LINC

Learning about Interacting Networks in Climate

Marie Curie Initial Training Networks (ITN)

FP7- PEOPLE - 2011- ITN

Grant Agreement No. 289447

WP1: Network Construction and Analysis

Deliverable D1.5

D1.5 - Report on reconstruction and analysis of evolving networks of the climate

Jürgen Kurths, PIK

Release date: 1 November 2014

Status: Public

Executive Summary

Objectives of the WP1 - "Network construction and analysis" include the construction and analysis of the global climate networks, climate networks of physical phenomena such as Indian Summer Monsoon, El-Nino Southern Oscillations, Atlantic Meridional Overturning Circulation etc., and development of the new techniques for the climate network reconstruction and analysis.

In *D1.5 Report on reconstruction and analysis of evolving networks of the climate* we present in particular our advances in reconstruction and analysis of evolving climate networks of the Indian Summer Monsoon (ISM) – a global climate phenomenon that affects half of the world's population - on intra-seasonal and inter-annual time scales. In order to do that, we constructed evolving networks of extreme rainfall events during the ISM, the pre-monsoon, and the post-monsoon period. Through the analysis of various complex-network metrics, we uncovered typical repetitive patterns that can be used as indicators of the ISM variability. These patterns appear during the pre-monsoon season, evolve during the ISM season, and disappear during the post-monsoon season. Some of the obtained pattern were previously known by climatologists, however, we have found new geographical region that significantly affects synchronization of the extreme rainfall over the Indian subcontinent. We suggest that this region may serve as a marker of Western disturbances and shed light on the interplay between the ISM wind system and westerlies. We suggest that the obtained spatial patterns are important meteorological features that need further attention and may be useful in ISM timing and strength prediction. Using evolving climate networks we improved our understanding of the physical mechanisms that influence the ISM.

We also report newly developed technique to track the evolution of climate networks on different time scales - the common component evolution function (CCEF), which characterizes network development over time. We illustrate the performance of this technique on several model systems and real data, including Erdős–Rényi networks, analytically derived flow-based networks, transient simulations from the START model and reanalysis observational data for the Asian monsoon domain provided by NCEP/NCAR. CCEF is developed to track changes in the links of the evolving in time

networks and allows to determine episodes of the breaking down of the links. Analysis of the evolving climate networks of the Asian Monsoon domain have shown that phasing of these episodes coincides with years of strong El Niño/Southern Oscillation phenomena. The main advantage of the proposed technique is that it can be applied not only to climate networks, but for any type of evolving networks, where the set of nodes is fixed but links are changing in time. Therefore, CCEF may be particularly useful to characterize nonstationary evolving systems using complex networks.

In a related deliverable (D1.4 - Toolbox for identification of most important teleconnections and their stability) we extended the previous deliverables toolbox for the network construction and analysis (PyUnicorn) with additional codes which allow to track climate network evolution as well as identify important telecoections in climate system and their stability. This toolbox with additional codes can be found on the LINC website.

Contribution of the deliverables to the goals of LINC: WP1 develops methods for the climate networks and software that are used by LINC partners ESRs and ERs in their research in order to analyze climate data using complex network approach.

Impact on other WPs: Besides WP1, these deliverables are relevant for WP3, WP4 and WP5, since the methods and techniques presented in the report can be applied to a variety of climatic networks. This will help to detect physical processes, that influence natural climate variability from interannual to interdecadal time scales, as well as allow the tracking of transitions between different climate regimes, and provide a better understanding of the future climate change.

Target audience of the deliverable: all partners, ESRs and ERs of the LINC project.

Deliverable Identification Sheet

Grant Agreement No.	**PITN-GA-2011-289447**
Acronym	**LINC**
Full title	**Learning about Interacting Networks in Climate**
Project URL	http://climatelinc.eu/
EU Project Officer	Lucia PACILLO

Deliverable	**D1.5- Report on reconstruction and analysis of evolving networks of the climate**
Work package	**WP1**

Date of delivery	**Contractual**	M 36	**Actual**	01-Nov-2014
Status	version. 1.00		final ☑ draft ☐	
Nature	Prototype ☐ Report ☑ ☐ Database ☐			
Dissemination Level	Public ☑ Consortium ☐			

Authors (Partner)	Juergen Kurths			
Responsible Author	Juergen Kurths		**Email**	kurths@pik-potsdam.de
	Partner	PIK	**Phone**	<optional>

Abstract (for dissemination)	*In D1.5 Report on reconstruction and analysis of evolving networks of the climate* we present in particular our work in reconstruction and analysis of evolving climate networks of the Indian Summer Monsoon (ISM), a global climate phenomenon that affects half of the world's population on intra-seasonal and inter-annual time scales.
Keywords	Climate, networks, evolution, LINC

Version Log			
Issue Date	**Rev No.**	**Author**	**Change(s)**
01-11-2014	001	J. Kurths	Final report

CONTENTS

Table of Contents

1 INTRODUCTION

The application of complex network theory to analyze different climate phenomena is a new but rapidly growing area of research, where a number of studies have been carried out recently (Donges et al., 2009a, b; Malik et al., 2010, 2011; Tsonis and Roebber, 2004; Tsonis et al., 2006, 2008, 2010; Gozolchiani et al., 2008, 2011; Yamasaki et al., 2009; Paluš et al., 2011; Barreiro et al., 2011; Deza et al., 2013, 2014; Martin et al., 2013; Tirabassi and Masoller, 2013). While most of these studies are focused on global climate networks of temperature fields and precipitation (Donges et al., 2009a, b; Tsonis and Roebber, 2004; Tsonis et al., 2006; Gozolchiani et al., 2011; Yamasaki et al., 2008, 2009; Scarsoglio et al., 2013), others consider smaller, regional networks that focus on a specific climate phenomenon of interest, such as El Niño (Tsonis and Swanson, 2008; Gozolchiani et al., 2008), Rossby waves (Wang et al., 2013), continental rainfall in Germany (Rheinwalt et al., 2012), the South American Monsoon System (SAMS) (Boers et al., 2013), and the Indian Summer Monsoon (Malik et al., 2010, 2011; Rehfeld et al., 2012). In the work by Rehfeld et al. (2012) on Indian Summer Monsoon (ISM) dynamics, a paleoclimate network approach has revealed a strong influence of the ISM on the East Asian Summer monsoon during the late Holocene period, but with varying strength according to the warm vs. cold epochs. In Malik et al. (2010, 2011), it was demonstrated that by combining the climate network approach with the event synchronization method, it is possible to identify regions that receive rainfall only during the most active phase of the Indian Summer Monsoon.

In this report we summarize our results regarding study of the evolving climate networks in two major parts. In the first part we describe evolving networks of the Indian Sumer Monsoon on seasonal time scale. In the second part we present a method which allows to characterize the evolution of climate networks with time.

The first part of the report is devoted to the analysis of the evolving climate networks of the extreme rainfall synchronization: before, during and after the Indian Summer Monsoon (ISM). The Indian monsoon is a large-scale atmospheric pattern and one of the active components in the global climate system in the tropics. It is an important weather phenomenon for other parts of the globe as well, because of the monsoon's coupling with climate drivers such as the El Niño–Southern Oscillation, the Indian Ocean Dipole and the Equatorial Indian Ocean Oscillation (Sankar et al., 2011; Achuthavarier et al., 2012; Sabeerali et al., 2011; Wu and Kirtman, 2003; Wu et al., 2012; Gadgil, 2004). Understanding the structure of the spatial heterogeneity of extreme rainfall during the ISM plays a crucial role in the daily life and prosperity of the Indian population. Spatial rainfall distribution and especially the distribution of extreme rainfall events is a significant question for Indian agriculture and economy, as extreme rainfall is a common cause of floods on the Indian subcontinent. It is therefore essential to track the origins and dynamics of the extreme events. In this report, we describe such origins in the topology of networks of extreme rainfall events during the ISM, and the periods before and after the monsoon.

In the second part of the report, we introduce a method, which allows to characterize the evolution of the climate networks on different time scales - the common component evolution function (CCEF). We apply this method to several model systems, such as, Erdo˝s-Re´nyi-networks, analytically derived flow-based networks, and transient simulations from the START model for which we control the change of single parameters over time. Then we construct annual climate networks from NCEP/NCAR

reanalysis data for the Asian monsoon domain for the time period of 1970 - 2011 C.E. and use the CCEF to characterize the temporal evolution in this region.

2 DATA

2.1 Indian Summer Monsoon: extreme rainfall synchronization network evolution

We use observational satellite data from 1998 to 2012 from the Tropical Rainfall Measuring Mission (TRMM 3B42V7) (Huffman et al., 2007; TRMM, 2014), with a spatial resolution of 0.25^0 - 25 km and a temporal resolution of 3 h aggregated to daily data. We have extracted data for the South Asian region (62.375–97.125^0 E, 5.125–39.875^0 N; see Table 1 and Fig. 1). This data set is the most recent precipitation data product available. It has a high spatial resolution and covers both land and sea. By the time of this study, however, there are only 15 years of data. To confirm the significance of the results of this study, we analyze reanalysis gridded daily rainfall data for a time period of 57 years (1951–2007) (Asian Precipitation Highly Resolved Observational Data Integration Towards the Evaluation of Water Resources, APHRODITE) (Yatagai et al., 2009). This data set forms a reliable reanalysis product; however, it is given only for the land masses, excluding the ocean. We have extracted the data for the same region (62.5 - 97.5^0 E, 5–40^0 N) with a resolution of 0.5^0 - 56 km (APHRO-V1003R1) (see Table 1). This data set was used previously in Malik et al. (2010, 2011) to study the spatial variability of the extreme Indian rainfall during the ISM period using the method

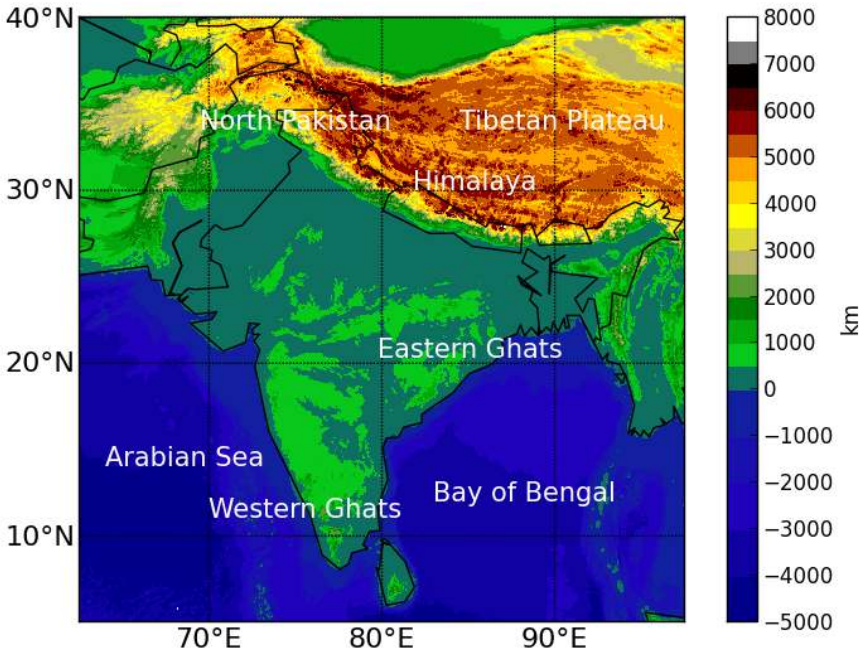

Figure 1. Regional overview of the Indian subcontinent. Major topographic and political features referred to in the text are labeled in white, and the background image is based on the Shuttle Radar Topography Mission (SRTM30) gridded digital elevation model, which is available at GEBCO (2014).

Table 1. Properties of the data sets. APHRODITE: interpolated rain-gauge data (Yatagai et al., 2009). TRMM 3B42V7: satellite-derived data (Huffman et al., 2007). NCEP/NCAR: reanalysis data (Kalnay et al., 1996).

Properties	APHRODITE	TRMM	NCEP/NCAR
Period	Jan 1951–Dec 2007	Jan 1998–Dec 2012	Jan 1949–Dec 2012
Geographical coverage	(62.5–97.5° E, 5–40° N)	(62.375–97.125° E, 5.125–39.875° N)	(62.5–97.5° E, 5–40° N)
Spatial resolution	0.5° × 0.5°	0.25° × 0.25°	2.5° × 2.5°
Temporal resolution	Daily precipitation	3 hourly, resampled to daily precipitation	Daily temperature and pressure anomalies, seasonal mean of winds
Number of grid points	4900	19 600	196

of Event Synchronization (ES). In this study, we compare results with previous findings for the ISM season and determine geographic patterns of extreme precipitation over the Indian subcontinent during the pre-monsoon, monsoon, and post-monsoon seasons. We compare geographic patterns based on precipitation data sets with patterns of temperature and pressure networks derived from reanalysis gridded daily data provided by the National Center for Environmental Prediction and the National Center for Atmospheric Research (NCEP/NCAR) (Kalnay et al., 1996; NCEP/NCAR, 2014). In addition, we compare our results with wind fields using reanalysis gridded seasonal mean wind data, also provided by NCEP/NCAR (Kalnay et al., 1996; NCEP/NCAR, 2014). The spatial resolution of the data is 2.5^0. The data are extracted for (roughly) the same region as for the APHRODITE data (see Table 1), and contain both land and ocean points in the considered region.

2.2 Method for characterization of the evolution of climate networks

2.2.1 Testing the method on random networks

To test our method we generate a set of T Erdos-Renyi graphs (Erdos and Renyi, 1959) with a fixed number of n nodes and a fixed connection probability p. We artificially impose a linear ordering on the set, such that we can index them with $i \quad (1, T)$. We compute the CCEF for Erdos - Renyi-graphs with 100 nodes and link probabilities of 0.3, 0.5 and 0.9. The resulting functions, shown in Fig. 1, decrease from 1 to a plateau at $CCEF(\ ) \quad p \ for \quad > 0$ for each link probability p. For this example we can analytically compute the expected *CCEF*, since for $= 0$ each network is compared with itself and therefore *CCEF* = 1. For all other values of two random matrices with n nodes and connection probability p are compared. Then the number of totally possible links is $n(n \ \ 1)/2$, and the expectation value of the number of links in each of the networks is $pn(n \ \ 1)/2$. As the probability of each of the edges in one network to also appear in the other network is p, the total number of common links is $p^2n(n \ \ 1)/2$, which with the normalization leads to *f(p)*, the ratio of total number of common links and the expectation value of the number of links to be $f(p) = \frac{p^2 n(n \ \ 1)}{pn(n \ \ 1)} = p$. The *CCEF* for each linking probability therefore lies close to the expected value p.

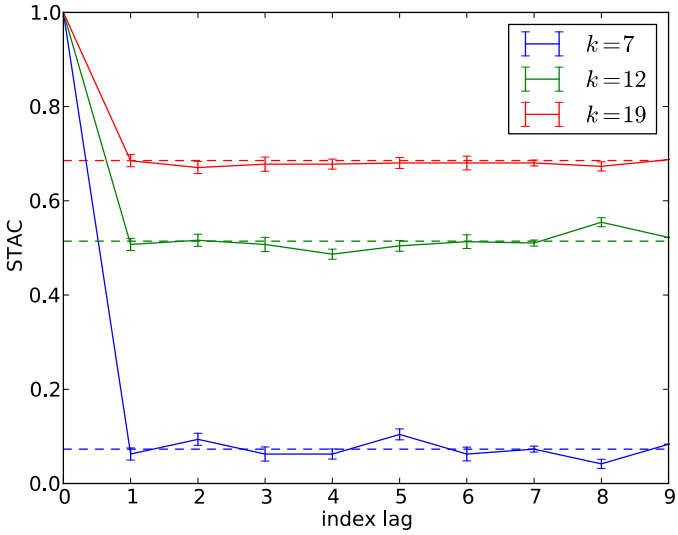

Figure 2. Mean CCEF for 10 indexed sets of random Erd˝os-Renyi networks with 100 nodes and different linking probabilities *p*. The dashed lines correspond to the analytical CCEF levels, the error bars give the 1 standard deviation for each index lag.

2.2.2 Test models

To characterize our approach further we investigate simulations of more complex, spatially embedded processes. We obtained networks (i) analytically from flow fields, as described in (Molkenthin et al., 2014a) and (ii) from the Spatio-Temporally Autocorrelated Time series model START.

Networks from flows

The flow networks are constructed directly from a velocity field using a correlation measure based on the temperature profiles resulting from a temperature peak via advection and diffusion (Molkenthin et al., 2014b,a). The velocity function considered here is:

$$v(x,y) = \frac{e^{(y\ 0.5x)^2/c}}{0.5e^{(y\ 0.5x)^2/c}} \qquad (1)$$

and we vary the parameter c for the flow-width from 200 to 2000 in 10 steps, thus gradually changing the flow network. The positions and the number of nodes are kept constant. For each value of c we obtain a correlation matrices C_1, …, C_{10}, and thresholding these matrices by different critical values we obtain set of adjacency matrices.

Networks from the START model

As a more complex test case we consider two transient simulations for the STARTmodel (Rehfeld et al., 2014). Networks generated from START undergo a

distinct transition when the forcing parameter F is changed: For $F = 1$ the network is partitioned into two vertical connected areas. For $F = 0$ horizontal cross-links have appeared and link the two sections. At maximal forcing, for $F = 1$, there is one large, horizontally oriented component. We performed two transient simulations with a 6*7 sampling grid (Rehfeld et al., 2014) for 20000 time steps and 100 ensemble members each. In the first run the forcing parameter was increased linearly from the start to the end of the simulation. In the second run we periodically changed the forcing parameter $F(t) = sin(t=2000)$. Networks were constructed based on the 20% strongest links in the correlation matrices obtained for each 100-step-long time window. Due to the stochastic component, networks constructed for different ensemble members, but for the same time period may be quite different, networks for different periods of same ensemble member may be quite similar.

Asian monsoon data

In a real-world application we used daily NCEP/NCAR reanalysis temperature anomaly data (NOAA) for the Asian monsoon domain for the years 1970-2010 C.E. The spatial resolution was 2.5^0- 2.5^0, covering the area between 2.5^0 S to 42.5^0 N and 57.5^0 E to 122.5^0 E, resulting in time series for 468 nodes. Networks were constructed using Pearson correlation in windows for each full year and by thresholding the correlation matrix such that we obtain a link density of 5%. The same dataset and time period was used in Molkenthin et al. (2014b) to investigate the influence of changing node topologies in space on the estimates of node degree and betweenness.

3 RESULTS

3.1 Indian Summer Monsoon: extreme rainfall synchronization network evolution

3.1.1 Network construction

We consider separately time series of rainfall events, temperature and pressure fields for three time periods: (i) pre-monsoon, (ii) ISM, and (iii) post-monsoon. The pre-monsoon period (March, April, May – MAM) is defined as the period from 1 March

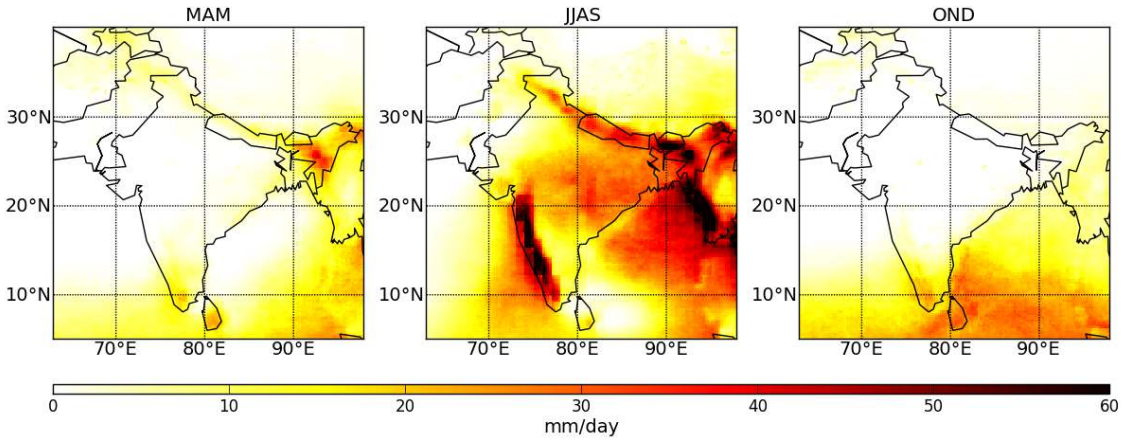

Figure 3. 90th percentile of daily rainfall amounts for the pre-monsoon (MAM), monsoon (JJAS), and post-monsoon (OND) periods for the APHRODITE (top) and

TRMM (bottom) data sets using the same color scale.
to 31 May for each year, and consists of 92 days per year. This number multiplied by the number of years in the data set gives the length of the time series for each node in the MAM period. The monsoon period (JJAS) is defined as the period from 1 June to 30 September for each year, and consists of 122 days per year. The post-monsoon period (OND) is defined as the period from 1 October to 31 December for each year, and consists of 92 days per year. The time series at each node for the JJAS and OND periods are computed in the same manner as for the MAM period.

Considering rainfall time series, it is important to choose an appropriate method to infer similarity of dynamics between different geographical sites. Rainfall time series are not as smooth and continuous as those for temperature or pressure fields, but often contain a high-frequency component. In this study, we use ES as a method for climate network construction from precipitation data, as proposed in Malik et al. (2010). This method has advantages over other time-delayed correlation techniques (e.g., Pearson lag correlation), specifically for studying precipitation data, as it allows us to define extreme event series of rainfall, depending on the kind of extreme, and uses a dynamic time delay. The latter refers to a time delay that is adjusted according to the two time series being compared, which allows for better adaptability to the region of interest. Another advantage of this method is that it can also be applied to a non-Gaussian and event-like data sets. We begin by extracting an event series, which is merely a time series that includes only the highest daily rainfall amounts, which we call "extreme events". In the realm of hydrology and climate sciences, extreme rainfall events are days that receive rainfall amounts that exceed the 90th percentile for all days in the time series at a given grid point (see Fig. 3). This threshold thus gives a unique amount of rainfall per day for each grid site. Only the daily rainfall events that are above this threshold for a given site are considered to be extreme rainfall events, and are what make up an event series.

Let us consider two grid sites i and j. An event l that occurs at grid site i at time t_l^i is called synchronized with an event m that occurs at grid site j at time t_m^j within a time lag $\pm\ {}^{ij}_{lm}$ (eq.1) if $0 < t_l^i \quad t_m^j < {}^{ij}_{lm}$. Here, l = 1,2,…, s_i, m = 1,2,…, s_j, where s_i, s_j - number of events in ith and jth grid sites. Then, for each grid site we count the number of times $c\left(i|j\right)$ when an event occurs at i after j, and vice versa $c\left(j|i\right)$ (eq.2). Here, J_{ij} is an event that happens at place j after place i within the time lag $\pm\ {}^{ij}_{lm}$(eq.3).

$$ {}^{ij}_{lm} = \min\left\{ t^i_{l+1} \quad t^i_l, t^i_l \quad t^{ij}_{l\ 1}, t^j_{m+1} \quad t^j_m, t^j_m \quad t^j_{m\ 1} \right\} / 2 \qquad (2)$$

$$ c\left(i|j\right) = \sum_{l=1}^{s_i} \sum_{m=1}^{s_j} J_{ij} \qquad (3)$$

$$ J_{ij} = \begin{cases} 1, & if \quad 0 < t_l^i - t_m^j < \tau_{lm}^{ij} \\ 1/2 & if \quad t_l^i = t_m^j \\ 0 & else. \end{cases} \qquad (4)$$

Table 2. Network measures. N – total number of nodes. D_j – degree of a node j. B_v – betweenness of a node v. L_{ij} – geographical link length between nodes i and j. ALL_i – average geographical link length of a node i. MLL_i – maximal geographical link length of a node i. $\sigma_v(i,j)$ – the number of shortest paths between nodes i and j passing through node v. $\sigma(i,j)$ – total number of shortest paths between i and j. α_{ij} – angular geographical distance. R – radius of the Earth, 6371.009.

Degree, D_i	Betweenness, B_v	Average geographical link length, ALL_i	Maximal geographical link length, MLL_i
$D_i = \frac{\sum_{n=1}^{N} A_{ij}}{N-1}$	$B_v = \sum_{i \neq j \neq v \in \{V\}}^{N} \frac{\sigma_v(i,j)}{\sigma(i,j)}$	$\mathrm{ALL}_i = < L_{ij} >_j = < \alpha_{ij} A_{ij} R >_j$	$\mathrm{MLL}_i = \max(L_{ij})_j = \max(\alpha_{ij} A_{ij} R)_j$

$$Q_{ij} = \frac{c(i|j) + c(j|i)}{\sqrt{(s_i - 2)(s_j - 2)}} \qquad (5)$$

Then, we define the strength of synchronization, Q_{ij}, between events at different grid sites i and j (eq. 5) and normalize it to be $0 < Q_{ij} < 1$. Here, $Q_{ij} = 1$ means complete synchronization, and $Q_{ij} = 0$ is the absence of synchronization.

After repeating this procedure for all pairs $(i \neq j)$ of grid sites, we obtain a square, symmetric correlation matrix, which represents the strength of synchronization of the extreme rainfall events between each pair of grid sites. Applying a threshold or certain link density, we yield an adjacency matrix A_{ij} (eq. 6).

$$A_{ij} = \begin{cases} 1, & if \;\; Q_{ij} > \frac{Q}{ij} \\ 0 & else. \end{cases} \qquad (6).$$

To include only statistically significant correlations, we threshold adjacency matrix in such a way so to choose the 5% strongest correlations and construct climate networks. The adjacency matrix represents a climate network, and complex network theory can more subsequently be employed to reveal properties of the given network such as degree, betweenness and the maximal geographical distance of links (see Table 2).

2.3 Results: evolving climate networks of the extreme rainfall

First, we determine general features that correspond to the chosen time period, the so-called "dominant" patterns of the entire time period, by comparing climate networks for the pre-monsoon, ISM and post-monsoon periods using the TRMM data. Figure 4 shows the degree, betweenness and the average geographical link length of the networks for the TRMM data (see Tables 1 and 2):

- Degree: during the pre-monsoon period, there are five regions with a high degree: (i)Western Ghats (WG) and the Arabian Sea, (ii) North Pakistan (NP), (iii) the Himalaya, (iv) Eastern Ghats (EG) and the Bay of Bengal, and (v) the Tibetan Plateau (TP). The regions (i) and (iii) have an especially high degree during the pre-monsoon season. During the monsoon season, the highest degree is in NP, EG and TP, while during the post-monsoon season, it is in the Himalaya, TP and NP.

- Betweenness: high values are observed for the pre-monsoon, monsoon and post-monsoon seasons for the same regions, as specified above with a high degree.

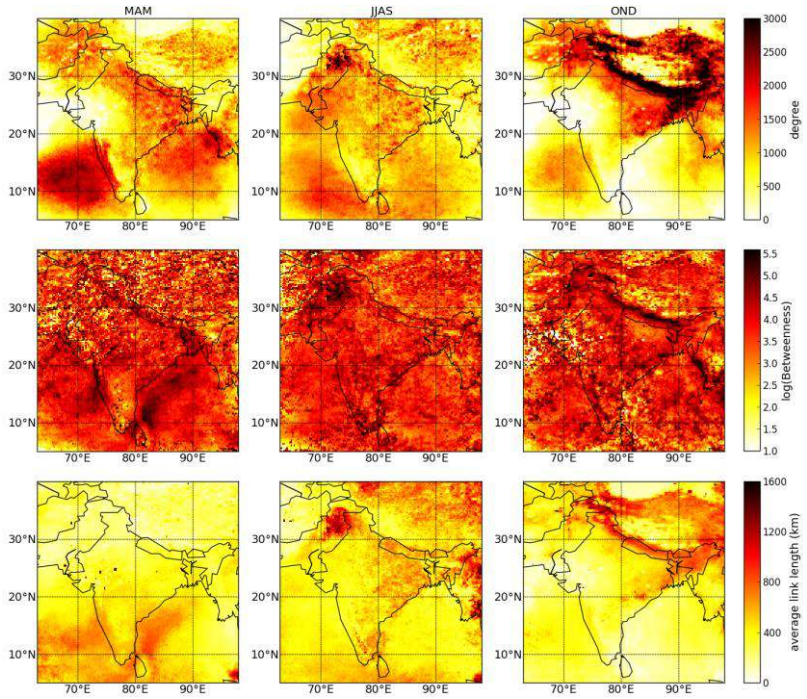

Figure 4. Common network measurements for the three time periods based on the TRMM data: pre-monsoon (MAM), monsoon (JJAS), and post-monsoon (OND). From top to bottom: degree, betweenness, average geographical link lengths.

– Average and maximal geographical link lengths: comparatively long links can be found in the Arabian Sea and the Bay of Bengal during the pre-monsoon period; in NP, TP, EG and the Himalaya during the monsoon season; and in TP and the Himalaya during the post-monsoon period.

Employing the network measures, we look into the role of hubs (super nodes) in the organization of the spatial structure of the pre-monsoon, monsoon, and post-monsoon periods. We define the hubs of a network as the 10% of grid points with the highest degree, betweenness, and average link length. We infer seven main patterns that have either a high degree, betweenness, or average link length during at least one of the three time periods (see Table 4). Three of these patterns are believed to be dominant during the ISM. First, the NP region shows a high degree, betweenness, and average link length during the monsoon season, indicating a large number of long connections in this region during this time. The high betweenness of the NP region shows its importance in the synchronization of extreme rainfall events over the Indian subcontinent specifically during the monsoon period. The EG and TP regions also have a large number of long connections during the monsoon season, with TP showing a similarly large number in the post-monsoon season as well.

These regions, therefore, play an important role in the organization of the structure of extreme rainfall event synchronization, primarily during the monsoon season. Using a similar analysis the monsoon season. Using a similar analysis for the other time periods, we uncover the main patterns in the networks of extreme rainfall event synchronization for each season: (i) pre-monsoon season: WG, the Himalaya – over the land, and the Arabian Sea and the Bay of Bengal; (ii) monsoon: NP, EG and TP; (iii) post-monsoon: the Himalaya and TP.

Table 4. Hubs of the extreme rainfall networks. Columns are hubs in degree, betweenness, and average link length. "+" – the network measure in the given region is in the top 10 % for the given period. "–" – lower than the top 10 %. The order of "+" or "–" means the season: pre-monsoon, monsoon, and post-monsoon, respectively. For example, the intersection of first row and first column (North Pakistan (NP) and degree) – "+/+/+" – means that one can observe a high degree in NP during the pre-monsoon, monsoon, and post-monsoon seasons, respectively.

	Degree, D_i	Betweenness, B_v	Average geographical link length, ALL_i
North Pakistan, NP	+/+/+	−/+/−	−/+/−
Tibetan Plateau, TP	+/+/+	−/−/−	−/+/+
Eastern Ghats, EG	+/+/+	−/+/−	−/+/−
Himalaya	+/−/+	+/−/+	−/+/+
Western Ghats, WG	+/−/−	+/−/−	+/−/−
Arabian Sea	+/−/−	+/−/−	+/−/−
Bay of Bengal	+/−/−	+/−/−	+/−/−

Second, we present a visualization of the links of the dominant patterns of ISM during the three seasons, and analyze the seasonal evolution of these patterns (Fig. 5). In order to understand which atmospheric processes cause synchronization of extreme rainfall events, we compare the visualization of links of these patterns with wind fields.

Links stemming from the NP pattern show local synchronization of extreme rainfall events during the pre-monsoon season. This phenomenon is not surprising, given the mountainous terrain of the region. During the ISM, however, long range connections appear; extreme rainfall events in the NP region synchronize with events in TP, the Himalaya, EG and Burma (Fig. 5). These clusters are of great interest for further research, as they appear only during the ISM season, and are possibly linked to monsoonal trends in wind strength. The post-monsoon season sees a decrease in the number of connections to points in the NP region, but maintains connections to TP and the Himalaya.

The TP pattern also shows local synchronization of extreme rainfall events during the pre-monsoon season. Once again, this coincides with the topography of the region that prevents rainfall from synchronizing with other regions. However, the TP has many links to both the NP and EG regions during the ISM. These connections are long range and, possibly, caused by large-scale atmospheric processes such as ISM winds and Western Disturbances. During the post-monsoon season, the connections are more locally clustered within the TP, and there are also a lot of connections between TP and the Himalaya.

The EG pattern shows the most intricate behavior of extreme rainfall event synchronization: the shape of the pattern varies noticeably from season to season. During the premonsoon season, extreme rainfall events are synchronized within the EG region, and with the WG region and the Arabian Sea. During the ISM, there are three

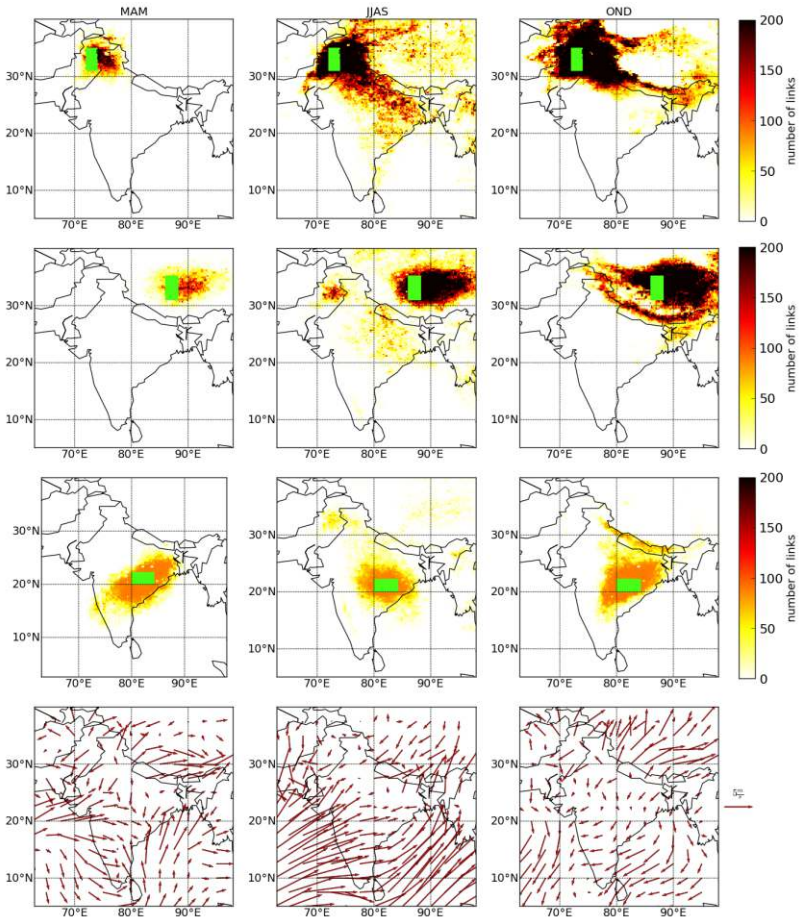

Figure 5. Links between a set of 153 reference grid points and other grid points, and the surface wind vector mean between 1998 and 2012. From top to bottom: North Pakistan (NP), Tibetan Plateau (TP), Eastern Ghats (EG) (TRMM) and the surface wind vector mean (NCEP/NCAR).

main clusters of synchronized extreme rainfall that are connected to the EG region: NP, TP and WG. During the post-monsoon season, the synchronization resembles that of the pre-monsoon season (see Fig. 5), with additional connections to the Himalaya. This feature has a climatological interpretation, since the Himalaya receive heavy rainfall during the post-monsoon season.

Third, we compare the results of the network analysis for the APHRODITE and TRMM data sets. The obtained results show that dominant patterns of the ISM are observed in both data sets.

Finally, we build networks from the temperature and pressure fields over the Indian peninsula for three seasons, and compare the structure of these networks with dominant patterns of the ISM extreme rainfall. We observed several common features in the networksof extreme precipitation, temperature and pressure fields during the three seasons, as well as unique features for each season. Common features of the networks of extreme precipitation are a high degree and betweenness over NP and EG during the ISM, and a high average link length in NP, EG, TP and the Himalaya also during the ISM, confirming their role in ISM dynamics. In particular, we find that the Himalaya play more than just the role of an orographic barrier, blocking ISM winds from blowing

rainfall into China, Afghanistan, and Russia. For the temperature and pressure networks, the common features are a high degree over both NP and TP (see Table 4).

Climatological interpretation of the results

All of these patterns coincide with the topography of the Indian peninsula, and four of them – EG, WG, TP, and the Himalaya – play a governing role in the onset and spatial organization of the ISM extreme rainfall (Wang, 2006; Pai and Nair, 2009; Malik et al., 2010, 2011). However, we show that the NP pattern also strongly influences the network of extreme rainfall over the Indian subcontinent, and should be taken into account when studying the ISM dynamics.

The EG, WG, TP and Himalaya patterns are known as areas that influence the ISM dynamics, mostly because of the intricate topography of these regions that forces orographic lifting and high rainfall amounts. The Western Ghats are the first highlands of the Indian subcontinent encountered by the ISM winds. The mountains rise abruptly from the western coastal plains of the subcontinent, creating an orographic barrier for the monsoonal winds. The WG, together with the Himalaya, is the main cause of the substantial orographic precipitation all across the Indian subcontinent during the ISM season (Bookhagen and Burbank, 2010). During the winter season, however, it is the Eastern Ghats that create an orographic barrier for the rainfall during the post-monsoon season. The Tibetan Plateau is another important region during the winter months, as precipitation in the region is caused by winter westerlies. It was shown that persistent warming on the Tibetan Plateau for the last three centuries coincides with intensification of the ISM, and the cool/warm epochs in TP also coincide with a weak/strong ISM (Feng, 2005). Our results from temperature, pressure, and extreme precipitation networks confirm the important role of the above mentioned regions in the ISM dynamics.

The formation of the NP pattern during the ISM is caused by the topography of the region that, together with the ITCZ and the Arabian Sea branch of the ISM, promotes the formation of wind convergence in this region. This can explain why the NP region is characterized by both high degree and high betweenness: the convergence of surface winds is caused by the low pressure in the NP area. The high degree, betweenness, and average link length that we have observed in NP supports the theory that it is not the differential heating of the land and sea that is the governing mechanism of the ISM, but the monsoonal winds along with the Himalaya that act as a high shield that stops Siberian dry and cold air from penetrating the Indian subcontinent (Webster, 1998; Chakraborty, 2002; Sinha et al., 2013; Boos and Kuang, 2010). Additionally, rainfall in the NP region is mostly caused not by the monsoonal rainfall, but by Western Disturbances, which transport moisture from the Caspian Sea to the low-pressure area of the NP. In this study, we have shown that the NP region influences extreme rainfall synchronization across a large area of the Indian subcontinent. We suggest that since this region reflects Western Disturbances and influences extreme rainfall over most of the Indian subcontinent, it can serve as an indicator region for the interaction between the ISM system and Western Disturbances. Therefore, it might deserve careful consideration as one of the key regions, along with the EG (because of their connection), for the analysis of the ISM interaction with Western Disturbances, as well as for the analysis of floods in Pakistan.

3.2 Method for characterization of the evolution of climate networks

The spatial-temporal developments in a given network set can be too complex to be captured by eye, and systematic approaches to quantify changes are needed. While Berezin et al. (2012) investigated the origins of the climate network stability such as the spatial embedding and physical coupling between climate in different locations using the correlation between correlation matrices, other studies describe how the network graph is changing over time to understand the behaviour of the underlying dynamical system (e.g. Rehfeld et al., 2013).

One of the most common dissimilarity measures which has been used for network comparison is Hamming distance. It was introduced by Hamming (1950) as a measure for comparing strings of symbols and was used for measuring the distance between the networks. Given the adjacency matrices A_N and A_M of two graphs N and M, their Hamming distance is determined from the sum over the number of links which are found in one, but not the other network: $H(N,M) = \sum_{i,j} \left| A_{ij}^N \quad A_{ij}^M \right|$.
However, although Hamming distance can be generalized for directed networks with possibly differing node numbers, two networks M and N may have the same Hamming distance to the fixed network K while having different topology themselves. The Hamming distance therefore may not enough to detect topological changes.

Here, we propose a common component evolution function (CCEF) based on the common set of links in pairs of networks to evaluate graph changes quantitatively in space and time. We characterize the method using Erdos–Rényi networks (Erdös and Rényi, 1959), analytically derived flow networks (Molkenthin et al., 2014a, b) and transient simulations from the START model (Rehfeld et al., 2014) for which we control changes of individual parameters over time. Then we construct annual climate networks from NCEP/NCAR reanalysis data for the Asian monsoon domain and use the CCEF to characterize the temporal evolution in the monsoon system.

3.2.1 Derivation of the common component evolution function

We consider unweighted and undirected networks, for which n nodes are joined in pairs by edges, or links. The linking structure is given in the adjacency matrix A, a binary $n \quad n$ matrix with zeros on the diagonal, as we do not allow for self-loops. An element is non-zero, $A_{ij} = 1$, if and only if the vertices i and j are connected, and zero otherwise.
Let us consider a linearly ordered set of T evolving in time networks: $N_1,...,N_T$. Then the common component network for two of these networks N_i and N_j, $CC(N_i, N_j)$, is a network on the same nodes, where the set of edges is present in both original networks. If N_i and N_j have adjacency matrices A_i and A_j, the number of edges in the common component network $CC(N_i, N_j)$ is the number of nonzero elements above the diagonal in the binary sum of adjacency matrices A_i and A_j. This common component network can be generalized for any $k+1$ networks by induction: $CC(N_i,...,N_{i+k+1}) = CC(CC(N_i,...,N_{i+k+1}), N_{i+k})$. The common component function $CCF(N_i,...,N_{i+k})$ counts the number of links in a common component network of k

networks: $CCF(N_i,...,N_{i+k}) = \|CC(CC(N_i,...,N_{i+k-1}), N_{i+k})\|$, where by $\|N_i\|$ we mean the number of links in the network N_i, and the common component function $CC(N_i, N_k)$ gives the number of coinciding edges in the graphs of N_i and N_k, $i,k \in \{1,2,...,T\}$. We set $CCF(N_i) = CC(N_i, N_i)$. The common component function $CCF(N_i)$ takes values in [0, max CCF(Ni)] and is in the following normalized to [0, 1] using the maximal number of links in the networks.

In analogy to covariance estimation (Chatfield, 2004) and similar to Berezin et al. (2012), we take the mean over the CCFs with the same time lags to estimate the non-normalized common component evolution function , $CCEF^*$, as

$$CCEF^*(\ ) = \frac{1}{T} \sum_{i=1}^{T} CCF(N_i, N_{i+\ }), \tag{7}$$

where is the time lag between the networks, and $[0, T\ \ 1]$. The maximum value of the $CCEF^*$ is given by for zero lag, as an average number of links in the set of network

$$CCEF^*(0) = \frac{1}{T} \sum_{i=1}^{T} CCF(N_i), \tag{8}$$

and we use it to obtain the normalized *common component evolution function*

$$CCEF(\ ) = \frac{CCEF^*(\ )}{CCEF^*(0)}, \tag{9}$$

which we will use exclusively in the following. As an estimation of the CCEF uncertainty we use the standard deviation over all CCEF values.

3.3 Applications of the common component evolution function

3.3.1 Flow networks

We computed the CCEF for flow networks with linearly increasing flow-width parameter c. As Fig. 6 shows, the common component size decreases monotonically with the width parameter difference of the networks. The higher the threshold of the correlation matrix is, the faster the CCEF decays but the general shape does not change.

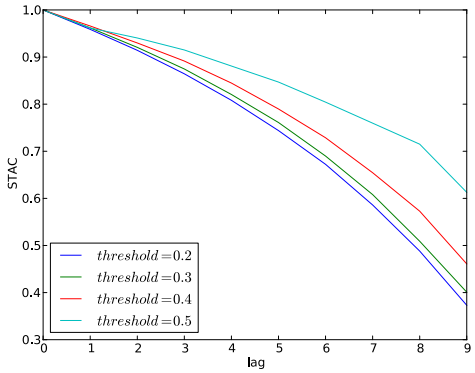

Figure 6. CCEF for the indexed flow networks with increasing flow-width parameter c, and for different threshold values of adjacency matrixes of the flow networks.

3.3.2 START-model networks

The START model undergoes a more distinct transition from a network with two distinct parts through a connected stage with three regions to one single component (Rehfeld et al., 2014) in response to a single forcing parameter F. To characterize the $CCEF$ response to different network evolution patterns we use two test cases, in which we vary F from its minimum to its maximum. In the first example, the forcing parameter is varied linearly along time. The $CCEF$ response is a slow decline from its maximum $CCEF(0)=1$ to a minimum value $CCEF(99)=0.4$, as shown in Fig. 7. In the second test the forcing parameter was varied periodically as a function of time, $F=\sin(\frac{2}{P}t)$, with $P=10$. In response to the sinusoidal forcing, periodic behavior is also observable in the $CCEF$ and with the same period as the forcing parameter.

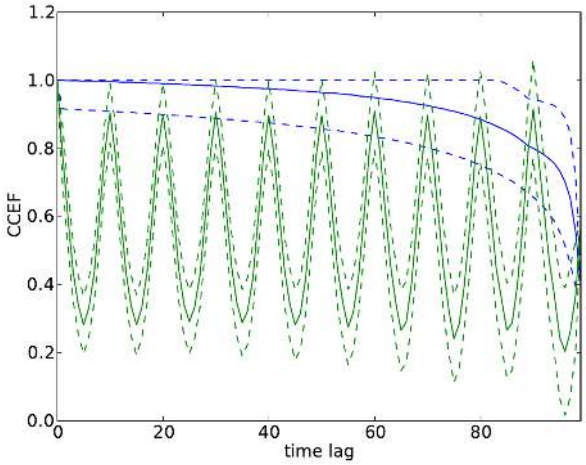

Figure 7. CCEF for networks from the START model with periodic (green) and linearly increasing (blue) forcing parameter F. The error bars indicate the standard deviation of the CC size estimates.

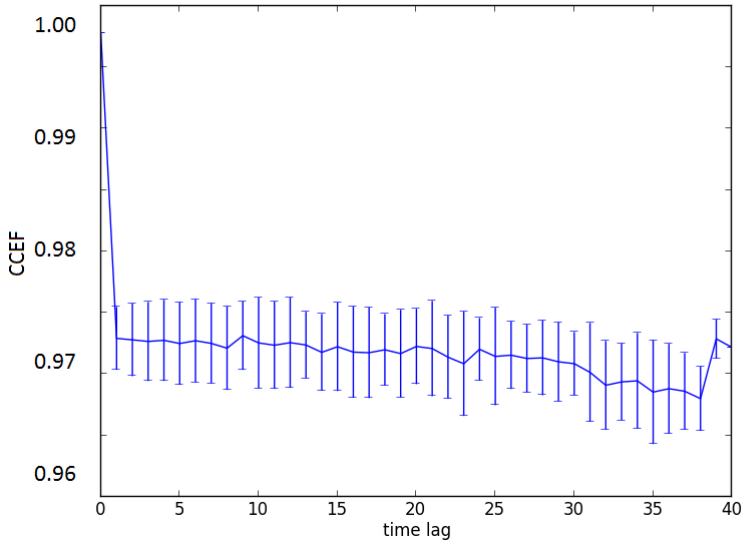

Figure 8. CCEF of annual climate networks for the time period 1970–2011 CE. Error bars are presented as CCEF standard deviation of the respective time lag in years.

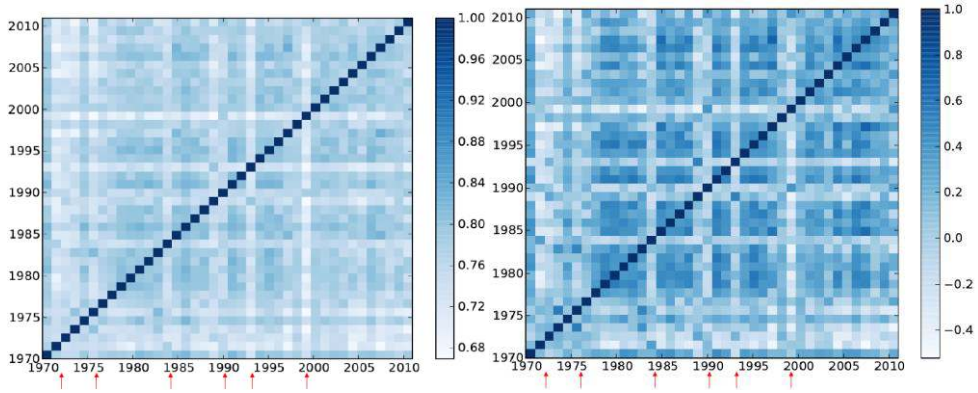

Figure 9. Common-links-recurrence-diagram (a) and correlation matrix of the common link evolution (b). Each point (*i*,*j*) in this diagram corresponds to the value of the correlation coefficient *corr*(*i*,*j*) between the common component functions $CCF(N_i, N_k)$ and $CCF(N_j, N_k)$. Lines with low values (marked at the bottom with arrows) are observable around strong ENSO years.

3.3.3 Application to the Asian Monsoon domain

Finally, we used the *CCEF* to investigate the evolution of climate networks from observations. The networks were constructed using a link density of an annual basis for 41 years, 1970-2011 C.E. The obtained *CCEF* in Fig. 8 is reminiscent of the Erdos-Renyi networks described above, with an initial quick decline followed by a plateau.

However, while in the ER-case (Fig. 6) the baseline is equal to the set link density, it is significantly higher than the link density here. We thus conclude that a high degree of

persistence and a low amount of spatio-temporal variance can be found in climate networks from the Asian Monsoon domain at annual time-scale.

The spatial domain selected for this study is the host of very distinct seasonal dynamics during the Asian Monsoon seasons (Wang et. al., 2006). At inter-annual timescales, however, teleconnections such as that to the El-Nino-Southern Oscillation (ENSO) phenomenon plays a significant role (Turner and Annamalai, 2012, Clarke, 2008). In order to identify reasons for the variability of climate networks in this region we compared the variation of common component functions $CCF(N_i, N_j)$ for different $i, j \quad \{1970, 2011\}$, where i and is kept fixed. This way, we obtained a common-links-recurrence-diagram, illustrated in Fig. 9 a, with maximum values on the diagonal. Each pair *(i,j)* for $i, j \quad \{1970, 2011\}$ corresponds to the value of the common component function $CCF(N_i, N_j)$, as in Eq. 9.

Fig. 8 a shows rows and columns with distinct lower values for the CCF. In these lines the overall sum, $S_i = \sum_j CCF(N_i, N_j)$, takes smaller values in the years $i \quad \{1971\text{-}1973, 1975, 1984, 1989, 1993, 1999\}$. We compared this sequence with a list with strong El-Nino phenomena according to the El-Nino 3.4 index (Trenberth, 1997), and observed that 1972, 1982, 1988, 1992 and 1997 were the strongest ENSO event years in this time period. At the same time, the correlation between the *CCF* functions of these years and all others, given in Fig. 9 b , also takes on very low values. Around stronger El-Nino years the surface temperature networks have less common links, and the correlation of their CCF with all others is considerably lower.

3.3.4 Climatological interpretation of the results

The CCEF enables us to investigate the evolution of linearly ordered, or evolving, network sets quantitatively. We tested its response to three different types of model networks and find that the responses enable us to characterize their evolution.

The year-long daily temperature anomaly networks of the Asian Monsoon domain show a high degree of spatiotemporal persistence. This is consistent with the results of Berezin et al. (2012), who found similarly high values over large regions of the South Atlantic and the Equatorial Pacific. This points towards a highly non-random, deterministic general structure in the network on which the inter-annual variability is imprinted. Links in this network are comparatively stable but loose some of their stability when the external disturbance of an El-Nino-event is added. This agrees well with the findings of Gozolchiani et al. (2008) and Tsonis and Swanson (2008), who showed that, for global networks, fluctuations, or “blinking” of links could be related to the global signature of ENSO variability. Therefore, despite the large persistence in the monsoon network, the monsoon-ENSO teleconnection is also visible in the common link recurrence. To check whether the main changes in the climate networks investigated here occurred due to changes in the degrees of ”supernodes” (nodes with higher degree), we investigated the variability of the degree for each node and find that, indeed, degree variability is high (low) where node degree is high (low). Computing the correlation between time series of node degrees, we obtain a network of degree variability. The degree variability is low, over mainland India, we find high degrees in

the degree network, which means that links in this region are mostly persistent. Where the degree variability is high, over the adjacent Indian Ocean and the South China Sea, we also observe high degree values in the network of the degrees, suggesting that degree changes here are large, but synchronized. In the northern part of the Asian monsoon domain considered here, spanning from Afghanistan through Pakistan, the Himalayas to China, we find a higher degree variability with less synchronized degree changes. Desynchonization in this region may occur due to the additional effect of the continental Westerlies and the large altitudinal gradients.

4 SUMMARY

In the first part of the report, we presented our results on reconstruction and analysis of evolving climate networks of the Indian Summer Monsoon. The intraseasonal, interannual variability, and the spatial distribution of the ISM have a tremendous socio-economic impact on subcontinental India, especially in the fields of agriculture and health. We used complex networks to identify dominant spatial patterns that govern the seasonal evolution of the organization of extreme rainfall over the Indian Monsoon domain. Through the analysis of various complex-network metrics, we describe typical repetitive patterns that can be used as indicators of the ISM variability: North Pakistan (NP), Western Ghats (WG), Eastern Ghats (EG), and Tibetan Plateau (TP). These patterns appear during the pre-monsoon season, evolve during the ISM season, and disappear during the post-monsoon season. We compared the obtained results with wind fields, temperature, and pressure networks in this region derived from re-analysis data provided by the National Center for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR). The areas of Eastern Ghats, Western Ghats, and Tibetan Plateau were previously known as areas that influence the ISM dynamics. The Western Ghats, specifically, the Kerala region, is commonly used by climatologists for the prediction of the onset of the ISM. However, North Pakistan has not been previously considered as an important region for the analysis of the ISM variability. In our study, we have shown that this region plays an important role in the extreme rainfall organization during the ISM. Taking into account that North Pakistan is strongly influenced by Western Disturbances during the pre-monsoon season and the monsoon season, which strongly affects the ISM variability, we suggest that this region may serve as a marker of Western disturbances and shed the light on the interplay between the ISM wind system and westerlies. We suggest that the obtained spatial patterns are important meteorological features that need further attention and may be useful in ISM timing and strength prediction.

In the second part of the report, we have presented a generic approach to characterize the evolution of networks. With model tests we established that it is possible to use it to distinguish random, deterministic and periodic evolution in a set of networks. The new quantity to measure variability and persistence in networks is suitable for different network types. For example, the network set may be linearly ordered by time - or by parameter difference. The method can be extended in a straightforward manner, but it currently requires that the node structure and link density remain constant. Applying the CCEF analysis to data from the Asian Monsoon domain we found that El-Nino years are accompanied by a distinct network imprint, leading to small common components with non-ENSO years and high agreement with ENSO years. In future, the CCEF could be a particularly useful tool in the investigation of change points in network evolution.

5 REFERENCES

[1] V. Stolbova, P. Martin, B. Bookhagen, N. Marwan, J. Kurths, Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka, Nonlin. Processes Geophys., 21, 901–917, 2014, doi:10.5194/npg-21-901-2014

[3] L. Tupikina, K. Rehfeld, N. Molkenthin, V. Stolbova, N. Marwan, and J. Kurths, Characterizing the evolution of climate networks, Nonlin. Processes Geophys., 21, 705–711, 2014, doi:10.5194/npg-21-705-2014

[3] N. Molkenthin, K. Rehfeld, V. Stolbova, L. Tupikina, and J. Kurths, On the influence of spatial sampling on climate networks, Nonlin. Processes Geophys., 21, 651–657, 2014, doi:10.5194/npg-21-651-2014

[4] Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: The backbone of the climate network, Europhys. Lett., 87, 48007, doi:10.1140/epjst/e2009-01098-2, 2009a.

[5] Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: Complex networks in climate dynamics, Eur. Phys. J. Spec. Top., 174, 157– 179, 2009b.

[6] Gozolchiani, A., Yamasaki, K., Gazit, O., and Havlin, S.: Pattern of climate network blinking links follows El Niño events, Europhys. Lett., 83, 28005, doi:10.1209/0295 5075/83/28005, 2008.

[7] Gozolchiani, A., Havlin, S., and Yamasaki, K.: The Emergence of El Niño as an Autonomous Component in the Climate Network, Phys. Rev. Lett., 107, 14, doi:10.1103/PhysRevLett.107.148501, 2011.

[8] 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.

[9] 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, 2011.

[10] Rehfeld, K., Marwan, N., Breitenbach, S. F. M., Kurths, J., and Dynamics, C.: Late Holocene Asian Summer Monsoon dynamics from small but complex networks of palaeoclimate data, Clim. Dynam., doi:10.1007/s00382-012-1448-3, 2012.

[11] Tirabassi, G. and Masoller, C.: On the effects of lag-times in networks constructed from similarities of monthly fluctuations of climate fields, Eur. Phys. Lett., 102 59003, doi:10.1209/0295- 5075/102/59003, 2013.

[12] TRMM: http://disc.sci.gsfc.nasa.gov/precipitation/documentation/, last access: 25 February 2014.

[13] Tsonis, A. and Roebber, P.: The architecture of the climate network, Physica A, 333, 497–504, 2004.

[14] Tsonis, A. A. and Swanson, K.: Topology and Predictability of El Niño and La Niña Networks, Phys. Rev. Lett., 100, 228502, doi:10.1103/PhysRevLett.100.228502, 2008.

[15] 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, 2006.

[16] Tsonis, A. A., Swanson, K. L., andWang, G.: On the Role of Atmospheric Teleconnections in Climate, J. Climate, 21, 2990–3001, 2008.

[17] 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, 2010.

[18] Wang, B.: The Asian Monsoon, Springer Praxis Books, Springer, Berlin, Germany, 2006.

[19] Wang, Y., Gozolchiani, A., Ashkenazy, Y., Berezin, Y., Guez, O., and Havlin, S.: Dominant Imprint of Rossby Waves in the Climate Network, Phys. Rev. Lett., 111, 138501, doi:10.1103/PhysRevLett.111.138501, 2013.

[20] 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.

[21] Yamasaki, K., Gozolchiani, A., Havlin, S., and Gan, R.: Climate Networks Based on Phase Synchronization Analysis Track El- Niño, in: Progress of Theoretical Physics Supplement, 178–188, doi:10.1143/PTPS.179.178, 2009.

[22] Yatagai, A., Arakawa, O., Kamiguchi, K., and Kawamoto, H.: A 44-Year Daily Gridded Precipitation Dataset for Asia, B. Am. Meteorol. Soc., 5, 3–6, 2009.