Clim Dyn (2014) 42:1665–1670
DOI 10.1007/s00382-013-1761-5

A climate model intercomparison at the dynamics level

Karsten Steinhaeuser · Anastasios A. Tsonis

Received: 17 November 2012 / Accepted: 30 March 2013 / Published online: 7 April 2013

Abstract Until now, climate model intercomparison has focused primarily on annual and global averages of various quantities or on specific components, not on how well the general dynamics in the models compare to each other. In order to address how well models agree when it comes to the dynamics they generate, we have adopted a new approach based on climate networks. We have considered 28 pre-industrial control runs as well as 70 20th-century forced runs from 23 climate models and have constructed networks for the 500 hPa, surface air temperature (SAT), sea level pressure (SLP), and precipitation fields for each run. We then employed a widely used algorithm to derive the community structure in these networks. Communities separate “nodes” in the network sharing similar dynamics. It has been shown that these communities, or sub-systems, in the climate system are associated with major climate modes and physics of the atmosphere (Tsonis AA, Swanson KL, Wang G, J Clim 21: 2990–3001 in 2008; Tsonis AA, Wang G, Swanson KL, Rodrigues F, da Fontura Costa L, Clim Dyn, 37: 933–940 in 2011; Steinhaeuser K, Ganguly AR, Chawla NV, Clim Dyn 39: 889–895 in 2012). Once the community structure for all runs is derived, we use a pattern matching statistic to obtain a measure of how well any two models agree with each other. We find that, with the possible exception of the 500 hPa field, consistency for the SAT, SLP, and precipitation fields is questionable. More importantly, none of the models comes close to the community structure of the actual observations (reality). This is a significant finding especially for the temperature and precipitation fields, as these are the fields widely used to produce future projections in time and in space.

Keywords Climate networks · Large-scale dynamics · Climate variability · Model intercomparison · Spatial pattern analysis

1 Introduction

Today there are more than two dozen different climate models which are used to make climate simulations and future climate projections. While they all share the basic formulation based on the Navier–Stokes equations, they differ in several aspects such as heat transport schemes, aerosol modeling, cloud parameterization, representation of terrestrial processes, ice sheet dynamics, oceanic dynamics, and other processes. These models are used as tools to understand climate variability (control runs) and to simulate how climate change will affect the planet (forced runs) not only at the annual/global average level but over specific areas of the globe.

In these models the evaluation of large-scale variability such as the North Atlantic Oscillation (NAO), the Pacific Decadal Oscillation (PDO), the El Niño/Southern Oscillation (ENSO), and the Pacific/North American (PNA) pattern is done at the component level. To evaluate how well the models reproduce ENSO, for example, the average

Electronic supplementary material The online version of this article (doi:10.1007/s00382-013-1761-5) contains supplementary material, which is available to authorized users.

K. Steinhaeuser (✉)
Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA
e-mail: ksteinha@umn.edu

A. A. Tsonis
Atmospheric Sciences Group, Department of Mathematical Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA

temperature of the NINO3 area (5°S to 5°N, 90°W to 150°W) is computed for all models and the corresponding power spectra are compared to that of the actual observations (reality) (AchutaRao and Sperber 2006; Randall 2007). This component-level evaluation gives an idea of how well a certain mode or pattern is simulated by the models. It does not, however, give an indication of how well the models generate the interplay of a set of modes.

The above-mentioned oscillations as well as other modes are major atmospheric and oceanic signals in the temperature and pressure (sea and upper levels) fields. They are coupled, they often synchronize, and their collective behavior defines the large scale variability of climate at interannual and decadal time scales (Tsonis et al. 2007; Swanson and Tsonis 2009; Wyatt et al. 2012). Thus, if a model adequately simulates ENSO but not PDO, to which is coupled, then the model does not adequately simulate their interplay and thus the dynamics. We will address the issue of comparing climate models at the "dynamics" level with a new approach involving climate networks.

2 Materials and methods

A network is defined by a set of nodes and their links. In the last 15 years, networks have found applications in many areas of science and recently have been applied to climate data organized as networks with impressive results (Tsonis et al. 2006, 2007, 2008, 2011; Gozolchiani et al. 2008; Tsonis and Swanson 2008; Yamasaki et al. 2008; Swanson and Tsonis 2009; Wang et al. 2009; Steinhaeuser et al. 2012; Wyatt et al. 2012). The topology of the network can reveal important and novel features of the system it represents (Strogatz 2001; Albert and Barabasi 2002; da Costa et al. 2007). One such feature is communities (Newman and Girvan 2004).

Communities represent groups of densely connected nodes with only a few connections between groups. It has been conjectured that each community represents a low-order subsystem and their collective behavior determines the dynamics of the complete system (Arenas et al. 2006). Thus, identification of these communities can offer useful insights about dynamics. Indeed it has been shown (Tsonis et al. 2006, 2008, 2011; Steinhaeuser et al. 2012) that climate networks are characterized by supernodes and a small number of communities, which relate to major teleconnection patterns/climate modes such as the NAO, ENSO, PNA, and PDO.

We start by considering 500 hPa, SAT, SLP, and precipitation fields generated in 28 pre-industrial control runs from 23 climate models (see supplementary material Table S1 for a list of model runs availability). All fields used here are arranged on a grid with a resolution of 5° latitude × 5° longitude. This results in 72 points in the east–west direction and 35 points in the north–south direction (leaving out the North and South poles, see supplementary material for details) for a total of $n = 2{,}520$ points. These 2,520 points are assumed to be the nodes of the network.

For each grid point monthly values for the 50-year period from 1950 to 1999 were considered. This period is the common period between all runs and observations. From the monthly values we produced anomaly values (actual value minus the climatological average for each month) and removed any trend (see supplementary material for details). Thus, for each grid point we have a time series of 600 anomaly values.

In order to define the links between the nodes, the correlation coefficient at lag zero (r) between the time series of all possible pairs of nodes $[n(n-1)/2 = 3{,}173{,}940$ pairs] is estimated. If between a pair $|r| \geq 0.5$ then this pair is connected (see more on this issue in supplementary material). Once a network is constructed we find its communities. We have employed a widely used algorithm to achieve this (Clauset et al. 2004; see supplementary material for description). Then we calculated a similarity measure, namely the Adjusted Rand Index (ARI; see supplementary material for definition) between the community structures of a run and the community structure of every other run and produced the similarity matrix shown in Fig. 1. The first row gives the ARI between run 1 and runs 1–28 and the last row the ARI between run 28 and runs 1–28 (see supplementary material Table S2 for correspondence between numbers and model runs). The ARI between a run with itself is equal to one (red diagonal).

3 Results

A simple visual inspection of Fig. 1 indicates that, with the exception of the 500 hPa field (top left panel) where there is good agreement between all models (owing to the strong division between barotropic tropics and baroclinic extratropics), the models are in significant disagreement when it comes to their SLP (top right), SAT (bottom left), and precipitation (bottom right) community structure. Table 1 (second column) gives the average ARI for the four fields. While the average ARI for the 500 hPa field is high, for the remaining fields it drops by a factor of two. This reflects the fact that the complexity of community structure (as indicated by the number of communities) increases from 500 hPa to SLP to SAT and to precipitation, a fact that has been observed in previous studies (Tsonis et al. 2008; Steinhaeuser et al. 2012). Another interesting observation is that the models may agree on one field but not on another. In other words, a set of models may delineate the

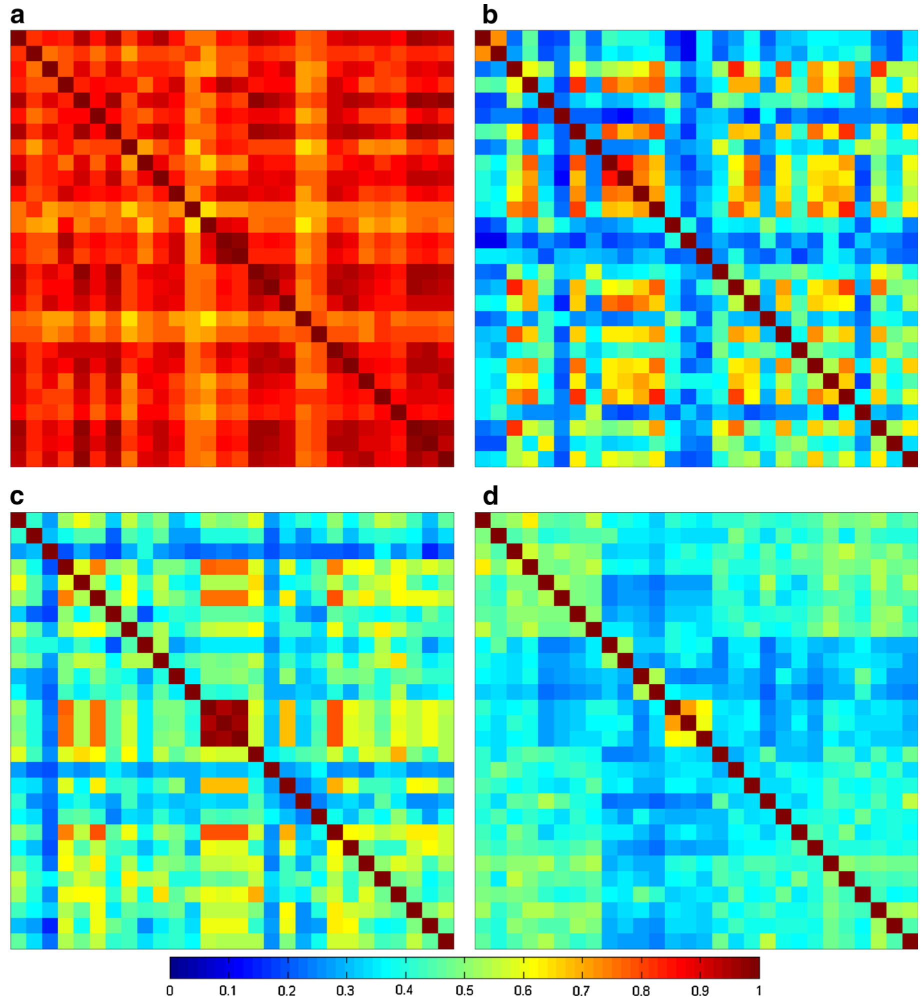

Fig. 1 We considered 28 model runs from 23 different climate models. For each run we considered four fields: **a** the 500 hPa field, **b** the Sea Level Pressure (SLP) field, **c** the surface air temperature (SAT), and **d** the precipitation field. For each run and each field we constructed the network and delineated its communities. We then estimated the Adjusted Rand Index (ARI) between a model run and all other available model runs The *top left panel* corresponds to the 500 hPa field, the *top right* to the SLP field, the *bottom left* to the SAT field, and the *bottom right* to the precipitation field. The *top row* is the comparison of run 1 with all other runs and the *bottom row* is the comparison of run 28 with all other runs. (Table S2 in the supplementary material provides the correspondence between numbers and model runs). The ARI between a run with itself is equal to one (*red diagonal*). See text for discussion of the results

Table 1 Average Adjusted Rand Index (ARI) for the pre-industrial control runs and for the forced 20-century runs; standard deviation and range are also shown for comparison

	Pre-industrial control runs	20-century forced runs	NCEP
500 hPa	0.84 $\delta = 0.072$	0.84 $\delta = 0.060$	0.86 ($\delta = 0.054$)
	(0.65,0.98)	(0.67,0.99)	(0.95, ECHAM5 run 1)
PSL	0.43 $\delta = 0.18$	0.41 $\delta = 0.14$	0.43 ($\delta = 0.17$)
	(0.13,0.86)	(0.13,0.81)	(0.68, GISS E-R run 2)
SAT	0.43 $\delta = 0.13$	0.43 $\delta = 0.12$	0.47 ($\delta = 0.11$)
	(0.18,0.96)	(0.17,0.82)	(0.65, MIROC HiRes run 1)
Precipitation	0.40 $\delta = 0.083$	0.39 $\delta = 0.083$	0.33 ($\delta = 0.068$)
	(0.20,0.64)	(0.22,0.67)	(0.46 MRI_CGCM2_3_2A run)

For multiple runs from the same model we first calculated the model average and then the average of all models. In the fourth column the average ARI between the models and reality (NCEP) is given. In the parenthesis in the same column the model with the best agreement with NCEP is indicated

community structure of a field relatively well but not that of another field.

We then considered 70 20-century forced runs from the same 23 models and repeated the analysis (see supplementary material Tables S3 and S4 for a list of model run availability and for the correspondence between numbers and model runs). These runs attempt to reproduce the observed climate of the twentieth century by considering the forcing of CO_2 and of aerosols. Here we also include the National Center for Environmental Prediction/National

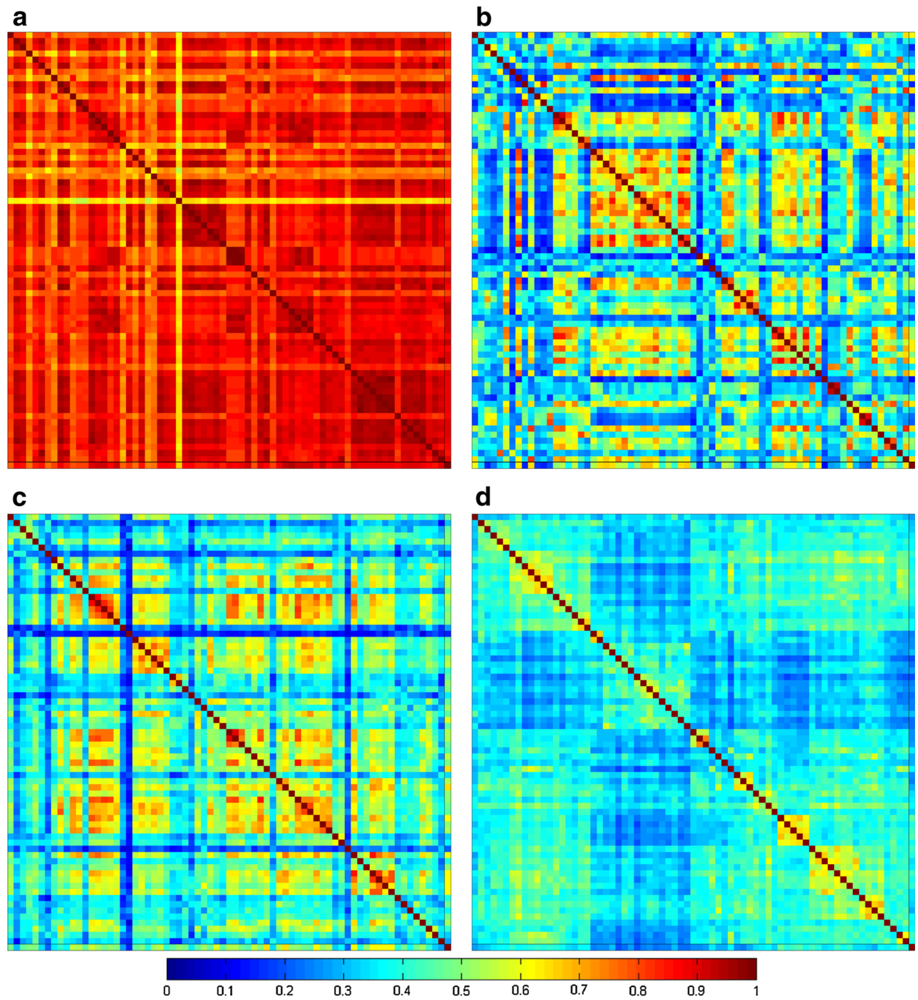

Fig. 2 Same as Fig. 1 but for 70 20-centuty forced runs from the same 23 models. Here NCEP (reality) is also included as number 71 (*last row/column*) separated from the runs by a thin black line (see Table S4 for correspondence between numbers and model runs). Qualitatively and quantitatively the results are very similar to those in Fig. 3 (see also Table 1)

Center for Atmospheric Research (NCEP/NCAR) Reanalysis 1 data set (Kistler et al. 2001) as a proxy for reality, assigned the number 71 (henceforth called NCEP for brevity).

The same picture emerges here as well (Fig. 2). Most models do well in simulating the features of the upper atmospheric flow, but not well in simulating the SLP, SAT, or precipitation fields. Note that, as in Fig. 1, different runs from the same model have a high ARI between them (which is rather expected; square areas of high ARI along the diagonal), but it is more visible here because now there is more runs available for the same models. Remarkably, the average ARI values (third column in Table 1) are basically identical to those for the pre-industrial controls. The fact that the community structure in forced and control simulations turns out to be, on average, very similar would then indicate that the only effect of forcing is to introduce a linear trend. Since the community approach is nonlinear, if the effect of forcing were nonlinear then removal of a linear trend would still include nonlinear effects, which will result in different community structures from those of the control runs. Therefore, we conjecture that the effect of forcing in the models is to "linearize" climate, which is highly questionable.

Also interesting is the fact that these values remain more or less the same when we calculate the average ARI between NCEP and all runs (fourth column in Table 1). The ARI between model runs and NCEP is shown on the last row and last column in Fig. 2, separated from the model runs by a thin black line. We note that, once again with the exception of the 500 hPa field, not only do the models not agree well with each other, but they do not agree with reality. Thus, our results indicate that the models are not capable to simulate the spatial structure of the temperature, sea level pressure, and precipitation field in a reliable and consistent way. This is an issue especially for SAT and precipitation, as those are the fields that are predicted to get projections of regional temperature and precipitation changes under forced scenarios. Even if the models manage to agree on global averages, they surely do not agree on regional changes.

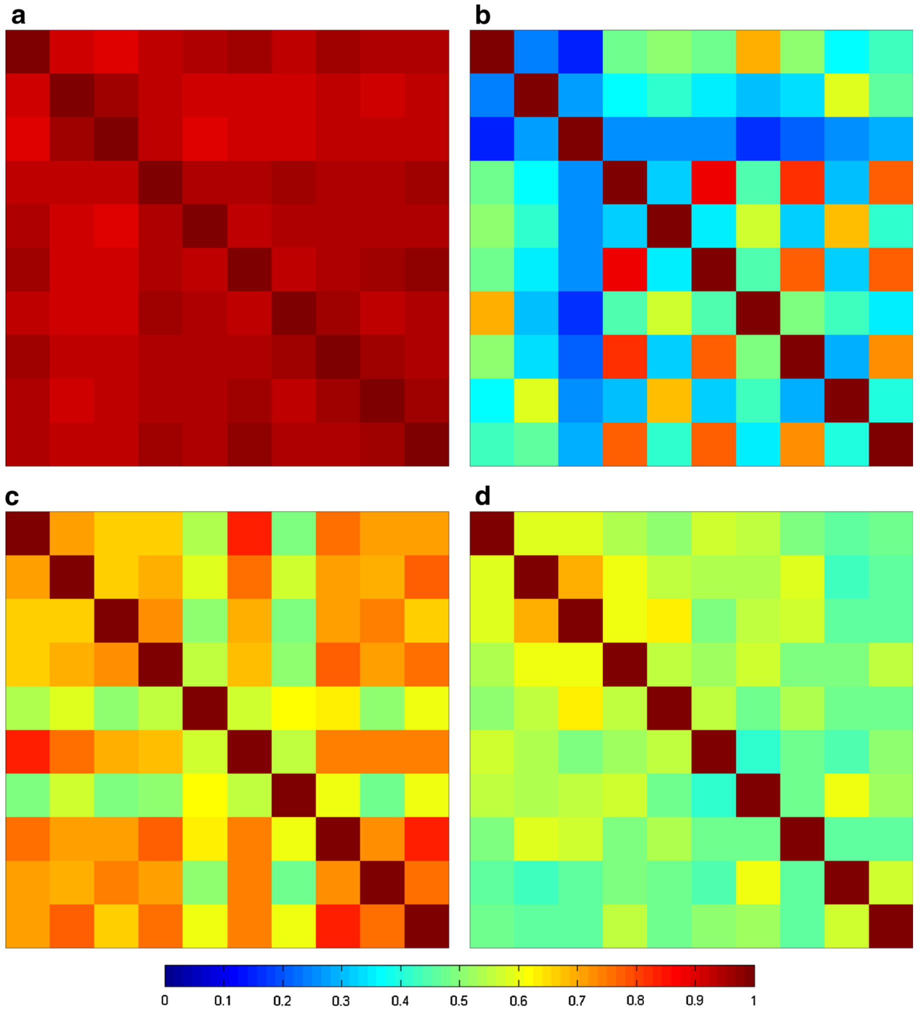

Fig. 3 This figure is similar to Fig. 1 but for 50-year sub-intervals of a 500-year control run (ECHAM5). The results indicate that the community structure is consistent throughout the 500-year period with higher and more uniform (with the exception of the PSL field) Adjusted Rand Indices than those in Fig. 1

To further ascertain the statistical significance of the measure of similarity between communities, we performed a Monte Carlo simulation wherein random partitions with the same number and size of communities was generated and the ARI re-computed; this experiment was repeated 1,000 times for each combination of runs (i.e., each cell in the similarity matrix shown in Fig. 1). The ARI of these random partitions was consistently close to zero, so we can be confident that the models do not produce random community structure. This is encouraging; however, the models are inconsistent and the agreement between them (except for the 500 hPa field) is not great.

4 Discussion

We don't mean to suggest that climate models do not have value. The parentheses in the fourth column in Table 1 give the models that agree most with reality (NCEP) for the four different fields. While no model or models emerge as superior, the ECHAM5 model agrees with NCEP 500 hPa to a high degree (ARI = 0.95), and the GISS and MIROC HiRes models have reasonable representation of the PSL (ARI = 0.68) and SAT (ARI = 0.65) fields, respectively. The MRI_CGCM does the best job for precipitation (ARI = 0.46).

Another encouraging result is shown in Fig. 3, which is similar to Fig. 1 but for 50-year sub-intervals of a 500-year pre-industrial control run (ECHAM5). The results indicate that, with the exception of the PSL field, the community structure is consistent throughout the 500-year period with higher and more uniform Adjusted Rand Indices than those in Fig. 1. This in turn suggests that models may not have a serious problem handling long-term simulations. Maybe the time has come to correct this modeling Babel and to seek a consensus climate model by developing methods which will combine ingredients from several models or a supermodel made up of a network of different models. For example, it has been shown that coupling imperfect copies of a simple dynamical system representing climate, which are assumed to be the truth, produces improved forecasts (Mirchev et al. 2012). Such an approach might draw on methodologies and experience from weather forecasting, where ensemble prediction methods have been developed and extensively used in operation for nearly two decades (Houtekamer and Derome 1995; Gneiting and Raftery 2005).

Acknowledgments AAT is supported by Department of Energy USA grant DE-0005305. KS is supported by the National Science Foundation grant IIS-1029711. Access to computing facilities was provided by the Minnesota Supercomputing Institute.

References

AchutaRao K, Sperber KR (2006) ENSO simulation in coupled ocean-atmosphere models: are the current models better? Clim Dyn 27:1–15

Albert R, Barabasi A-L (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47–101

Arenas A, Diaz-Guilera A, Perez-Vicente CJ (2006) Synchronization reveals topological scales in complex networks. Phys Rev Lett 96:114102

Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. Phys Rev E 70:066111

da Costa LF, Rodrigues FA, Travieso G, Villas Boas PR (2007) Characterization of complex networks: a survey of measurements. Adv Phys 56:167–242

Gneiting T, Raftery AE (2005) Weather forecasting with ensemble methods. Science 310:248–249

Gozolchiani A, Yamasaki K, Gazit O, Havlin S (2008) Pattern of climate network blinking links follow El Nino events. Europhys Lett 83:28005

Houtekamer PL, Derome J (1995) Methods for ensemble prediction. Mon Weather Rev 123:2181–2196

Kistler R et al (2001) The NCEP/NCAR 50-year reanalysis: monthly means, CD-ROM and documentation. Bull Am Meteorol Soc 82:247–267

Mirchev M, Duane G, Tang W, Kocarev L (2012) Improved modeling by coupling imperfect models. Commun Nonlinear Sci Numer Simul 17:2741–2751

Newman MEJ, Girvan M (2004) Finding and evaluating community structure in Networks. Phys Rev E 69:026113

Randall DA et al (2007) Cilmate models and their evaluation. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

Steinhaeuser K, Ganguly AR, Chawla NV (2012) Multivariate and multiscale dependence in the global climate system revealed through complex networks. Clim Dyn 39:889–895. doi:10.1007/s00382-011-1135-9

Strogatz SH (2001) Exploring complex networks. Nature 410:268–276

Swanson KL, Tsonis AA (2009) Has the climate recently shifted? Geophys Res Lett 36:L06711. doi:10.1029/2008GL037022

Tsonis AA, Swanson KL (2008) Topology and predictability of El Nino and La Nina networks. Phys Rev Lett 100:228502

Tsonis AA, Swanson KL, Roebber PJ (2006) What do networks have to do with climate? Bull Am Meteorol Soc 87:585–595. doi:10.1175/BAMS-87-5-585

Tsonis AA, Swanson KL, Kravtsov S (2007) A new dynamical mechanism for major climate shifts. Geophys Res Lett 34:L13705. doi:10.1029/2007GL030288

Tsonis AA, Swanson KL, Wang G (2008) On the role of atmospheric teleconnections in climate. J Clim 21:2990–3001

Tsonis AA, Wang G, Swanson KL, Rodrigues F, Da Costa LF (2011) Community structure and dynamics in climate networks. Clim Dyn 37:933–940. doi:10.1007/s00382-010-0874-3

Wang G, Swanson KL, Tsonis AA (2009) The pacemaker of major climate shifts. Geophys Res Lett 36:L07708. doi:10.1029/2008GL036874

Wyatt MG, Kravtsov S, Tsonis AA (2012) Atlantic multidecadal oscillation and northern hemisphere's climate variability. Clim Dyn 38:929–949. doi:10.1007/s00382-011-1071-8

Yamasaki K, Gozolchiani A, Havlin S (2008) Climate networks around the globe are significantly affected by El Niño. Phys Rev Lett 100:228501