



# Testing the detectability of spatio–temporal climate transitions from paleoclimate networks with the START model

K. Rehfeld<sup>1,2,3</sup>, N. Molkenthin<sup>1,2</sup>, and J. Kurths<sup>1,2</sup>

<sup>1</sup>Potsdam Institute for Climate Impact Research, P.O. Box 601203, 14412 Potsdam, Germany

<sup>2</sup>Department of Physics, Humboldt-Universität zu Berlin, Newtonstr. 15, 12489 Berlin, Germany

<sup>3</sup>Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Telegrafenberg A43, 14473 Potsdam, Germany

Correspondence to: K. Rehfeld (krehfeld@awi.de)

Received: 17 December 2013 – Revised: 18 March 2014 – Accepted: 22 April 2014 – Published: 25 June 2014

**Abstract.** A critical challenge in paleoclimate data analysis is the fact that the proxy data are heterogeneously distributed in space, which affects statistical methods that rely on spatial embedding of data. In the paleoclimate network approach nodes represent paleoclimate proxy time series, and links in the network are given by statistically significant similarities between them. Their location in space, proxy and archive type is coded in the node attributes.

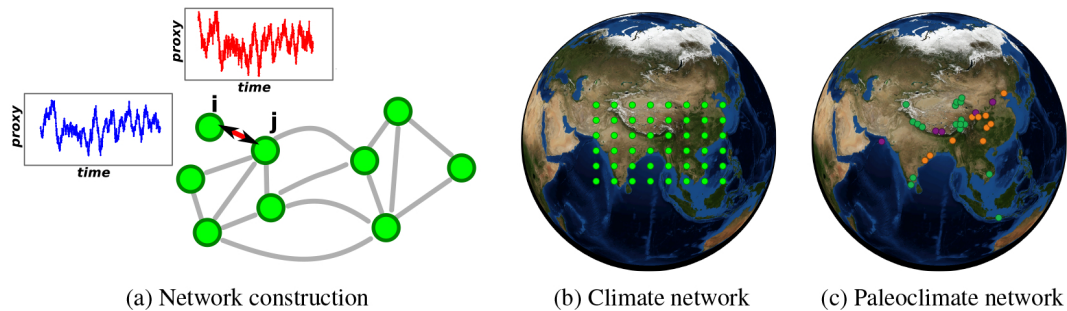
We develop a semi-empirical model for Spatio-Temporally AutocoRrelated Time series, inspired by the interplay of different Asian Summer Monsoon (ASM) systems. We use an ensemble of transition runs of this START model to test whether and how spatio–temporal climate transitions could be detectable from (paleo)climate networks. We sample model time series both on a grid and at locations at which paleoclimate data are available to investigate the effect of the spatially heterogeneous availability of data. Node betweenness centrality, averaged over the transition region, does not respond to the transition displayed by the START model, neither in the grid-based nor in the scattered sampling arrangement. The regionally defined measures of regional node degree and cross link ratio, however, are indicative of the changes in both scenarios, although the magnitude of the changes differs according to the sampling.

We find that the START model is particularly suitable for pseudo-proxy experiments to test the technical reconstruction limits of paleoclimate data based on their location, and we conclude that (paleo)climate networks are suitable for investigating spatio–temporal transitions in the dependence structure of underlying climatic fields.

## 1 Introduction

A growing number of paleoclimate records from environmental archives are available for past climate reconstruction. Fundamentally, this should increase the potential for successful reconstructions of the spatial and temporal features of past climatic changes, and thus enhance the general understanding of the climate system.

The paleoclimate network approach (Rehfeld et al., 2013), illustrated in Fig. 1, is a tool adapted to the challenges in environmental data analysis. As in the climate network approach (Tsonis et al., 2006; Donges et al., 2009b), *nodes* are identified with positions for which climate time series are available, and *links* are drawn between the nodes if statistically significant associations are found. The *adjacency matrix*  $\mathbf{A}$  is then a sparse binary matrix with the  $(i, j)$ th entry being non-zero if (and only if) the time series representing nodes  $i$  and  $j$  are *significantly* associated. Network statistics can then reflect global and local characteristics of the underlying data: the importance of a node, for example, can be measured by its degree, i.e., how many links the individual node has, or more abstract measures such as betweenness centrality (Opsahl et al., 2010; Barthélemy, 2011). Complex networks have been used to investigate the behavior of the climate system from reanalysis data (Donges et al., 2009a, 2011; Gozolchiani et al., 2011; Steinhäuser et al., 2010; Tsonis et al., 2010; Yamasaki et al., 2008) or recent observations (Ge-Li and Tsonis, 2009; Malik et al., 2010, 2011). The network methodology can, however, not be applied directly to paleoclimate data.



**Figure 1.** (a) Schematic illustration of paleoclimate network construction: nodes in the network represent paleoclimate archives with proxy time series. If the time series similarity estimate between a node pair is significant, these two nodes are considered to be linked. The connectedness of the obtained graph reflects the underlying dependencies created by climate processes. While the (b) climate network approach is based on gridded and dense observations of climatic parameters such as temperature or precipitation, the (c) paleoclimate network combines paleoclimate proxy records that are heterogeneously and sparsely distributed on the Earth’s surface.

The first reason is that the availability of paleoclimate data is heterogeneous in time, which implies that standard time series similarity measures are not applicable, since bias effects distort the results (Schulz and Stattegger, 1997; Rehfeld et al., 2011, 2013). For regular data the standard definitions for mutual information (MI, Cover and Thomas, 2006) or Pearson cross correlation (CC, Chatfield, 2004) can be used to obtain correlation matrices for the network. For paleoclimatic data sets, adapted techniques that are more robust against time series irregularity, such as Gaussian kernel-based cross correlation (Rehfeld et al., 2011), mutual information (Rehfeld et al., 2013) or the event synchronization function (ESF, Rehfeld and Kurths, 2014) need to be employed.

The effect of the spatially heterogeneous node distribution on network measures has also not received much attention so far, beyond the studies of Heitzig et al. (2010) and Rheinwald et al. (2012). In particular for node-based network measures that depend on non-local network topology, changes in network structure could cause non-trivial bias effects. A systematic investigation of the effects of spatial sampling on spatially embedded (climate) networks can be found in the paper of Molkenhain et al. (2014b).

In this paper we first review the paleoclimate network approach and identify potentially suitable network measures. Then we investigate how different network measures reflect distinct changes (transitions) in the underlying climate fields. To this end we develop and use the semi-empirical simplified model **START** (name derived from **Stream Transported Autocorrelated Time series**, or **Spatio-Temporally Autocorrelated Time series**) to simulate characteristic changes in a spatially extended domain. The model is based on an approximated solution of the advection-diffusion equation, which describes how temperature fluctuations are dissipated through stationary flow fields (Molkenhain et al., 2014a). A single forcing parameter can be varied to change the underlying flow, resulting in distinctly different fields, and in distinctly different climate network evolution

patterns (Tupikina et al., 2014). Transitions, in this context, are large-scale dependence changes due to changes in the atmospheric flow patterns. Using this computationally efficient and reduced model we test which climate network measures are suitable for characterizing the transition in the underlying climate field. In particular we evaluate if it is possible to infer the degree of interaction between different regions in the network and how spatially heterogeneously distributed nodes affect the analysis of spatio-temporal dynamics.

## 2 The paleoclimate network approach (PAN)

A schematic illustration of a paleoclimate network is given in Fig. 1a. Its nodes are given by paleoclimate proxy archives, its links by significant statistical association between the archives’ time series. As a method, climate networks ask the question “How dependent, linearly or nonlinearly, are the climate changes in place A on climate changes in another place B – and vice versa?”, rather than “Was the temperature in A strongly linearly correlated with the temperature in B at the same point in time?”, which is the case for standard empirical orthogonal function analysis (EOFs were used, for example, in Mayewski et al., 2004; Sinha et al., 2011; Yi et al., 2011). If, say, a proxy for local temperature in China co-varied significantly with reconstructed rainfall in India, this is caused by either (a) a common driving phenomenon (for example through the North Atlantic Oscillation, Wu et al., 2009 or solar forcing, Agnihotri et al., 2002), (b) local convective phenomena (for example internal ASM dynamics (Wang et al., 2010)), (c) an artifact in the reconstruction, for example non-climate-related common trends in the time series, or (d) a false positive (Rehfeld and Kurths, 2014).

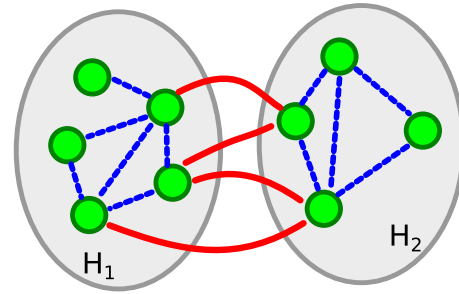
Consider a paleoclimate network as a graph  $G = (V, E)$  on a set of  $N$  vertices or *nodes*  $V$ , which are connected by a set of edges or *links*  $E$ . The *nodes* in graph  $G$  are embedded in space. Nodes have certain differing *properties* that include

their *position* on the Earth’s surface, the corresponding *type* of paleoclimate archive (e.g., tree, stalagmite, marine sediment) and *proxy* type (e.g., isotope ratios, lithogenic grain size or annual ring width). There exist time series whose archive sources are (a) distributed over a large area (Yi et al., 2011), or are (b) not Earth-bound, as for insolation (Steinhilber et al., 2009). Pragmatically, in the former case, the nodes should be placed in the center of the considered region. The latter should be considered as a node in a different *subnetwork* (see the illustration in Fig. 2 and the paragraph on subnetworks below).

Technically, each node is associated with at least one time series of an environmental proxy. If age uncertainty is considered, an ensemble of observation time vectors  $T_i$ ,  $i = 1, \dots, N_{\text{ens}}$  is generated and all ensemble members are considered equally likely realizations of the proxy time series of this node. For modeled network results, an ensemble of model realizations for the observation times  $t$  can be considered. The spatial distribution of nodes is also time dependent: while high-resolution archives (such as trees, Anchukaitis et al., 2006; Sano et al., 2011; Singh et al., 2009, but increasingly also stalagmites, Kennett et al., 2012; Hu et al., 2008; Tan et al., 2009; Fleitmann, 2004) predominantly cover more recent periods at annual timescales, others (e.g., marine or lacustrine sediment sequences, stalagmites or ice cores) might grow at very slow rates, and over the course of millennia (Dykoski et al., 2005; Rodbell et al., 1999; Fleitmann et al., 2007; Wang et al., 2001). A node will only be incorporated into the network evaluation for a time window  $W$  if it fulfills the minimal sampling requirements of  $> 50$  observations per window.

In the considered paleoclimate networks, *links* are undirected and weighted. A link between node  $i$  and node  $j$  exists, if the link *weight*, given by the link strength, is greater than zero (Rehfeld and Kurths, 2014). Generally, links in the paleoclimate network approach can not be assigned by simple thresholding of a similarity value  $S(X, Y)$  when the time series are irregular, of different length and/or age-uncertain (Rehfeld and Kurths, 2014). Therefore, significant similarity is established using surrogate data. The surrogate time series have the same temporal resolution as the original time series, but the observed proxy values are replaced using autocorrelated noise (c.f. Rehfeld et al., 2011; Rehfeld and Kurths, 2014). Using  $N_{\text{sim}}$  different similarity measures  $S$  with different characteristics and algorithms increases the robustness of the link detection, as the proxy-climate relationship might be nonlinear, weak or even erratic (Rehfeld and Kurths, 2014).

Fundamentally, for each pair of nodes  $i$  and  $j$ , time series ensemble member  $k$  and similarity measure  $l$ , a similarity  $S_{i,j}^{l,k}$  is calculated.  $N_{\text{sur}}$  autocorrelated but mutually uncorrelated time series surrogates are employed to obtain a distribution of surrogate similarity values  $S_{i,j}^{*l,k}$ . The *fundamental adjacency matrix* entry  $A_{i,j}^{l,k}$  consequently results from



**Figure 2.** To test (paleo)climatic hypotheses, the considered domain can be split and the nodes in the different subdomains associated with different sub-networks. This offers the possibility of differentiating statistically between associations within sub-networks (autolinks, in blue) and those connecting different sub-domains (cross links, in red).

thresholding the similarity value  $S_{i,j}^{l,k}$  using the chosen critical values  $S^*(q_{\text{low}})$  and  $S^*(q_{\text{hi}})$ , corresponding to the quantiles  $q_{\text{low}} = 0.05$  and  $q_{\text{hi}} = 0.95$  of the distribution  $S_{i,j}^{*l,k}$ :

$$A_{i,j}^{l,k} = 1 \text{ if } S_{i,j}^{l,k} < S^*(q_{\text{low}}) \text{ or } S_{i,j}^{l,k} > S^*(q_{\text{hi}}) \quad (1)$$

for asymmetric measures that distinguish between positive and negative similarity (e.g., CC), and

$$A_{i,j}^{l,k} = 1 \text{ if } S_{i,j}^{l,k} > S^*(q'_{\text{hi}}) \quad (2)$$

for symmetric measures that consider only an association strength (e.g., MI and the ESF), where  $q'_{\text{hi}} = 0.90$ .

The weight of a link between nodes  $i$  and  $j$  is given by the ratio of the number of significant associations between them for all  $N_{\text{ens}}$  ensemble realizations and  $N_{\text{sim}}$  similarity measures. This is summarized in the link weight matrix  $LW$ :

$$LW(i, j) = \frac{\sum_{k=1}^{N_{\text{sim}}} \sum_{l=1}^{N_{\text{ens}}} A_{i,j}^{l,k}}{N_{\text{ens}} \cdot N_{\text{sim}}} \quad (3)$$

### 2.1 Subnetworks

The nodes in the paleoclimate network  $G$  have different properties (e.g., archive type or geographic origin from a specific region) that may influence its role within the network. To investigate regional dynamics, these nodes are considered here as lying in different *subnetworks*.

A subnetwork  $H_1$  is formed by a subset of nodes  $V(H_1)$  and links  $E(H_1)$  from network  $G$ , where all nodes in  $V(H_1)$  fulfill a certain property (e.g., geographic location in the region of 60–100° eastern longitude and 0–40° northern latitude) and all links between these nodes. If, and how, the nodes are sorted into subnetworks depends on the research question that is being asked. Splitting the domain as above is motivated by the different ASM subsystems that are thought

to influence the regions west and east of 100° E. To investigate the dependence of the ASM on solar irradiance, a proxy record of insolation (e.g., Steinhilber et al., 2009) could then be considered as a node in a separate subnetwork.

Links within the subnetwork,  $E(H_1)$ , are *internal links*, and links from nodes  $V(H_1)$  to nodes in another subnetwork  $H_2$ ,  $V(H_2)$ , are *cross links*  $E(H_1, H_2)$ . Thus, the overall link set is the unity of internal and cross links between the subnetworks,  $E(G) = \cup_{i,j} E(H_i, H_j)$ , similar to Donges et al. (2011).

## 2.2 PAN measures

An abundance of graph-theoretical measures, i.e., statistics that are supposed to reflect characteristic node, link or network properties (Tsonis et al., 2006; Gozolchiani et al., 2011), is available. In the paleoclimate context, data-based challenges require an adaptation and careful evaluation of commonly employed (climate) network measures, because of age uncertainty and temporal and spatial heterogeneity. Here we test how well spatio-temporal changes are reflected in network measures, and to what extent they are influenced by spatially heterogeneous node distribution. Network measures considered for the test in the following include (a) the average *link density* (or connectivity), (b) the *regional degree* and *cross link ratio*, and (c) shortest path *betweenness centrality*.

In the following paragraphs we derive and review these basic network measures.

### 2.2.1 Average link density

For general complex networks, the link density, or connectivity, of a graph  $G$  with  $N_{\text{no}}$  nodes is simply the ratio of realized links between the nodes vs. the number of possible links

$$\text{LD}'(G) = \frac{\sum_{i,j} \mathbf{A}_{i,j}}{(N_{\text{no}} - 1) N_{\text{no}}}, \quad (4)$$

which is between zero and one.

However, the number of nodes of a paleoclimate network can vary if the minimal overlap between the time series is not always fulfilled. Therefore the node number may differ between the ensemble realizations  $N_{\text{no}} = N_{\text{no}}^l$ , and this results in a link density that depends on the realization number  $l$  and the link weight  $\text{LW}$ :

$$\text{LD}(l) = \frac{\sum_{i,j} \text{LW}_{i,j}^l}{(N_{\text{no}}^l - 1) N_{\text{no}}^l}. \quad (5)$$

This expression is averaged to obtain the *average link density* for the considered ensemble of time series,

$$\text{LD} = \frac{\sum_{l=1}^{N_{\text{ens}}} \text{LD}(l)}{N_{\text{ens}}}. \quad (6)$$

As all time series in this study are consistently sampled the network is thresholded such that the 20 % strongest links are considered significant.

### 2.2.2 Cross link probability

Assume network  $G$  consists of  $N_{\text{no}}^G$  nodes and  $N_{\text{ed}}^G$  edges. Let us partition this network into nodes in two subnetworks, say,  $H_1$  and  $H_2$  with  $N_{\text{no}}^{H_1}$  or  $N_{\text{no}}^{H_2}$  nodes each such that  $N_{\text{no}}^G = N_{\text{no}}^{H_1} + N_{\text{no}}^{H_2}$ . Accordingly, the sum of edges is partitioned into the sum of edges within  $H_1$  and  $H_2$ ,  $N_{\text{ed}}^{H_1}$  and  $N_{\text{ed}}^{H_2}$ , and edges from  $H_1$  to  $H_2$ ,  $N_{\text{ed}}^{1-2}$ .

The relative frequency of *realized* cross edges,

$$P'_{1-2} = \frac{N_{\text{ed}}^{1-2}}{N_{\text{no}}^{H_1} \cdot N_{\text{no}}^{H_2}}, \quad (7)$$

gives the *cross link probability*  $P_{1-2}$

$$P_{1-2} = \frac{\sum_{i \in H_1, j \in H_2} \text{LW}_{i,j}}{N_{\text{no}}^{H_1} \cdot N_{\text{no}}^{H_2}} \quad (8)$$

with a link weight matrix  $\text{LW}$  and  $\text{LW}_{i,j} \in [0, 1]$ .

### 2.2.3 Cross link ratio

The cross link ratio  $\text{CLR}(H_1, H_2, G)$  is given by the cross link probability divided by the overall link probability

$$\text{CLR}(H_1, H_2, G) = \frac{P_{1-2}}{\text{LD}}. \quad (9)$$

### 2.2.4 Average and regional node strength

In classical complex network theory, the *degree*  $D$  of a node  $i$  is a measure of the presumed importance of a node, given by the number of its links to all other nodes  $j = 1, \dots, N_{\text{no}}$ ,  $j \neq i$ :

$$D'(i) = \sum_{i,j \neq i} \mathbf{A}(i, j). \quad (10)$$

We consider the links to be weighted (cf. Opsahl et al., 2010, and references therein), and therefore the notion of the *degree* of a node was replaced by that of a *node strength*, also called the *vertex strength* (Gozolchiani et al., 2011; Opsahl et al., 2010). Using link weights this gives a *node strength*  $D^l$

$$D^l(i) = \sum_{i,j \neq i} \text{LW}_{i,j}^l \quad (11)$$

for each ensemble realization  $l$ .

### 2.2.5 Shortest path betweenness centrality

Betweenness centrality has been regarded as a measure of local dynamical information flow (Opsahl et al., 2010;



Barthélemy, 2011). The shortest path betweenness of a node  $k$  is calculated from the number of *shortest paths*  $\sigma_{ij}$  between all nodes  $i$  and  $j$ , and the number of these paths that pass through node  $k$ ,  $\sigma_{ij}(k)$ :

$$BC'(k) = \sum_{j \neq k} \frac{\sigma_{ij}(k)}{\sigma_{ij}}. \quad (12)$$

If the node number changes over time, we can standardize the betweenness to obtain a measure that is independent under such changes:

$$BC(k) = \frac{BC'(k)}{(N_{no} - 1)(N_{no} - 2)}. \quad (13)$$

### 3 START: spatio-temporally autocorrelated time series

#### 3.1 Model philosophy

START is a simple semi-empirical model for the propagation of climate variability through flows in a spatially extended domain. In the current implementation it is a statistical toy model with three independent spatial components that react differently to applied external forcing. Asian Summer Monsoon (ASM) dynamics are determined by the interplay between the Indian Summer Monsoon (ISM) and the East Asian Summer Monsoon (EASM). The predominant regions of influence, and main wind directions, of ISM, EASM and the continental westerlies are given in Fig. 3. They can be roughly divided at 100° eastern longitude (Wang et al., 2010). The dependence – and interplay – of the subsystems on each other is a topic of ongoing research (Wang et al., 2010; Yihui and Chan, 2005; Cao et al., 2012).

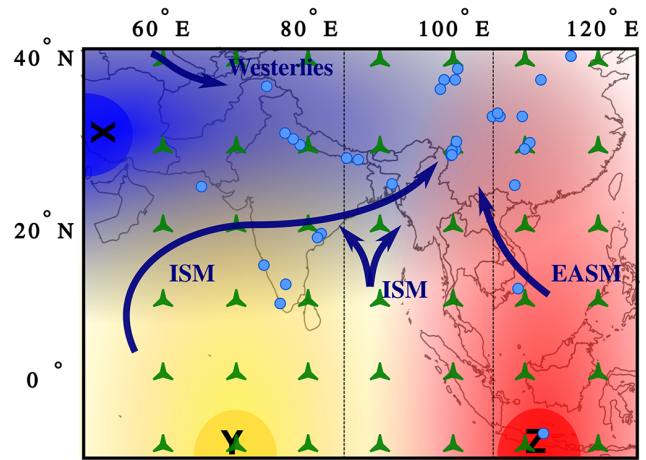
Assuming that we can view the model domain in summer as a region with three main wind systems (cf. Fig. 3) that each extend only zonally, i.e., laterally or longitudinally. While the true wind pattern might be significantly more complex, we argue that this could be viewed as a statistical decomposition of the mean summer surface wind field into simplified lateral/longitudinal components  $V_X(\mathbf{p})$ ,  $V_Y(\mathbf{p})$  and  $V_Z(\mathbf{p})$ . Each of these fields are assumed to be a Gaussian-modulated unidirectional front with a velocity at position  $\mathbf{p}$  and time  $t$

$$v_X(t, \mathbf{p}, m_X, W) = m_X e^{-(p_x - p_{0,x})^2 / 2W}, \quad (14)$$

with a full width at a half maximum of  $2\sqrt{W \log 2}$ . The maximal amplitude of the velocity,  $m_X$ ,

$$m_X(t, F) = m_X = B_X + \alpha F, \quad (15)$$

is found in the center of the Gaussian front, as in Fig. 3. Here,  $B_X$  is the baseline strength of the component's flow, and  $\alpha$  is its amplitude, or susceptibility to the external forcing, represented by the parameter  $F$ ,  $F = [-1, \dots, 1]$ . The velocities



**Figure 3.** Map showing the main wind directions of the Indian and East Asian summer monsoon systems. Inflow corridors are modeled as sources of variability: the Indian Summer Monsoon (ISM) with a longitudinal ( $X$ , in blue) and a latitudinal component ( $Y$ ) and the East Asian Summer Monsoon ( $Z$ , EASM) with a latitudinal component. The dynamics in the model, governed by the respective strengths of the source flows, are sampled at the grid locations (triangles) and where paleoclimate data is available (circles). Dashed lines bracket the intermediate domain.

for sources  $Y$  and  $Z$  are defined analogously, and the chosen values are given in Table 2.

Each of the fields originates from a source at a position  $\mathbf{p}_{src}$ , and each of the sources  $X$ ,  $Y$  and  $Z$  is associated with a climate process  $X_t$ ,  $Y_t$  and  $Z_t$  that represents the annual mean of a hypothetical climate variable, for example surface temperature or precipitation anomalies, in the year  $t$ . By definition the model is restricted to modeling inter-annual variability, and the construction rationale is that local variability can be modeled as a superposition of variability mediated by atmospheric flows and local variability. The amount of dynamical information about the climatic process at the source that flows along one of these fields, say, from source  $X$  to a point at a position  $\mathbf{p}$  in its region of influence, is approximated by a *variance factor*  $f_X(\mathbf{p}, F)$ . By construction, the square of the variance factor is proportional to the amount of variance shared between the source  $X$  and the time series at point  $\mathbf{p}$ :  $f_X^2(\mathbf{p}, F) \propto \sigma_X(\mathbf{p})$ .

We assume three sources for the underlying flow system, where random climate variability originates and which is then transported via advection and diffusion along the paths. The position and transmission direction of the sources and the observation points are illustrated schematically in Fig. 3. At each point in the ASM region a local time series of climate variability is computed as the sum of the noise contributions from each of the three sources. These components are scaled with a factor that quantifies the amount of information that is preserved from the source to the point of observation:

$$R_i = f_X(F, i)R_X + f_Y(F, i)R_Y + f_Z(F, i)R_Z + R_{noise} \quad (16)$$

where  $R_i$  is the signal at point  $i$  obtained from the superposition of  $R_X$ , the signal of the longitudinal ISM component,  $R_Y$ , the latitudinal ISM component and  $R_Z$ , the signal of the EASM source.  $R_{\text{noise}}$  is local observation noise. The factors  $f_X(F, i)$ ,  $f_Y(F, i)$  and  $f_Z(F, i)$  scale the contributions to the overall variance of the time series obtained at point  $i$  at time  $t$  and a potential forcing  $F$ .

### 3.2 Derivation of the scaling factors

The system we are looking at is a two-dimensional boundary-less fluid of constant diffusivity  $\chi$  with a stationary flow described by the velocity field  $\mathbf{v}(\mathbf{x})$ . Temperature transport in the system is governed by the advection–diffusion equation, which states how the change in temperature over time is determined by the spatial temperature change and the velocity:

$$\frac{\partial T}{\partial t} = \chi \Delta T - \nabla \cdot (\mathbf{v}(\mathbf{x}) T), \quad (17)$$

which is obtained by inserting the advective and diffusive flux

$$\mathbf{j} = \mathbf{j}_{\text{diff}} + \mathbf{j}_{\text{adv}} = -\chi \nabla T + \mathbf{v} T \quad (18)$$

into the sourceless continuity equation for temperature

$$\frac{\partial T}{\partial t} = -\nabla \cdot \mathbf{j}. \quad (19)$$

$T(\mathbf{p}, t)$  is the temperature value at position  $\mathbf{p}$  at time  $t$ .

To compute the scaling factors we approximate and solve Eq. (17) with temperature spikes along the source origin line, perpendicular to the flow direction. These spikes are propagated along a Gaussian-modulated velocity field, and the scaling factors are determined from the remaining spike height at position  $\mathbf{p}$ . We use this temperature  $\delta$  peak as a tracer of the flow. The initial condition is a Gaussian-shaped temperature front of unit height (in the  $x$  or  $y$  direction)

$$T(\mathbf{p}, 0) = e^{-\frac{(p-p_0)^2}{s}}, \quad (20)$$

where  $s$  is the Gaussian front width and  $p_0$  the position of the source. Local temperature is computed as a function

$$T(\mathbf{p}, t) = \sqrt{\frac{s}{s + 4\chi t}} e^{-\frac{(p-p_0-vt)^2}{s+4\chi t}} \quad (21)$$

of time and space. For the constant  $v$  this can be solved analytically. The diffusion constant  $\chi$  is set to unity. We neglect the derivative of the velocity field but replace  $v$  by  $\mathbf{v}(\mathbf{p})$  and thereby get an approximate solution for velocity fields with a slow spatial variation. We also assume that this velocity field depends on a given forcing  $F$ . As this can be seen as a statistical description of how one original disturbance would dissipate over space, we use it to define the local variance

factors:  $f_X^*(T, i) = T(i, t_{\text{max}})$  with the front in the  $y$  direction and  $p_0 = y_X$ ,  $f_Y^*(T, i) = T(i, t_{\text{max}})$  with the front in the  $x$  direction and  $p_0 = x_Y$ ,  $f_Z^*(T, i) = T(i, t_{\text{max}})$  with the front in the  $x$  direction, and  $p_0 = x_Z$ .

The factors depend on the observation position  $\mathbf{p}$ , source positions  $\mathbf{p}_X$ ,  $\mathbf{p}_Y$  and  $\mathbf{p}_Z$  and the velocity component in flow direction,  $v_X(t, F)$ :

$$f_X^*(\mathbf{p}, F) = f(\mathbf{p}_X, \mathbf{p}, v_X, \chi, s). \quad (22)$$

### 3.3 Obtaining the START time series

At each point in space and for each time point  $t$  the factors are standardized by the factor sum  $g(t, F)$ ,

$$g(t) = f_X^*(F) + f_Y^*(F) + f_Z^*(F) + f_{D,t}^*, \quad (23)$$

i.e.,  $f_X(F) = \frac{f_X^*(F)}{g(t)}$ . Here, the processes  $X_t$ ,  $Y_t$  and  $Z_t$  are uncoupled AR(1) processes of unit variance, and with a persistence time of  $\tau = 6.4$  years. This value for  $\tau$  was chosen as compatible with the order of magnitude estimated for ASM paleoclimate proxy records spanning the last millennium (Rehfeld et al., 2011). At each point  $\mathbf{p}$  in the model domain, the local climate “history”  $S_i(\mathbf{p})$  is computed as a superposition of source and noise terms (Eq. 16).

### 3.4 Model setup

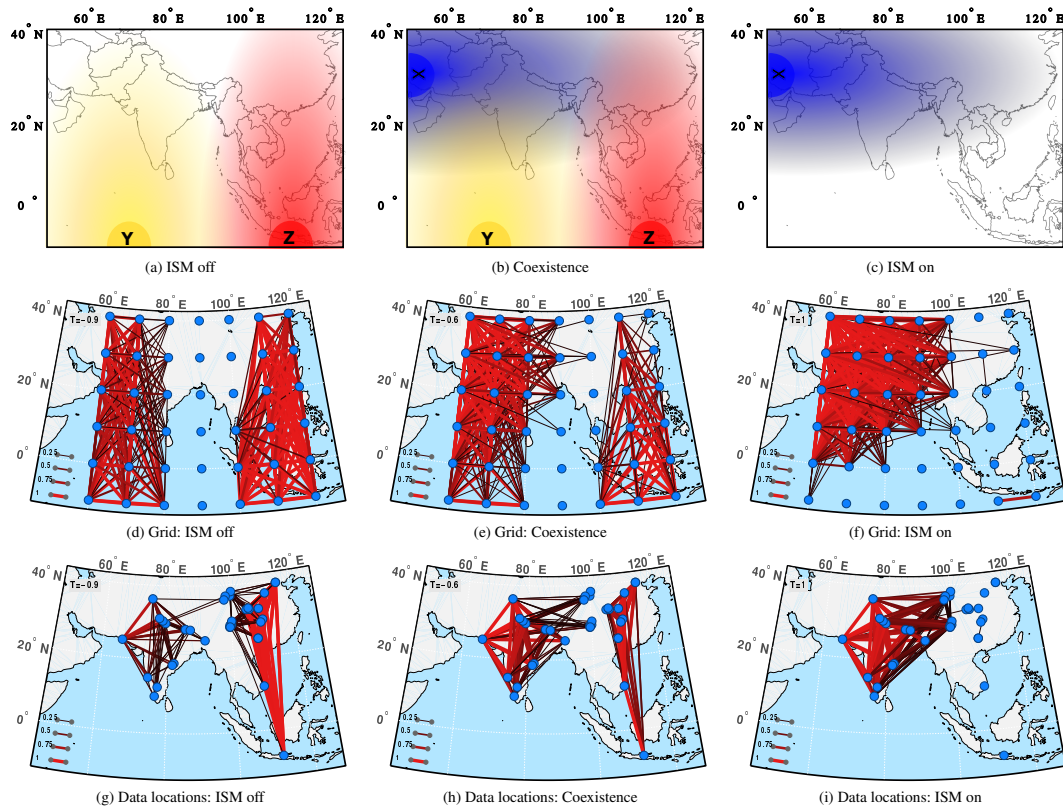
START generates synthetic time series at locations in its integration domain. The interdependences between the time series depend on their relative position in the considered flows and the given forcing. Although each time series is distinct, as the influence of the different components is location dependent, time series located in close proximity are similar, and the amount of variance shared with the components’ sources is both location and forcing dependent.

The transient forcing model runs are sampled for two spatial sampling types, a grid and a data set of locations of paleoclimate records (c.f. Table 1 and Fig. 3) throughout the ASM domain. The regional span of the spatial sampling and the node numbers is comparable, although the archive locations are spaced closer at the center of the ASM domain, while the grid also samples the areas over the Indian Ocean.

Networks are computed for twenty 200-year-long time slices of a 4000 year transient simulation, during which the model forcing parameter,  $F$ , was increased consistently from its minimum value,  $-1$ , to  $1$ . To ensure the robustness of the intended spatial inference against estimation errors for these relatively short time series, 100 simulation realizations were analyzed separately.

## 4 Validation of PAN using START

To validate the PAN methodology we run START with different forcing parameters and investigate the resulting network topology. In transient runs we vary the parameter continuously. Three stages are distinguished in particular: “ISM



**Figure 4.** Extreme points of the modeled Asian monsoon dynamics: schematic illustration of input variance factors (a)–(c) and networks reconstructed on grid (d)–(f) as well as the actual data positions (g)–(i). Only the 20 % strongest links are shown.

**Table 1.** Spatial setup for the START model experiments.

	Grid	Heterogeneous
Number of nodes	42	36
Regional span [latitude]	[−10; 40]	[−7; 39.5]
Regional span [longitude]	[60; 120]	[66; 115.5]
Node distribution	Regular grid	Paleoclimate record origins, c.f. Rehfeld et al. (2013)

off”, equivalent to forcing  $F = -1$ , “Coexistence” with  $F = 0$  and “ISM on” with  $F = 1$  (c.f. Fig. 4).

The velocities in the Gaussian-shaped fronts in Eq. (22) are modulated through the forcing parameter  $F$  in the START model. Consequently the amount of variance conserved along a flow and the synchronizing reach of the three “wind components” change with changing forcing, as illustrated schematically in Fig. 4a–c. The components are tuned to react proportionally to the effected forcing, but compete at each point in space due to the standardization in Eq. (23).

Therefore, the fraction of variance explained by the components in each location changes in a nonlinear way. With parameter  $F$ , the reach of the  $X$  component increases while the other components lose relevance.

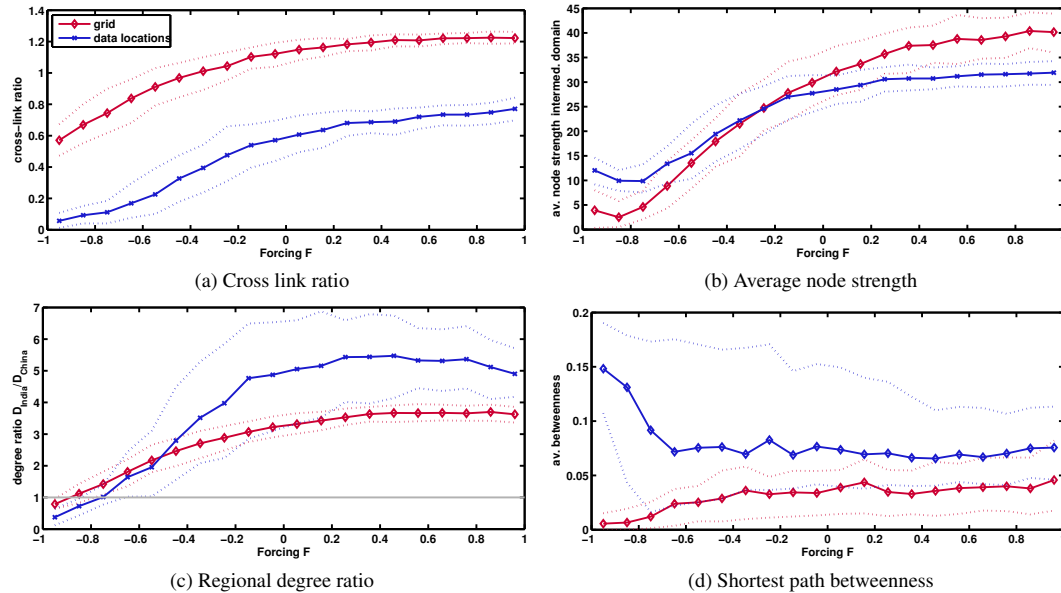
#### 4.1 Topology of the observed networks

The networks obtained from the 20 % strongest time series correlations are given in Fig. 4d–f for the grid-based time series, and in Fig. 4g–i for the proxy locations. In the first, “ISM off” case in Fig. 4d and g the latitudinal components  $Y$  and  $Z$  are strong, resulting in two clearly separated network components. The  $Y$  component covers the longitudes 60–80° E, while the  $Z$  component covers the longitudes 100–120° E, and no strong links appear between the two.

As the forcing increases, a “coexistence” stage is reached: the longitudinal  $X$  component strengthens, and the latitudinal components  $Y$  and  $Z$  lose in relation, as they have opposite sensitivities to forcing (c.f. Table 2). As Fig. 4e and h show, the reconstructed network is still split, as the strongest links are in the core region of the  $Y$  and  $Z$  flow parts. However, strong links, originating far in the west, connect to the 90° E grid points, and weaker links extend even beyond. The  $Z$ -component half of the network has retracted southward, in agreement with the decreased relative forcing strength.

**Table 2.** START source and flow attributes.

	Source	Position	Strength $B$	Ampl.	Width $W$
ISM (long. component)	$X$	(30, 55)	85	$\alpha = 70$	1200
ISM (lat. component)	$Y$	(-15, 70)	10	$\beta = -5$	100
EASM	$Z$	(-15, 112.5)	30	$\gamma = -12$	100
White noise	$N$	all	–	Const.	–



**Figure 5.** Most network measures reflect the increasing change in network structure occurring with changing ISM dominance in START (a)–(d), but they show different sensitivities. The spatial distribution of node positions influences the absolute value of the network measure, but the general trends are consistent for the underlying grid and the heterogeneous data positions. Broken lines indicate the 10 and 90 % quantiles of the ensemble results.

With maximal forcing the full “ISM only” stage is reached in Fig. 4f and i. The  $X$  component dominates, and the reconstructed network has only a single core region extending to and across  $100^\circ$  longitude. The changing strength of the synchronizing components due to the varied forcing is reflected by the extent of the synchronized regions in which the grid-point time series show the strongest similarities. Although differences exist, the networks reconstructed from grid and heterogeneous locations bear a large resemblance and relate to the underlying forcing structure.

## 4.2 Validation of the network measures

Network measures are statistical estimators that reflect properties of individual nodes as well as regional or global characteristics of the network. The spatio-temporal changes, visible as the START model is driven with different forcing, should effect consistent transitions in measures suitable for the investigation of spatio-temporal changes. Furthermore, these transitions should also be detectable under varying node distributions, i.e., for a grid structure as well as for

heterogeneous node distribution. The gradually increased forcing results in a transition from a bimodal, laterally synchronized network with separated components to a widely connected state. Figure 5 shows the network measure expectation values from 100 transient START forcing runs.

### 4.2.1 Cross link ratio

In the grid-based network reconstruction the *cross link ratio* (Fig. 5a), calculated as the ratio of cross link probability over overall link probability, increases monotonously for the grid-based networks and saturates well before maximal forcing is reached. It crosses the threshold of equal probabilities, transitioning from values of 0.6 to 1.2. This means that while in the early, bimodal, stage the cross link density is about 60 % of the average link density, it reaches 120 % when the synchronized region spans the whole network. Thus, at low forcing the cross link density is significantly *lower* than the overall link density, with the network being effectively partitioned. At high forcing the co-varying region spans across the former separate parts, resulting in a *higher* cross link density



than average link density. For the data-based network, the expected path of transition for the heterogeneous network differs from that observed for the grid topology by a negative offset, but parallels it otherwise.

#### 4.2.2 Average node strength in the intermediate domain

The *intermediate domain* (85–105° E) between the defined *model* ISM/EASM core regions is not synchronized by an external source of variability for low forcing values. Thus, nodes within this domain are expected to have no, or few, connections, resulting in weak node strength, as illustrated in Fig. 5b. This changes, however, as the longitudinal X component strengthens with increasing forcing, when these nodes fall into its region of influence. Please note that this does not tie in exactly with the actual ISM/EASM transition region, which is situated farther to the east. This adaptation was necessary because the implemented flow paths in START are modeled for simplicity either laterally *or* longitudinally, and the core interaction region around 100° E should be reachable for both flows. Indeed, as shown in Fig. 5b, the average node strength in the intermediate domain rises from a low level to full connectivity (node strength equal to  $N_{\text{no}} - 1$ ) for the grid-based network measures after an initial short decline. The amplitude of the change is, however, lower for the heterogeneous locations.

#### 4.2.3 Regional node strength ratio

The regional node strength ratio (Fig. 5c) highlights the different regional degrees in subregions of the network. Here it is computed using Eq. (10) as the ratio of the average node strengths in India vs. that in China. If both subnetworks are equally well connected within the overall network, their average node strength should be similar, and the node strength ratio should equal unity, as indicated by the gray line in Fig. 5c.

Starting off at equally well-connected subdomains, the western, “Indian” part of the network gains importance with increased forcing and the node strength ratio settles after a short growth slightly below 4 for the grid-based network. Thus, nodes in “India” are associated with four times the link weight when compared to those in “China”. For the heterogeneous sampling scheme the line of equal node strength is crossed *later* than for the gridded data. The following incline, contrastingly, is sharp and the plateau reached gives fivefold node strength for “India” vs. “China”, accompanied by a large uncertainty in this estimate. This, c.f. Fig. 3, is consistent with the stronger representation of “China” in the used grid.

#### 4.2.4 Shortest path betweenness

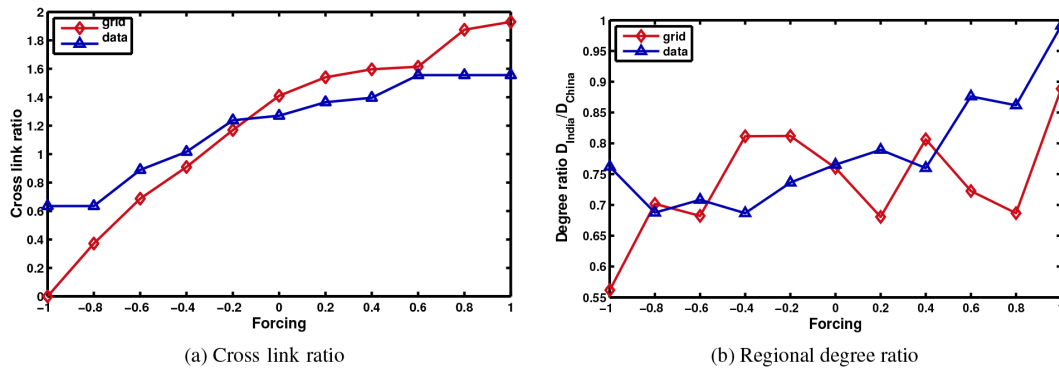
*Shortest path betweenness centrality* is a measure developed to infer the presumed relative importance of nodes and regions (Gozolchiani et al., 2011; Opsahl et al., 2010; Barthélemy, 2011; Donges et al., 2009a). Furthermore,

Donges et al. (2009b) found that “betweenness centrality allows to measure the importance of localized regions on the earth’s surface for the transport of dynamical information within a climatological field in the long term mean”, and stated that “information is transported by advective processes, where the assumption of information traveling on shortest paths can be substantiated by extremalization principles”. As such it should, in principle, also be an interesting measure for paleoclimate network applications.

As a node-based measure it is not possible to compare it directly when using different spatial sampling schemes. Still, the characteristic dynamical changes should be reflected in regional properties. Based on this presumption, betweenness centrality estimates for nodes in the intermediate and central region (85–105° E) are averaged to obtain a domain estimate. The results, shown in Fig. 5d, are inconsistent for the first segment of the transition experiment, in which the largest dynamical changes occur. On the one hand, the betweenness estimates for the grid *increase* slowly, albeit with a comparatively large uncertainty. On the other hand, the estimates for the heterogeneous locations *decrease* initially from large initial values to then stay on a flat plateau.

#### 4.3 Comparison to flow network results

For comparison we also computed the networks analytically, directly from the flow using a continuously defined cross-correlation analog and a solution of the ADE from Molkenhain et al. (2014a). Figure 6 shows the results obtained for the cross link ratio and the regional degree ratio, where the average node strengths West/East of the 100° longitude boundary are compared. Both grid-based and data location-based results agree well with those obtained from the START time series for the cross link ratio, showing the gradual increase in the cross link density as the forcing parameter increases. The regional degree ratio shows substantially more variability than for START, though an overall positive trend is apparent. This may be due to the fact that the flow networks, by construction, have an overall higher connectivity due to a higher correlation level than the START networks. Please note that by construction the absolute forcing values are not equivalent for START and the flow networks and therefore the results can not be compared quantitatively. The main construction difference is that flow networks do not have designated fluctuation sources, but average over all possible source locations. By contrast, sources in START (e.g., X, Y and Z in Fig. 3) are the origin of variability. Links closer to the sources are more likely than those far away from them in START.



**Figure 6.** Cross link ratio (a) and average regional degree ratio (b) for networks computed directly from the flow data (Molkenthin et al., 2014a, b). Note that due to the necessarily different implementations the results can not be quantitatively compared to those obtained with START.

## 5 Discussion

### 5.1 Network measures

The cross link ratio reflects the transition in the model simulations well. While the paths for different node distributions differ substantially in their amplitude, baseline and speed of increase, both show the overall increase in subdomain-connecting cross links. The initial decline observed for the data topology results from spurious cross links, as the overall similarity level (not shown) is much lower than at higher forcing levels. These cross links appear randomly at low forcing, but disappear if the link density is chosen more conservatively. Heterogeneous sampling apparently also does not strongly affect the node strength in the intermediate domain. Both sampling schemes reflect the transition of these nodes from being irrelevant to heavily tied to the rest of the network.

The regional node strength ratio is more difficult to interpret. The expected transition path for the relative average strength of nodes in India vs. that of those in China is significantly different for differing sampling schemes. The general feature, however, that repeats itself is that proximity to the source clearly results in higher node strength. There is nevertheless large uncertainty associated with the transition path for the data-based network.

The betweenness centrality estimates are inconsistent for the different sampling schemes, and the error margins spread widely. This inconclusiveness could be due to hypersensitivity of the measure and the comparatively low number of nodes in our network.

### 5.2 Regional changes and inter-regional information flow

Three major regions are relevant for START dynamics: the ISM region, the intermediate region and the EASM region. While the first and the latter start off as independent

subdomains of the network, they are connected by the longitudinal ISM component at increased forcing. The distinct dynamical features include (i) bimodality vs. later unimodality, (ii) increasing size of the spatially synchronized region and (iii) increased flow of dynamical information through the increasing strength of the longitudinal ISM component.

The initial *bimodality* of the network is directly visible in the network, and in the low cross link ratio and average node strength in the intermediate domain. Although the absolute values of the between cross link ratio are not on the same scale as those for the grid, a significant increase occurs. Thus, the change in model dynamics can be inferred if the node topology does not change and if the sampling bias (defined here as a systematic offset due to an other-than-regular node topology) can be quantified. Node strength in the transitional region shows smaller sampling-dependent deviations and could thus be a more robust measure of the importance of intermediate regions. The regional degree ratio shows, however, that spatially biased sampling can have large effects: as only one node of the paleoclimate network samples close to the EASM source, and most of the rest of the clusters far north of it, the synchronizing influence of this source quickly disappears as the ISM region grows. Located at the fringe of the ISM component, these nodes also get less ISM input than all its westwards neighbors. This results in few links and an under-representation of the Chinese part of the network, caused by a combination of model shortcomings and data sparsity. Such effects have to be addressed before a comparison of networks with changing node architecture is possible. The increasing strength of the ISM component is directly visible by its growth in the reconstructed network, and additionally reflected in the increasing cross link ratio. The source region of the modeled pathway is robustly characterized by a higher node degree. Increasing information flow from the ISM to the EASM core region is indecipherable using shortest path betweenness, which is sensitive to sampling changes and spurious links. Nevertheless, provided the potentially changing node topology is addressed, the cross

link ratio could be a sufficient measure to quantify changes in inter-regional dependences.

What remains to be investigated is the influence that varying node numbers – in space and time – have on estimated network measures. This could be done by simulating constant START “climate dynamics” and removing nodes iteratively. The relevance of the spatial position of a node can then be assessed by the discrepancy between the network measures estimated with/without it in comparison to the expected value for high-resolution sampling.

## 6 Conclusions

We conclude that it is also possible to reconstruct the spatio-temporal changes in the semi-empirical ASM model for heterogeneous node distribution in space, though the results are limited to qualitative statements if the node topology changes. Model dynamics are reflected in the reconstructed network, specifically its link strength distribution, and in network measures such as regional average node strength and node strength in the transitional zone, the intermediate domain outside the latitudinal source influence.

Inter-regional information flow, or, in the context of the START model, spatial distribution of variance, cannot be inferred using shortest path betweenness, as it is found to be too sensitive to irregularities, and no clear dynamical signature can be found in the transition experiments. The cross link ratio is a better alternative, though sampling biases have to be taken into account for its analysis.

Spatial heterogeneity, in general, has strong effects, both on the reconstructed network and on the network measures. It manifests itself in (i) biases in network measures that can be negative (cross link ratio) as well as positive (regional degree ratio), (ii) increased variance in the estimates (betweenness centrality, regional degree ratio), and (iii) the amplification of effects due to node clustering (regional degree, reconstructed network). Therefore, if networks with varying spatial sampling are investigated, care has to be taken to perform adequate significance tests to ensure that spurious sampling effects can be distinguished from real climate processes.

The developed ASM model START is a toy model that can not be expected to reflect actual monsoon dynamics. In reality, local, global and external forcing influences local climate processes. In the START model world, information transfer and local climate processes are governed solely by physical flows. Processes external to the ASM domain are not considered but can, in nature, lead to increased correlation in the whole or parts of the ASM region. One of the desired features to improve realism, for example, would be the inclusion of regional sources of variance, i.e., by region-dependent noise terms. Then the propagation of information could be considered serially, and causality-sensitive directed measures (Granger causality, ESF) could be tested. In the model’s simplicity, however, also lies its strength, because it

is possible to interpret the results with respect to its dynamics, a task that is much more complicated if such pseudo-proxy experiments are conducted with actual global climate models (GCMs) (Smerdon, 2012; von Storch et al., 2004; Mann and Rutherford, 2002). Unlike GCMs, START is computationally inexpensive; therefore large ensembles of time series for pseudo-proxy experiments can be generated. Using START, hypotheses concerning local vs. global drivers of climate dynamics can be tested directly based on the paleoclimate data, because START models the propagation of local climate variability through advection and diffusion. For large-scale dynamical and coupled GCMs with their multitude of output variables and parameters, cause and effect are more difficult to discern. Thus, it provides a good opportunity to assess whether and how spatio-temporal dynamics of a given paleoclimate data set are affected by age uncertainty, spatio-temporal heterogeneity and sparsity.

*Acknowledgements.* The authors thank Liubov Tupikina, Veronika Stolbova and Norbert Marwan for stimulating discussions. Funding by the German Science Foundation (DFG graduate school 1539), the German Federal Ministry for Education and Research (BMBF project PROGRESS) and via Helmholtz grant VG-900NH is gratefully acknowledged.

Edited by: J. Davidsen

Reviewed by: two anonymous referees

## References

- Agnihotri, R., Dutta, K., Bhushan, R., and Somayajulu, B.: Evidence for solar forcing on the Indian monsoon during the last millennium, *Earth Planet. Sc. Lett.*, 198, 521–527, doi:10.1016/S0012-821X(02)00530-7, 2002.
- Anchukaitis, K. J., Evans, M. N., Kaplan, A., Vaganov, E. A., Hughes, M. K., Grissino-Mayer, H. D., and Cane, M. A.: Forward modeling of regional scale tree-ring patterns in the southeastern United States and the recent influence of summer drought, *Geophys. Res. Lett.*, 33, 2–5, doi:10.1029/2005GL025050, 2006.
- Barthélemy, M.: Spatial networks, *Phys. Rep.*, 499, 1–101, doi:10.1016/j.physrep.2010.11.002, 2011.
- Cao, J., Hu, J., and Tao, Y.: An index for the interface between the Indian summer monsoon and the East Asian summer monsoon, *J. Geophys. Res.*, 117, 1–9, doi:10.1029/2012JD017841, 2012.
- Chatfield, C.: *The analysis of time series: an introduction*, 6th Edn., CRC Press, Florida, USA, 2004.
- Cover, T. and Thomas, J.: *Elements of information theory*, 2nd Edn., John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.
- Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: The backbone of the climate network, *Europhys. Lett.*, 87, 48007, doi:10.1209/0295-5075/87/48007, 2009a.
- Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: Complex networks in climate dynamics, *Eur. Phys. J. Spec. Top.*, 174, 157–179, doi:10.1140/epjst/e2009-01098-2, 2009b.

- Donges, J. F., Schultz, H. C. H., Marwan, N., Zou, Y., and Kurths, J.: Investigating the topology of interacting networks, *Eur. Phys. J. B*, 84, 635–651, doi:10.1140/epjb/e2011-10795-8, 2011.
- Dykoski, C., Edwards, R., Cheng, H., Yuan, D., Cai, Y., Zhang, M., Lin, Y., Qing, J., An, Z., and Revenaugh, J.: A high-resolution, absolute-dated Holocene and deglacial Asian monsoon record from Dongge Cave, China, *Earth Planet. Sc. Lett.*, 233, 71–86, doi:10.1016/j.epsl.2005.01.036, 2005.
- Fleitmann, D.: Palaeoclimatic interpretation of high-resolution oxygen isotope profiles derived from annually laminated speleothems from Southern Oman, *Quaternary Sci. Rev.*, 23, 935–945, doi:10.1016/j.quascirev.2003.06.019, 2004.
- Fleitmann, D., Burns, S. J., Mangini, A., Mudelsee, M., Kramers, J., Villa, I., Neff, U., Al-Subbary, A. A., Buettner, A., Hippler, D., Matter, A., and Alsubbary, A.: Holocene ITCZ and Indian monsoon dynamics recorded in stalagmites from Oman and Yemen (Socotra), *Quaternary Sci. Rev.*, 26, 170–188, doi:10.1016/j.quascirev.2006.04.012, 2007.
- Ge-Li, W. and Tsonis, A. A.: A preliminary investigation on the topology of Chinese climate networks, *Chinese Phys. B*, 18, 5091–5106, 2009.
- Gozolchiani, A., Havlin, S., and Yamasaki, K.: Emergence of El Niño as an Autonomous Component in the Climate Network, *Phys. Rev. Lett.*, 107, 1–5, doi:10.1103/PhysRevLett.107.148501, 2011.
- Heitzig, J., Marwan, N., Zou, Y., and Donges, J. F.: Consistently weighted measures for complex network topologies, *Europ. Phys. J. B*, 85, 1–16, doi:10.1140/epjb/e2011-20678-7, 2010.
- Hu, C., Henderson, G., Huang, J., Xie, S., Sun, Y., and Johnson, K.: Quantification of Holocene Asian monsoon rainfall from spatially separated cave records, *Earth Planet. Sc. Lett.*, 266, 221–232, doi:10.1016/j.epsl.2007.10.015, 2008.
- Kennett, D. J., Breitenbach, S. F. M., Aquino, V. V., Asmerom, Y., Awe, J., Baldini, J. U. L., Bartlein, P., Culleton, B. J., Ebert, C., Jazwa, C., Macri, M. J., Marwan, N., Polyak, V., Pruffer, K. M., Ridley, H. E., Sodemann, H., Winterhalder, B., and Haug, G. H.: Development and Disintegration of Maya Political Systems in Response to Climate Change, *Science*, 338, 788–791, doi:10.1126/science.1226299, 2012.
- Malik, N., Marwan, N., and Kurths, J.: Spatial structures and directionalities in Monsoonal precipitation over South Asia, *Nonlin. Processes Geophys.*, 17, 371–381, doi:10.5194/npg-17-371-2010, 2010.
- Malik, N., Bookhagen, B., Marwan, N., and Kurths, J.: Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks, *Clim. Dynam.*, 39, 971–987, doi:10.1007/s00382-011-1156-4, 2011.
- Mann, M. E. and Rutherford, S.: Climate reconstruction using Pseudoproxies, *Geophys. Res. Lett.*, 29, 1501, doi:10.1029/2001GL014554, 2002.
- Mayewski, P. A., Rohling, E. E., Curt Stager, J., Karlén, W., Maasch, K. A., David Meeker, L., Meyerson, E. A., Gasse, F., van Kreveld, S., Holmgren, K., Lee-Thorp, J., Rosqvist, G., Rack, F., Staubwasser, M., Schneider, R. R., and Steig, E. J.: Holocene climate variability, *Quaternary Res.*, 62, 243–255, doi:10.1016/j.yqres.2004.07.001, 2004.
- Molkenthin, N., Rehfeld, K., Marwan, N., and Kurths, J.: Networks from Flows – from Dynamics to Topology, *Sci. Reports*, 4, 4119, doi:10.1038/srep04119, 2014a.
- Molkenthin, N., Rehfeld, K., Stolbova, V., Tupikina, L., and Kurths, J.: On the influence of spatial sampling on climate networks, *Nonlin. Processes Geophys.*, 21, 651–657, doi:10.5194/npg-21-651-2014, 2014b.
- Opsahl, T., Agneessens, F., and Skvoretz, J.: Node centrality in weighted networks: Generalizing degree and shortest paths, *Soc. Networks*, 2010.
- Rehfeld, K. and Kurths, J.: Similarity estimators for irregular and age-uncertain time series, *Clim. Past*, 10, 107–122, doi:10.5194/cp-10-107-2014, 2014.
- Rehfeld, K., Marwan, N., Heitzig, J., and Kurths, J.: Comparison of correlation analysis techniques for irregularly sampled time series, *Nonlin. Processes Geophys.*, 18, 389–404, doi:10.5194/npg-18-389-2011, 2011.
- Rehfeld, K., Marwan, N., Breitenbach, S. F. M., and Kurths, J.: Late Holocene Asian Summer Monsoon dynamics from small but complex networks of palaeoclimate data, *Clim. Dynam.*, 41, 3–19, doi:10.1007/s00382-012-1448-3, 2013.
- Rheinwald, A., Marwan, N., Kurths, J., Werner, P., and Gerstengarbe, F.-W.: Boundary effects in network measures of spatially embedded networks, *Europhys. Lett.*, 100, 28002, doi:10.1209/0295-5075/100/28002, 2012.
- Rodbell, D., Seltzer, G., and Anderson, D.: An ~15,000-year record of El Niño-driven Alluviation in Southwestern Ecuador, *Science*, 283, 516–520, doi:10.1126/science.283.5401.516, 1999.
- Sano, M., Ramesh, R., Sheshshayee, M. S., and Sukumar, R.: Increasing aridity over the past 223 years in the Nepal Himalaya inferred from a tree-ring  $\delta^{18}\text{O}$  chronology, *Holocene*, 32, 245–251, doi:10.1177/0959683611430338, 2011.
- Schulz, M. and Stattgeger, K.: SPECTRUM: spectral analysis of unevenly spaced paleoclimatic time series, *Comput. Geosci.*, 23, 929–945, doi:10.1016/S0098-3004(97)00087-3, 1997.
- Singh, J., Yadav, R. R., and Wilmking, M.: A 694-year tree-ring based rainfall reconstruction from Himachal Pradesh, India, *Clim. Dynam.*, 33, 1149–1158, doi:10.1007/s00382-009-0528-5, 2009.
- Sinha, A., Berkelhammer, M., Stott, L., Mudelsee, M., Cheng, H., and Biswas, J.: The leading mode of Indian Summer Monsoon precipitation variability during the last millennium, *Geophys. Res. Lett.*, 38, 2–6, doi:10.1029/2011GL047713, 2011.
- Smerdon, J. E.: Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments, *Wiley Interdiscip. Rev. Clim. Chang.*, 3, 63–77, doi:10.1002/wcc.149, 2012.
- Steinhaeuser, K., Chawla, N. V., and Ganguly, A. R.: An exploration of climate data using complex networks, *ACM SIGKDD Explor. Newsl.*, 12, 25, doi:10.1145/1882471.1882476, 2010.
- Steinhilber, F., Beer, J., and Fröhlich, C.: Total solar irradiance during the Holocene, *Geophys. Res. Lett.*, 36, 1–5, doi:10.1029/2009GL040142, 2009.
- Tan, L., Cai, Y., Cheng, H., An, Z., and Edwards, R. L.: Summer monsoon precipitation variations in central China over the past 750 years derived from a high-resolution absolute-dated stalagmite, *Palaeogeogr. Palaeoclimatol.*, 280, 432–439, doi:10.1016/j.palaeo.2009.06.030, 2009.
- Tsonis, A. A., Swanson, K. L., and Roebber, P. J.: What Do Networks Have to Do with Climate?, *B. Am. Meteorol. Soc.*, 87, 585–595, doi:10.1175/BAMS-87-5-585, 2006.

- Tsonis, A. A., Wang, G., Swanson, K. L., Rodrigues, F. A., and Costa, L. D. F.: Community structure and dynamics in climate networks, *Clim. Dynam.*, 37, 933–940, doi:10.1007/s00382-010-0874-3, 2010.
- Tupikina, L., Rehfeld, K., Molkenthin, N., Stolbova, V., Marwan, N., and Kurths, J.: Characterizing the evolution of climate networks, *Nonlin. Processes Geophys.*, 21, 705–711, doi:10.5194/npg-21-705-2014, 2014.
- von Storch, H., Zorita, E., Jones, J. M., Dimitriev, Y., González-Rouco, F., and Tett, S. F. B.: Reconstructing past climate from noisy data, *Science*, 306, 679–82, doi:10.1126/science.1096109, 2004.
- Wang, Y., Liu, X., and Herzschuh, U.: Asynchronous evolution of the Indian and East Asian Summer Monsoon indicated by Holocene moisture patterns in monsoonal central Asia, *Earth-Sci. Rev.*, 103, 135–153, doi:10.1016/j.earscirev.2010.09.004, 2010.
- Wang, Y. J., Cheng, H., Edwards, R. L., An, Z. S., Wu, J. Y., Shen, C. C., and Dorale, J. A.: A high-resolution absolute-dated late Pleistocene Monsoon record from Hulu Cave, China, *Science*, 294, 2345–2348, doi:10.1126/science.1064618, 2001.
- Wu, Z., Wang, B., Li, J., and Jin, F.: An empirical seasonal prediction model of the East Asian summer monsoon using ENSO and NAO, *J. Geophys. Res.*, 114, 1–13, doi:10.1029/2009JD011733, 2009.
- Yamasaki, K., Gozolchiani, A., and Havlin, S.: Climate Networks around the Globe are Significantly Affected by El Niño, *Phys. Rev. Lett.*, 100, 228501, doi:10.1103/PhysRevLett.100.228501, 2008.
- Yi, L., Yu, H., Ge, J., Lai, Z., Xu, X., Qin, L., and Peng, S.: Reconstructions of annual summer precipitation and temperature in north-central China since 1470 AD based on drought/flood index and tree-ring records, *Climatic Change*, 110, 469–498, doi:10.1007/s10584-011-0052-6, 2011.
- Yihui, D. and Chan, J. C. L.: The East Asian summer monsoon: an overview, *Meteorol. Atmos. Phys.*, 89, 117–142, doi:10.1007/s00703-005-0125-z, 2005.