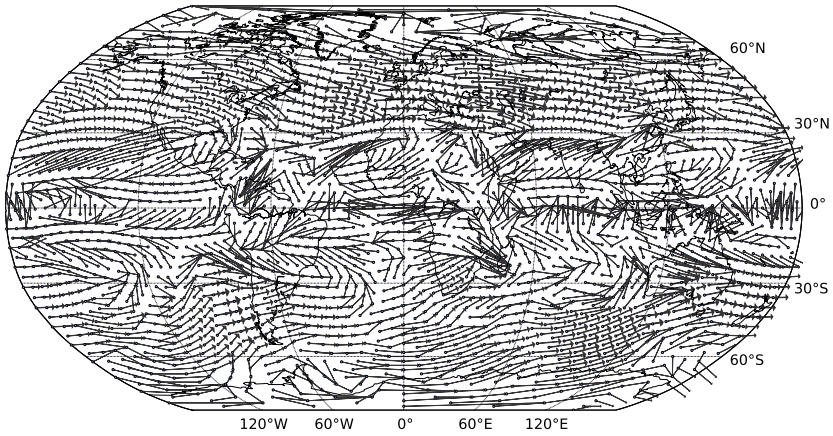

FIG. 2. (color online). Granger causality estimate for temperature field (at 1000hPa height level). Showing one strongest causality for each grid point for the period 1/1/1948 - 12/31/2012.

correlation coefficient or mutual information. However, to go behind a qualitative description of the observed graph structure, its properties need to be quantified and compared to an expected distribution under a reasonable null model.

To this end, we propose two approaches. The first one considers the flow-likeness itself – i.e. the sparse occurrence of sinks; and is thus framed purely graph-theoretically. It involves recording the distribution of the in-degree of the graph nodes, to quantify the flow-like/non-convergent character of the graph. Intuitively, a flow-like graph corresponding to a system with just a few sinks/sources would have an in-degree distribution with a sharp peak at in-degree 1 (in an extreme case, exactly one incoming link for each node).

The second measure focuses on the laminarity or smoothness of the causality direction field. Indeed, the paths through the observed causality graph appeared not only non-branching, but also surprisingly straight and locally parallel, with only minimal differences in direction between neighbouring nodes. To capture this, we study the distribution of the local average direction, e. g., the mean vector (i, j) between a grid point i and the six closest sourcing node j on a rectangular projection.

1. Non-convergence

As one can see from the histogram of the in-degree in the temperature data driven graph in Fig. 3 top in red, the in-degree 1 is roughly three times more probable than the in-degree 2, and the count of nodes with higher in-degree is negligible. This quantitatively corresponds to the visually observable flow-like structure of the graph.

To obtain a proper quantitative/statistical evidence, we generate a set of surrogate directed graphs with out-degree one, but randomly selected target nodes. The comparison with the random graph distribution (Fig. 3 in blue) shows that random graphs exhibit generally higher variance of the in-degree. We can now conclude that convergence is less frequent in the temperature graph than in the random graph. Note that this property is purely graph-theoretical; random graphs obtained by permuting the geospatial locations of the nodes have of course the same distribution of in-degrees as the original graph (Fig. 3 in yellow).

2. Smoothness

However, the information flow uncovered in Fig. 2 is not only flow-like in a graph-theoretical sense, but as a spatially embedded structure it is highly laminar in the sense of

smoothness of the direction field. In fact, the difference between the temperature causality graph and random graphs is even stronger when looking at the histogram of average angle differences (Fig. 3). The temperature data driven graph contains smaller mean local direction differences (mostly $\leq 10°$), meaning the causality flow is approximately smooth/laminar.

This tendency is evident when comparing the distribution of mean local angle differences of a random directed graph or a graph, where the spatial locations of the nodes were randomly reassigned (Fig. 3 in blue and yellow). Here, the most common mean local direction difference is between 50° and 60° for the used grid and orthogonal projection.

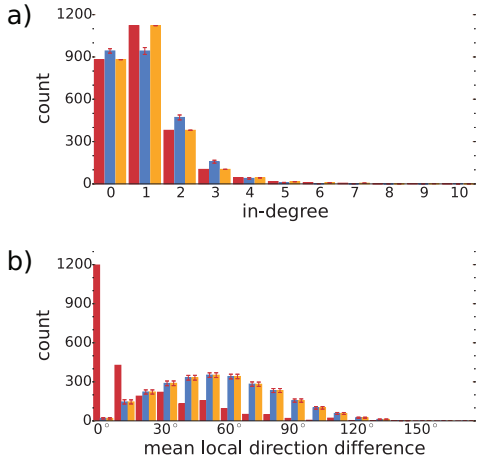

FIG. 3. (color online). Quantitative assessment of the flow-like structure of temperature causality graph: (a) histogram of in-degrees and (b) histogram of average local direction angle differences. Red: temperature causality graph, yellow: the same graph with randomly permuted spatial embedding, blue: random directed graph with all nodes with out-degree 1. For random graphs, the bars represent mean and the error-bars represent standard deviation of across 1000 realizations.

C. Wind direction

At this point we may consider the observation that the graph derived from causal influences in the temperature dynamics exhibits a smooth flow-like structure as quantitatively substantiated. However, the question of the origin of this structure emerges. Based on inspection of the specific pattern of the flow, we hypothesized that the causal influences in the temperature field may be caused by air mass transfer. If this was the case, it should at least partially comply with the prevalent wind direction. To test this hypothesis we determined the prevalent wind direction for each grid point and compared it to the causal influence direction.

For this purpose, the zonal (u-wind) and meridional (v-wind) wind fields were employed from the same data source - the NCEP/NCAR reanalysis[22]. The field parameters (pressure level, spatial and temporal sampling and span) were the same as for the temperature field.

The updated graph showing the temperature causality and relating it to wind direction is shown in Fig. 4. In total, 36% of the causal influences are in agreement with the wind direction (angle difference $\leq 30°$). Note that this is substantially higher than the 16.7% that would be in agreement for a random directed graph. The agreement with the wind direction is apparent especially in the extra-tropics, particularly on the northern hemisphere where the strongest causalities are present.

Not only is the causality field in line with the wind field in terms of direction, there is also a plausible relation in terms of the causal link length. Although when estimating the causal influence between each pair of grid points, no restriction on length of such influence was posed, only about 1% of the causal influence links (29 out of 2562) were longer than

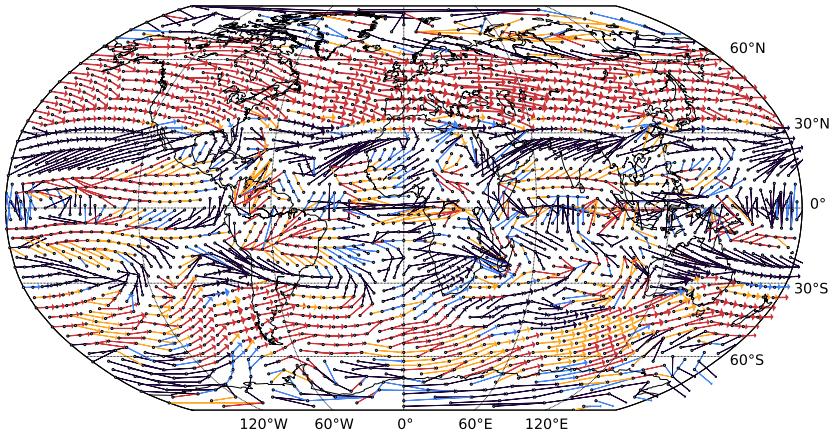

FIG. 4. (color online). Agreement between the direction of the estimated causal interaction (network as in Fig. 2) and the wind direction. Arrow color: red - difference $\leq 30°$, yellow - difference $\leq 60°$, blue - difference $\leq 90°$ and black - difference $\geq 90°$.

2400km, with vast majority of them pointing to the next grid point. This typical spatial scale of estimated causal influence is consistent with the spatiotemporal scale of the winds, as within 1 day the air mass is realistically able to travel the average distance between grid points, which is about 500km (corresponding to an average speed of 21kmh^{-1}).

While the 36% direction agreement is substantially higher than an expectable $\sim 16.7\%$ random agreement, it is quite far from a perfect alignment of the wind and causality fields. This suggests that other factors apart from air mass transport might play strong role in the surface temperature dynamics; or otherwise that the estimation procedure is imprecise or the data is too noisy. For instance, the causal influences in the equatorial belt ($< \pm 30°$N/S) are in general not in agreement with the wind direction, while the estimated influences here are small in comparison with extra-tropical areas.

To gather more insight into this problem, we further analyzed data from higher levels of atmosphere, where the spatiotemporal dynamics are in general simpler, unaffected by the surface orography. Interestingly, the agreement between the direction of causal influences and local prevalent wind direction grows with height. When the directed graph is estimated using the 500hPa temperature field and then compared with 500hPa wind field (i.e. about 5,500 meters (18,000 feet) above the see level), full 75.9% of the causal influences are in excellent agreement with the wind direction (angle difference $\leq 30°$), see Fig. 5.

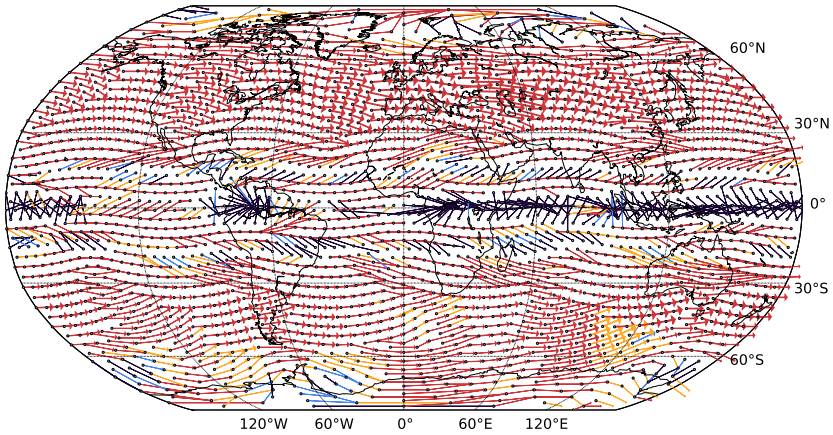

FIG. 5. (color online). Granger causality estimate for temperature field (at 500hPa height level). Visualization as in Fig. 4.

IV. CONCLUSION

We have proposed an approach to the construction of climate networks that provides a representation of the system that captures the directed interactions within the field, globally, robustly and with a high spatial resolution. To solve the problem of thresholding a graph with a highly heterogeneous degree, we introduced a winner-takes-all construction. For the case of air temperature field, our methodology clearly uncovers a smooth flow structure; evident both qualitatively and in quantitative comparison with appropriate random graphs. The climatological relevance is shown by the close relation to the air-mass flow. Our study is probably the first quantitative comparison of an abstract causality measure (information flow) with a measured physical quantity - the wind vector field. This new observation, not clearly represented in the correlation (or mutual information) networks, provides a major improvement in the interpretability of climate networks, opening new avenues of research in this field.

The presented result also adds to the ongoing discussions about the (non)physical nature of information. Paluš et al.[24] proposed information-theoretic formulation of the Granger causality for nonlinear processes using conditional mutual information. The latter functional become known as the transfer entropy due to Schreiber[25]. The equivalence of the two is shown by Palus and Vejmelka[26]. Barnett et al.[27] have shown analytically that the transfer entropy is equivalent to Granger causality for Gaussian processes. These equivalence enable us to interpret the Granger causality relation as an information flow between connected nodes of climate network. Prokopenko and Lizier[28] related the transfer entropy to fundamental physical quantities, namely to the thermodynamic entropy production in small-scale physical systems. In macroscopic systems such as the Earth atmosphere studied here, however, the interpretation as given by Ebeling and Feistel[29] is more appropriate: the information measuring dependence is an abstract, nonphysical quantity, however, information transfer is tied to the transfer of mass and energy. Detection of information transfer from instrumental and model climate data helps to shed new light on known climate processes as well as to uncover new phenomena such as the recently observed information transfer across time scales of atmospheric dynamics[30] which has considerable impact on interannual temperature variability[31].

ACKNOWLEDGMENTS

This work was supported by the Czech Science Foundation project No. P103/11/J068 and by the Ministry of Education, Youth and Sports of the Czech Republic within the Program KONTAKT II, project LH14001.

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).
[2] M. E. Newman, SIAM Rev. **45**, 167 (2003).
[3] A. A. Tsonis and P. J. Roebber, Physica A **333**, 497 (2004).
[4] A. A. Tsonis, K. L. Swanson, and P. J. Roebber, B. Am. Meteorol. Soc. **87**, 585 (2006).
[5] K. Yamasaki, A. Gozolchiani, and S. Havlin, Phys. Rev. Lett. **100**, 228501 (2008).
[6] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, Eur. Phys. J.–Spec. Top. **174**, 157 (2009).
[7] M. Vejmelka, L. Pokorná, J. Hlinka, D. Hartman, N. Jajcay, and M. Paluš, Clim. Dynam. **44**, 2663 (2015).
[8] A. G. Barnston and R. E. Livezey, Mon. Weather Rev. **115**, 1083 (1987).
[9] S. B. Feldstein, J. Climate **13**, 4430 (2000).
[10] J. Hlinka, D. Hartman, M. Vejmelka, D. Novotná, and M. Paluš, Clim. Dynam. **42**, 1873 (2014).
[11] A. Gozolchiani, S. Havlin, and K. Yamasaki, Phys. Rev. Lett. **107**, 148501 (2011).
[12] J. Ludescher, A. Gozolchiani, M. I. Bogachev, A. Bunde, S. Havlin, and H. J. Schellnhuber, P. Natl. Acad. Sci. USA **110**, 11742 (2013).
[13] D. Zhou, A. Gozolchiani, Y. Ashkenazy, and S. Havlin, Phys. Rev. Lett. **115**, 268501 (2015).
[14] C. W. Granger, Econometrica , 424 (1969).
[15] G. Tirabassi, C. Masoller, and M. Barreiro, Int. J. Climatol. **35**, 3440 (2015).
[16] J. I. Deza, M. Barreiro, and C. Masoller, Chaos **25**, 033105 (2015).
[17] I. Ebert-Uphoff and Y. Deng, J. Climate **25**, 5648 (2012).
[18] I. Ebert-Uphoff and Y. Deng, Geophys. Res. Lett. **39** (2012).

[19]J. Runge, V. Petoukhov, J. F. Donges, J. Hlinka, N. Jajcay, M. Vejmelka, D. Hartman, N. Marwan, M. Paluš, and J. Kurths, Nature Comm. **6** (2015).
[20]J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, Phys. Rev. Lett. **108**, 258701 (2012).
[21]J. Hlinka, D. Hartman, M. Vejmelka, J. Runge, N. Marwan, J. Kurths, and M. Paluš, Entropy **15**, 2023 (2013).
[22]E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, *et al.*, B. Am. Meteorol. Soc. **77**, 437 (1996).
[23]R. Heikes and D. A. Randall, Mon. Weather Rev. **123**, 1862 (1995).
[24]M. Paluš, V. Komárek, Z. Hrnčíř, and K. Štěrbová, Phys. Rev. E **63**, 046211 (2001).
[25]T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[26]M. Paluš and M. Vejmelka, Phys. Rev. E **75**, 056211 (2007).
[27]L. Barnett, A. B. Barrett, and A. K. Seth, Phys. Rev. Lett. **103**, 238701 (2009).
[28]M. Prokopenko and J. T. Lizier, Scientific Rep. **4** (2014).
[29]R. Feistel and W. Ebeling, "Self-organization of information and symbols," in *Physics of Self-Organization and Evolution* (2011) pp. 363–411.
[30]M. Paluš, Phys. Rev. Lett. **112**, 078702 (2014).
[31]N. Jajcay, J. Hlinka, S. Kravtsov, A. A. Tsonis, and M. Paluš, Geophys. Res. Lett. **43**, 902 (2016).