Spatial Analyses of Precipitation Climatology using Climate Networks

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

d o c t o r r e r u m n a t u r a l i u m
(Dr. rer. nat.)
im Fach Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Dipl.-Phys. Aljoscha Rheinwalt

Präsident der der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Elmar Kulke

Gutachter:

1. Prof. Dr. Dr. h.c. J. Kurths
2. Prof. Dr. I. Ebert-Uphoff
3. Prof. Dr. A. Feigin

Tag der mündlichen Prüfung: 20. Juli 2015

to my father

Abstract

The climate system is a complex dynamical system with many closely interacting subsystems, for instance the atmosphere. Although many attempts were, and are still taken to model the climate system as a whole using global circulation models, many issues and challenges persist. Prediction of precipitation is one of the major problems in dynamical models. In the following an approach to the analysis of spatial structures of precipitation event synchronizations is presented. By estimating the synchronicity of precipitation events between points in space, a spatial similarity network is constructed. These Climate Networks can be analyzed statistically in various ways. However, the similarity measure Event Synchronization that will be presented, as well as the concept of Climate Networks, is more general. Climate Network precipitation analyses are done in the applications part in order to present improvements to existing methodologies, as well as novel ones.

On one hand, the existing similarity measure Event Synchronization will be refined and extended to a weighted and continuous version, and on the other hand, new methods for statistical analyses of Climate Networks will be presented. Climate Networks are spatially embedded networks and the probability of a link between two nodes decreases with the distance between these nodes. In other words, Climate Network topologies depend on the spatial embedding. Often this effect is distracting and should be considered as a bias in Climate Network statistics. This thesis provides a methodology to estimate this bias and to correct network measures for it.

Furthermore, two novel graph statistics are introduced. First, the novel network measure Directionality, and second, a network coarse-graining approach that reduces Climate Networks to Climate Networks of teleconnections, i.e., long-ranged interrelations. This new approach is in contrast to existing Climate Network construction schemes, since commonly most links are short. The novel network measure Directionality provides a dominant direction of links in the embedding space. For undirected Event Synchronization networks this measure is applied for the estimation of Isochrones, i.e., lines of synchronous event occurrences.

Zusammenfassung

Das Klima System ist ein so komplexes dynamisches System aus mehreren eng ineinder verflochtenen subsystemen wie z.b. der Atmosphäre, dass es immer noch nicht vollständig gelungen ist dieses als ganzes erfolgreich zu modellieren. Die Vorhersage von Niederschlägen stellt dabei ein besonderes Problem dar. Im folgenden wird ein Verfahren dargestellt welches die Möglichkeit bietet komplexe räumliche Zusammenhänge zwischen Niederschlagsereignissen quantitativ in Klimanetzwerke zu fassen und diese auf vielfältige Arten und Weisen zu analysieren. In dem Maße wie synchronisiert Niederschlagsereignisse zwischen Raumpunkten auftreten, in dem Maße sind diese Raumpunkte in Event Synchronization Klimanetzwerken verbunden. Sowohl das Konzept der vorgestellten Ereignissynchronisation, als auch das Konzept Klimanetzwerke, ist jedoch viel allgemeiner. Die Analyse von Niederschlägen stellt nur eine Möglichkeit dar und wird vollzogen um Fortschritte in der allgemeinen Methodik exemplarisch zu zeigen.

Zum einen wird das bestehende Ähnlichkeitsmaß der Ereignissynchronisation verbessert und erweitert, und zum anderen werden verschiedene, zum Teil neue, statistische Methoden zur Netzwerkanalyse vorgestellt und erläutert. Klimanetzwerke sind räumlich eingebettete Netzwerke und die statistisch zu zeigende Abhängigkeit der Ähnlichkeit vom räumlichen Abstand führt zu einer vom Raum nicht unabängigen Netzwerkstruktur. Dies ist in einer Vielzahl von Fällen ein ungewünschter Effekt und es wird eine Methodik entwickelt wie dieser statistisch quantifiziert werden kann.

Des weiteren werden zwei weitere neue Netzwerkstatistiken vorgestellt. Einerseits das neue Netzwerkmaß Directionality und andererseits eine Netzwerkreduktion welche Klimanetzwerke auf Klimanetzwerke mit weitreichenden Verbindungen reduziert. Dieser neue Ansatz steht gewissermaßen im Gegensatz zur klassischen Klimanetzwerkkonstruktion die vor allem zu kurzreichweitigen Verbindungen führt. Das neue Netzwerkmaß Directionality gibt für jeden Raumpunkt des Netzwerks eine dominante Raumrichtung der Netzwerkverbindungen an und kann dadurch z.B. für bestimmte Event Synchronization Klimanetzwerke Isochronen abbilden.

List of publications

This dissertation is partly based on the following publications. The identifiers, *e.g.*, P1 or C1, given below are cited in the text to highlight passages that are connected to one or more of these publications.

Papers

P1 **A. Rheinwalt**, N. Marwan, J. Kurths, P. Werner, F.-W. Gerstengarbe, *Boundary effects in network measures of spatially embedded networks*, Europhysics Letters (2012).

P2 N. Boers, **A. Rheinwalt**, B. Bookhagen, H.M.J. Barbosa, N. Marwan, J. Marengo, J. Kurths, *The South American Rainfall Dipole: A Complex Network Analysis of Extreme Events*, Geophysical Research Letters (2014).

P3 **A. Rheinwalt**, N. Boers, N. Marwan, J. Kurths, F.-W. Gerstengarbe, P. Werner, *Non-Linear Time Series Analysis of Precipitation Events Using Regional Climate Networks for the Region of Germany*, Climate Dynamics (2015).

P4 **A. Rheinwalt**, N. Boers, B. Goswami, C. Price, N. Marwan, J. Kurths, *Kernel-based Lightning Synchronization: A Complex Network Analysis of Lightning Activity*, (in preparation).

Conference proceedings

C1 **A. Rheinwalt**, N. Marwan, J. Kurths, P. Werner, F.-W. Gerstengarbe, *Boundary Effects in Network Measures of Spatially Embedded Networks*, High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion (2012).

C2 **A. Rheinwalt**, B. Goswami, N. Boers, J. Heitzig, N. Marwan, R. Krishnan, J. Kurths, *Teleconnections in Climate Networks: A Network of Networks Approach to Investigate the Influence of Sea Surface Temperature Variability on Monsoon Systems*, Machine Learning and Data Mining Approaches to Climate Science: Proceedings of the Fourth International Workshop on Climate Informatics (2015).

C3 N. Boers, **A. Rheinwalt**, N. Marwan, J. Kurths, *A complex network approach to investigate the spatiotemporal co-variability of extreme rainfall*, Machine Learning and Data Mining Approaches to Climate Science: Proceedings of the Fourth International Workshop on Climate Informatics (2015).

Acknowledgements

This work was supported by the German Research Foundation by the DFG project "Interactions and complex structures in the dynamics of changing climate", as well as project MA 4759/4-1 and the German Federal Ministry for Education and Research (BMBF) via the Potsdam Research Cluster for Georisk Analysis, Environmental Change and Sustainability (PROGRESS).

Contents

List of Figures

Frequently used Abbreviations and Symbols

A	Adjacency matrix (typically NxN) with entries A_{ij}
s_{ij}	Estimated similarities between i and j
w_{ij}	Link weight for link A_{ij}
PCC	Pearson's correlation coefficient
ρ	Spearman's rho, Spearman's rank correlation coefficient
ES	Event Synchronization
TCN	Teleconnection Climate Network
SST	Sea surface temperature
PCA	Principal component analysis
EOF	Empirical orthogonal function

Network measures

DG	Degree
SG	Strength
MD	Mean Geographical Distance
LC	Local Clustering Coefficient
CC	Closenness Centrality
BC	Shortest-path Betweenness
DR	Directionality
NR??	Normalized ranks of measure ??
	e.g. NRDG are the normalized ranks of Degree
BN	Bottleneckness
LD	Long-range Directedness
??*	Node-splitting invariant version of measure ??
	e.g. CC^* is the node-splitting invariant CC
??$^-$	Spatially corrected version of measure ??
	e.g. CC^- is the spatially corrected CC

Regions in South America

SEBRA	Southeastern Brazil
SESA	Southeastern South America
SACZ	South Atlantic Convergence Zone
ITCZ	Intertropical Convergence Zone
SALLJ	South American Low-Level Jet

Regions in Germany

ML	Mecklenburg lowlands
RM	Rhenish Massif
TF	Thuringian Forest
RP	Rhineland-Palatinate
BF	Bavarian Forest
BA	Bavarian Alps

European atmospheric circulation patterns (Grosswetterlagen)

BM	Bridge Central Europe
HM	High Central Europe
TRM	Trough Central Europe
TRW	Trough Western Europe
NWZ	Northwest Cyclonic
WZ	West Cyclonic

Chapter 1.

Introduction

The study of spatially extended complex systems is a lively and growing field, for instance in astrophysics (Amelino-Camelia, 2011), biology (Hauert and Doebeli, 2004), Earth sciences (Davidsen and Paczuski, 2005), ecology (Blasius et al., 1999; Jones et al., 2008), or medical image analysis (Schmah et al., 2011). In the last decades, powerful tools of time series analysis have been proposed and developed, such as wavelet analysis (Torrence and Compo, 1998), empirical mode decomposition (Huang et al., 1998), singular spectrum analysis (Ghil et al., 2002), or recurrence plots (Marwan et al., 2007b). For a spatial analysis various tools are available, such as empirical orthogonal functions (EOFs) (Hannachi et al., 2007), tools adapted from time series analysis such as spatial recurrence plots (Marwan et al., 2007a; Agustí et al., 2011), or complex networks (Newman, 2003).

In many fields of research, complex networks have proven to be a successful concept for understanding complex systems, e.g. resilience studies of the Internet (Cohen et al., 2000) or power grids (Menck et al., 2014), transport optimization on street networks, power grids and supply chain networks (Danila et al., 2006; Chen et al., 2010), spread of epidemics within populations (Klovdahl et al., 1994; Tuckwell et al., 1998; N. M. Ferguson, 2000), relations from structure to function in brain networks (Sporns et al., 2004; Zhou et al., 2006; Stam, Reijneveld, et al., 2007; Bullmore and Sporns, 2009; Zamora-López et al., 2010; Bashan et al., 2012; Gallos et al., 2012), and even in the analysis of single time series by recurrence networks (Marwan et al., 2009; Donner et al., 2010). Recently, network theory has also been utilized in climate research by so called Climate Networks for understanding complex climate phenomena (Tsonis and Roebber, 2004a; Yamasaki et al., 2008a; Donges et al., 2009a; Steinhaeuser et al., 2010; Paluš et al., 2011, P2).

Especially in meteorology, climatology and oceanography EOFs are traditionally employed as a tool for spatial exploratory data analyses. Such EOFs are derived by a principal component analysis (PCA) (Pearson, 1901) of the covariance matrix of given data set. It is used as a form of dimensionality reduction for spatial feature extraction. It is assumed that relevant similarities between time series are linear and correctly estimated by the covariance matrix, and that relevant spatial structures are represented in its eigenvectors (EOFs). However, although these eigenvectors are linearly uncorrelated, they are in general not independent (Monahan et al., 2009). Hence, it is in general impossible to interpret a single EOF as being of individual dynamical meaning. This carries the disadvantage that one potentially has to combine

an a priori unknown subset of EOFs in an a priori unknown combination in order to identify a physical process of interest.

Climate Networks have certain advantages in such exploratory data analyses. The basic idea of the Climate Network approach is to represent time series of observables under consideration as network nodes, and interrelations among them as network links. Possibly unknown physical processes underlying the co-variabilities between the time series determine the topology of the resulting network and can then be inferred from suitable statistics on these networks. This approach is a more general and adaptable approach compared to EOFs. For instance, it is not restricted to a certain type of similarity matrix and it provides various feature extraction possibilities, e.g. graph statistics from graph theory.

The aim of this thesis is to extend and enhance existing techniques and methods of the Climate Network approach as well as to introduce new ones. The first part of this thesis covers the theoretical framework. It is divided into a chapter about Climate Network construction and a chapter about statistics on Climate Networks, i.e., about feature extraction from Climate Networks. The second part of this thesis contains several applications of the Climate Network approach to precipitation and lightning data. This part begins with the fourth chapter on the general applicability of Event Synchronization as a similarity measure for Climate Networks in analyses of precipitation events. In chapter five the novel network measure Directionality is applied to rainfall networks in South America. In chapter six the novel link-weighted and spatially corrected Directionality is applied to precipitation networks in Germany. Finally, in chapter seven a novel Climate Network coarse-graining approach is applied to an Asian precipitation network, to a global sea surface temperature network, as well as to a combined network of these Climate Networks in order to identify teleconnections within and between these systems.

Part I.

Theoretical Framework

This part also provides the theoretical foundation of associated publications P1 to C3, and follows in parts closely corresponding sections in these publications.

Chapter 2.

Climate Network Construction

Abstract

Climate Networks have proven to be a useful concept in various studies of climate systems. This chapter briefly reviews the Climate Network approach and introduces for this thesis relevant Climate Network construction schemes. Relevant similarity measures are introduced that serve as basis for the construction of Climate Networks. Furthermore, statistical network models are presented with which the statistical significance of found Climate Network topologies is tested.

2.1. Climate Networks

2.1.1. Introduction

The conception of Climate Networks is to represent estimated interrelations between a possibly large set of climatic time series as links of a network that can be analyzed statistically. This is in contrast to, for instance, recurrence networks from climatic time series where recurrences of a single time series are encoded in a network (Marwan et al., 2009; Donner et al., 2010). With the Climate Networks approach spatially extended systems are thus studied with respect to interrelations between the time series. Depending on the observable or observables of given climatic time series and the measures used to estimate interrelations between them, a resulting Climate Network encodes certain features of the underlying climate system. A measure that quantifies an interrelation between two time series is referred to as a *similarity measure*. Many different similarity measures have been studied in the context of Climate Networks, most prominently Pearson's correlation coefficient (PCC), Spearman's rho (ϱ), their cross-correlation versions and mutual information (MI). However, more recently the event-based similarity measure Event Synchronization (ES) (Quiroga et al., 2002) got applied in various forms to Climate Networks in studies of precipitation events (e.g. Malik et al., 2012a; Boers et al., 2013, P1, P2, P3, C1, C3).

2.1.2. Network Representation

In graph theory a *graph* G is defined as a set of *vertices* V together with a set of *edges* E connecting *vertices* (Newman, 2010):

$$G := (V, E) .$$

As it is more common in the physics literature, I will refer to *graphs* as networks, to *vertices* as nodes and to *edges* as links. Climate Networks are spatially embedded networks since nodes represent time series that have a one-to-one correspondence to geographical positions. In this thesis Climate Networks occur as node-weighted, link-weighted or unweighted networks. Node-weighted networks have a weight u_i attached to each node V_i and link-weighted networks have a weight w_{ij} attached to each link E_k connecting node V_i to node V_j. In this thesis Climate Networks can furthermore be directed or undirected. Undirected networks have undirected links and hence, if node V_i is connected to node V_j by a single undirected link, node V_j is also connected to node V_i by that same link. For undirected and link-weighted networks w_{ij} is equal to w_{ji} for all links.

In the following, the general and common representation of Climate Networks by their adjacency matrix A together with its corresponding spatial embedding and potential node weights u is utilized. The spatial embedding is provided by a pair of geographical coordinate vectors (X, Y) with a geographical position (X_i, Y_i) for each node V_i. A is defined as

$$A_{ij} := w_{ij} ,$$

with $w_{ij} = 0$ if the corresponding nodes are not connected. For unweighted networks A simplifies to,

$$A_{ij} = \begin{cases} 1 & \text{if } V_i \text{ is connected to } V_j , \\ 0 & \text{otherwise} . \end{cases}$$

2.2. Similarity Measures

An initial idea behind the Climate Network approach was to consider *climate as a network of many dynamical systems* (Tsonis and Roebber, 2004b). And further, to study the collective behavior of these dynamical systems by an investigation of the coupling architecture of this network. This idea was pursued in various studies by analyzing spatially extended sets of time series of climatic observables by similarity measures.

Depending on the value distribution of time series and the assumed functional form of relationships between time series, different similarity measures might be suitable. Although the climate system might have various non-linear interactions, the coupling between Climate Network nodes does not necessarily have to be assumed as non-linear. Complex dynamical behavior can also emerge from linearly coupled

dynamical systems. In fact, some Climate Networks constructed by a linear similarity measure are surprisingly similar to those constructed by a non-linear measure (Donges et al., 2009a).

Additionally, depending on the spatial scale of the Climate Network and the temporal resolution of time series, an assumed relationship between time series might occur with a certain delay. Traditionally, this is studied by a *lead-lag analysis* using cross-correlation. However, especially with higher temporal resolutions of time series covering climatic periods such a single delay for a pair of time series might be inappropriate. An interesting similarity measure in that context is Event Synchronization, which is designed to handle dynamical delays. The following similarity measures occur in this thesis:

Pearson's correlation coefficient (PCC): The probably most common similarity measure is Pearson's correlation coefficient (PCC). It quantifies linear dependencies and can for centered data distribution be interpreted geometrically as the cosine of the angle between the two time series as vectors. In this case the PCC can be seen as a normalized dot product. If $PCC = 0$ the two vectors are orthogonal and therefore linearly independent. If $PCC = 1$ or $PCC = -1$, the one time series is a linear combination of the other and vice versa. For two time series x and y of length l with means $\bar{x}$ and $\bar{y}$ and standard deviations σ_x and σ_y the Pearson's correlation coefficient is defined as

$$PCC_{x,y} := \frac{\mathrm{Cov}(x,y)}{\sigma_x \sigma_x} = \frac{\sum_{t=1}^{l}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{l}(x_t - \bar{x})^2}\sqrt{\sum_{t=1}^{l}(y_t - \bar{y})^2}} \in [-1, 1]\,. \tag{2.1}$$

Hence, for centered data $\hat{x}, \hat{y}$:

$$PCC_{\hat{x},\hat{y}} \equiv \frac{\hat{x} \cdot \hat{y}}{\|\hat{x}\|\,\|\hat{y}\|} = \cos(\theta_{\hat{x},\hat{y}})\,,$$

with the angle $\theta_{\hat{x},\hat{y}}$ between vectors.

Spearman's rho (ϱ): The Spearman's rho ϱ is defined as the Pearson's correlation coefficient between the ranked time series. It therefore provides a non-parametric measure of a monotonic relationship between the two time series. In that sense it is a non-linear measure. Unlike the Pearson's correlation coefficient it is robust to outliers.

Event Synchronization (ES): This non-linear synchronization measure first introduced by Quiroga et al., 2002 is defined on event series. It quantifies the synchronicity between events in two given event series e_i^{μ} and e_j^{ν} by counting the number of events that can be uniquely associated with each other within a prescribed maximum delay, while taking into account their temporal ordering.

Event series can be constructed from natural time series in various ways and in the applications part of this thesis several thresholding techniques are pursued. One can also think of natural event series like series of earth quakes or lightning strokes. In the following also a novel, kernel-based event synchronization measure is introduced that is used to quantify synchronicity of lightning activity (P4).

Consider two event series e_i^μ and e_j^ν, where e_i^μ (e_j^ν) denotes the time index of the μ-th (ν-th) event observed at grid point i (j). In order to decide if two events e_i^μ and e_j^ν with $e_i^\mu > e_j^\nu$ can be assigned to each other uniquely, we first compute the waiting time $d_{ij}^{\mu,\nu} := e_i^\mu - e_j^\nu$, and then define the *dynamical delay*:

$$\tau_{ij}^{\mu\nu} = \frac{1}{2}\min\{d_{ii}^{\mu,\mu-1}, d_{ii}^{\mu,\mu+1}, d_{jj}^{\nu,\nu-1}, d_{jj}^{\nu,\nu+1}\}\,. \tag{2.2}$$

In order to avoid nonphysical synchronizations of events in sparse event series, we further introduce a maximum delay τ_{max} which shall serve as an upper bound for the dynamical delay. If then $0 < d_{ij}^{\mu,\nu} \leq \tau_{ij}^{\mu\nu}$ and $d_{ij}^{\mu,\nu} \leq \tau_{max}$, we count this as a directed synchronization from j to i:

$$S_{ij}^{\mu\nu} = \begin{cases} 1 & \text{if} \quad 0 < d_{ij}^{\mu,\nu} \leq \tau_{ij}^{\mu\nu} \quad \text{and} \quad d_{ij}^{\mu,\nu} \leq \tau_{max}\,, \\ 0 & \text{else}\,. \end{cases} \tag{2.3}$$

Directed Event Synchronization (Boers et al., 2014a) from j to i is then given by

$$ES_{ij}^{dir} := \sum_{\mu\nu} S_{ij}^{\mu\nu}\,. \tag{2.4}$$

A symmetric version of this measure can be obtained by also counting events at the very same time as synchronous and taking the absolute value of the dynamical delay,

$$\overline{S}_{ij}^{\mu\nu} = \begin{cases} 1 & \text{if} \quad |d_{ij}^{\mu,\nu}| \leq \tau_{ij}^{\mu\nu} \quad \text{and} \quad d_{ij}^{\mu,\nu} \leq \tau_{max}\,, \\ 0 & \text{else}\,, \end{cases} \tag{2.5}$$

and computing the corresponding sum:

$$ES_{ij}^{sym} := \sum_{\mu\nu} \overline{S}_{ij}^{\mu\nu}\,. \tag{2.6}$$

A continuous version of this measure can be obtained by a kernel-based approach. Instead of a dynamical delay τ a dynamical delay kernel κ is incorporated into the synchronicity estimation. For a given kernel κ the Kernel-based Event Synchronization is defined as

$$KES_{ij} := \sum_{\mu\nu} \kappa_{ij}^{\mu\nu}\,. \tag{2.7}$$

A directed measure is achieved by an asymmetric kernel

$$ {}^{asym}\kappa_{ij}^{\mu\nu} = \begin{cases} f(d_{ij}^{\mu,\nu}, \dots) & \text{if} \quad d_{ij}^{\mu,\nu} > 0\,, \\ 0 & \text{else}\,, \end{cases} \tag{2.8} $$

where f is the actual function of the kernel with possible parameters in case of a dynamic delay kernel. A possible parameter could be for instance $\tau_{ij}^{\mu\nu}$. An undirected Kernel-based Event Synchronization would have a symmetric kernel, for instance a Gaussian:

$$ {}^{gauss}\kappa_{ij}^{\mu\nu} = \exp - \frac{\left(d_{ij}^{\mu,\nu}\right)^2}{2\left(\sigma_{ij}^{\mu,\nu}\right)^2}\,, \tag{2.9} $$

where $\sigma_{ij}^{\mu,\nu}$ is the standard deviation of the Gaussian. An undirected Kernel-based Event Synchronization using a dynamic delay Gaussian kernel is therefore given by

$$ KES_{ij}^{gauss} = \sum_{\mu\nu} \exp - \frac{\left(d_{ij}^{\mu,\nu}\right)^2}{2\left(\sigma_{ij}^{\mu,\nu}\right)^2}\,. \tag{2.10} $$

The Kernel-based Event Synchronization can be seen as a generalization of the common Event Synchronization. A dynamical step function as a delay kernel would correspond to the previously defined ES versions using a dynamical delay τ. However, as the kernel is in general a continuous function it does not rely on temporal sampling as in time series. Therefore, it is directly applicable to continuous time measurements such as lightning activity data (P_4).

2.3. Statistical Network Models

As links of Climate Networks are often estimated using traditional measures of correlation (dependence), the statistical inference approaches of statistical hypothesis testing and confidence interval estimation are also applicable to Climate Network topologies. For instance, a common question in the Climate Network approach is whether an estimated similarity between two time series (nodes) is statistically significant according to a certain significance level. In the case of Pearson's correlation coefficient and bivariate normally distributed time series, such a question can be answered analytically. However, with more complex similarity measures or non-normal distributed observables, a statistical null model often has to be computed numerically by surrogate data sets. Prominent examples are surrogates retrieved by permutation or bootstrapping, sometimes referred to as shuffled surrogates, Fourier surrogates (Schreiber and Schmitz, 1996; Schreiber and Schmitz, 2000), or twin surrogates (Thiel et al., 2006; Marwan et al., 2007b; Thiel et al., 2008; Romano et al., 2009) as also applied in the associated publication C2.

However, in studies of complex networks also other statistical models have been developed that are more network specific. Depending on the null hypothesis they range from random network models like the Erdős-Rényi network model (Solomonoff and Rapoport, 1951; Erdős and Rényi, 1959) over configuration models (Newman and Park, 2003; Squartini and Garlaschelli, 2011) to network models that respect the spatial embedding. An example of such a spatial network model will be introduced in the following chapter (see Sec. 3.3).

In order to infer the statistical significance of Event Synchronization estimates we will employ shuffled surrogates. The associated statistical model also reveals a bias in the measure Event Synchronization that depends on the event rate in event time series and is therefore called *event rate bias* (P3). Furthermore, we could show that this model is an analytical model for the case of no delay, i.e., for simultaneously occurring events.

2.3.1. Event Rate Bias

Given two event series e_i^μ and e_j^ν with events $\mu \in \{1, \ldots, m\}$ and $\nu \in \{1, \ldots, n\}$ with $n \geq m$, the maximum number of synchronizations in the measure Event Synchronizations is limited by the number of events m. The limit would correspond to the case that all m events in e_i^μ synchronize with events in e_j^ν. In other words,

$$\sum_{\mu\nu} S_{ij}^{\mu\nu} \leq m\,. \tag{2.11}$$

Additionally, the limit gets more likely the higher event rates are. Where event rates are given by $\frac{m}{l}$ and $\frac{n}{l}$ with the initial time series length l. Specifically, these synchronization counts are biased by the number of events and we refer to this bias as *event rate bias*. The initially proposed Event Synchronization by Quiroga et al., 2002 is normalized by $\sqrt{mn}$. However despite this normalization the measure remains biased. In the appendix A.3 a comparison between the original Event Synchronization (normalized by $\sqrt{mn}$) and the unbiased variant presented in chapter 6 is done.

For the general Event Synchronization with delay or the Kernel-based Event Synchronization the *event rate bias* has to be estimated numerically by shuffled surrogates. However, for synchronously occurring events the distribution of the expected number of synchronous events assuming a uniformly random distribution of events can be written as the following:

$$p(k) = \binom{m}{k} \frac{(l-m)!}{l!} \prod_{i=0}^{k-1} n - i \prod_{j=0}^{m-k-1} l - n - j\,, \tag{2.12}$$

where $p(k)$ is the probability to have k synchronizations between two event series of length l with n and m events, respectively, at uniformly random time positions. In the appendix A.2 we provide an efficient algorithm for the computation of the exact distribution $p(k)$ of the expected number of synchronization counts. The distribution

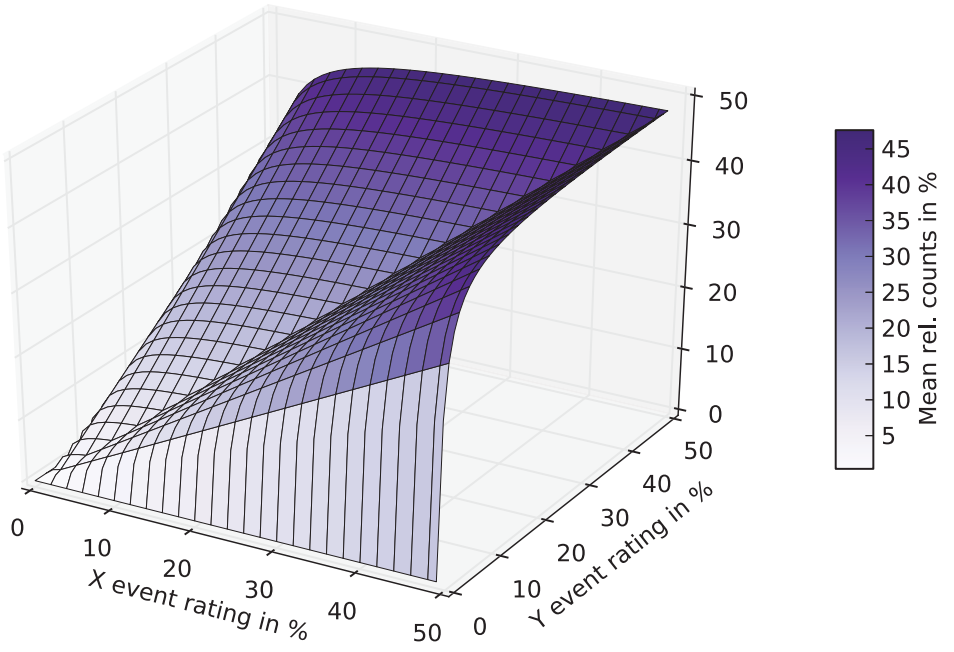

Figure 2.1.: *Event rate bias* for synchronously occurring events as a function of two random event series X and Y: We show the mean number of synchronizations in percent of the maximum possible number of synchronizations (mean rel. counts in %) given X and Y. The length of these series is 100, hence, the percentage is also the actual number of events in this example.

$p(k)$ of the expected number of synchronous events assuming a uniformly random distribution of events can serve as a statistical null model and will also be used as such.

The *event rate bias* is estimated by the expected number of counts $\langle k \rangle_p$, where $p(k)$ is either estimated numerically or given by Eq. (2.12). The higher the event rates, the more events synchronize at random and hence, the higher is the *event rate bias*. The expected number of counts $\langle k \rangle_p$ in percent of the maximum possible number of counts (m) for the synchronous (analytic) case is seen in Fig. 2.1.

2.4. Summary

Technically, the Climate Network approach can be divided into five subsequent steps:

1. **Data pre-processing**: Depending on, for instance the research question, the similarity measure or the data product itself, it might be necessary to pre-process given time series. The question of research might demand the analysis of anomalies in respect to a given climatology, or might call for a more

aggregated measure in time or space. As in the case of precipitation analyses in this thesis, event series for the similarity measure Event Synchronization are constructed in this step from the continuous observable precipitation time series. However, many more pre-processing steps are possible.

2. **Similarity estimation**: For a given set of N pre-processed time series $x_i, i \in \{1, \ldots, N\}$, all pair-wise similarities s_{ij} are estimated by a given similarity measure, leading to a NxN similarity matrix. Depending on the similarity measure used, this similarity matrix can be asymmetric, as with for instance Directed Event Synchronization:

   $s_{ij} = ES_{ij}^{dir}$.

3. **Similarity significance**: In order to remove spurious similarities from a retrieved similarity matrix a null hypothesis of statistical independence has to be formed. For most such null hypothesis in applications of Climate Networks no analytic models exist because of the similarity measures used or because of not normally distributed data. Null models are therefore usually computed numerically by surrogate data. Hence, this step often involves the generation of surrogate data in order to estimate the distributions of similarities in the null model. However, no matter whether these distributions are estimated numerically or given analytically, they are then employed in the hypothesis testing of each originally found similarity s_{ij}. Accordingly, each similarity s_{ij} is tested and then discarded if it fails to reject the null hypothesis with a predefined significance level, i.e., the discarded similarity is set to zero.

4. **Climate Network construction**: A similarity matrix that is cleared from all spurious similarities as described in the previous step, can directly be analyzed as a link-weighted Climate Network by identifying link weights with similarities:

   $w_{ij} = s_{ij}$.

   The similarity matrix can also be thresholded to an unweighted Climate Network by a threshold function $k(i, j)$ which is often chosen to be constant:

   $$A_{ij} = \begin{cases} 1 & \text{if } s_{ij} > k(i,j) \\ 0 & \text{otherwise} \end{cases} \quad .$$

5. **Climate Network feature extraction**: Which features of a Climate Network are quantified and how depends obviously on the research question. Possibilities include, but are not limited to, the study of Climate Networks by network measures (e.g., Directionality, see P2 and P3) or by coarse-graining approaches (cf. teleconnection coarse-graining in C2, e.g.) But many other criteria can be applied to Climate Networks in order to extract certain features. For instance one can remove non-causal links and thereby reveal to topology of causal

networks (Ebert-Uphoff and Deng, 2012a; Runge et al., 2012). The theoretical framework for this step in the Climate Network approach is covered in the next chapter about Climate Network statistics.

Chapter 3.

Climate Network Statistics

Abstract

Climate Networks are high dimensional and spatially embedded complex networks. However, physical processes in the underlying climate system can be inferred from Climate Networks by statistical analyses of their topologies and its spatial embedding. Common graph statistics are network measures. Those relevant for this thesis are introduced in this chapter. Among them also the novel network measure *Directionality* which measures dominant link directions in the embedding space. Structural properties of Climate Networks are influenced by the spatial embedding of the network, which leads to biases in network measures. This chapter proposes a methodology to estimate such *spatial biases* and to correct network measures for it. Apart from network measures, also another statistical approach is presented: a coarse-graining of complex networks that reduces a Climate Network to its *Teleconnection Climate Network*

3.1. Introduction

In accordance with the growth of computational power in science, also the size of networks studied increased. In order to analyze and interpret these networks various approaches have evolved. For instance, network rendering techniques that make complex networks visually interpretable by showing them in an equilibrium state regarding node positions in the visualization space (Kamada and Kawai, 1989; Fruchterman and Reingold, 1991; Wiese et al., 2004, e.g.). But also edge bundling techniques that merge similar links in order to reveal high-level link patterns (Holten and Van Wijk, 2009; Lambert et al., 2010; Pupyrev et al., 2011, e.g.), as well as coarse-graining methods that merge topologically similar nodes (Itzkovitz et al., 2005, e.g.) or community detection approaches (Girvan and Newman, 2002; Newman and Girvan, 2004; Duch and Arenas, 2005; Aldecoa and Marín, 2013, e.g.) have been proposed.

Climate Networks have the specialty that they are spatially embedded networks and hence, certain topological features of the networks are related to climatological features of the underlying climate system in the embedding space. This makes those network statistics favorable that respect the spatial embedding. An example of a class of very sparse Climate networks are Causal Climate Networks (Ebert-Uphoff

and Deng, 2012a; Runge et al., 2012). However, even those are usually analyzed using network statistics.

In Climate Network studies the most common network statistics are node-based network measures. Node-based measures have the advantage that they retrieve a score for each node of the network and hence, for each location in the embedding space. Such a field of network measure values can then be visualized on a map by use of geographical node positions (X, Y). Topological features of a Climate Network can thereby be directly related to the corresponding geographical region.

Although most Climate Network studies use network measures, it is often challenging to find straightforward climatological interpretation for measures which where developed in different research fields like for instance social sciences. Combining network measures in certain ways can sometimes clarify ambiguities by ruling out specific interpretations. Two examples are introduced in the next Section. Furthermore, new network measures especially for Climate Networks with a clear climatological interpretation are necessary in certain cases. In the associated publications P2 and P3 we put this idea forward and introduce such a network measure called Directionality.

3.2. Network Measures

In this Section all network measures used in this thesis are defined. These include the well established measures Degree, Strength, Local Clustering Coefficient, Closeness Centrality and Shortest-path Betweenness (Newman, 2010), but also recently developed climate network related measures like Mean Geographical Distance or Long-ranged Directedness (Boers et al., 2013) and the novel vector network measure Directionality (see P2 and P3). All measures are node-based, hence reveal statistics for each node $i \in 1, \ldots, N$ of the network. Except for Directionality all network measures are scalar measures. Link-weighted or node-weighted (Heitzig et al., 2012; Wiedermann et al., 2013) versions of network measures are only introduced for cases that occur in this thesis.

3.2.1. Scalar Measures

Degree (DG): The Degree of a node i is defined as the number of links that are attached to this node:

$$DG_i := \sum_{j=1}^{N} A_{ji} \,. \tag{3.1}$$

The node-weighted Degree (DG^*) is defined as:

$$DG_i^* := \sum_{j=1}^{N} u_j A_{ji} \,, \tag{3.2}$$

with node weights u. The link-weighted Degree is traditionally called Strength.

Strength (SG): Link-weighted networks have a weight attached to each link and are represented by the matrix of link weights with zero weights for no links (see Section 2.1.2). Therefore, with $A_{ij} = w_{ij}$ the Strength of a node i is defined as:

$$SG_i := \sum_{j=1}^{N} w_{ji} \,. \tag{3.3}$$

However, in this thesis the measure Strength occurs mainly in the context of directed networks where the Strength of a given node i is separated into the Strength for links pointing away from that node (SG_i^{out}) and towards that node (SG_i^{in}) (Newman, 2010):

$$SG_i^{out} := \sum_{j=1}^{N} w_{ji} \tag{3.4}$$

$$SG_i^{in} := \sum_{j=1}^{N} w_{ij} \,. \tag{3.5}$$

Mean Geographical Distance (MD): The Mean Geographical Distance of a node i is defined as the mean spatial length of all links attached to it:

$$MD_i := \frac{1}{DG_i} \sum_{j=1}^{N} dist(i,j) A_{ji} \,, \tag{3.6}$$

where $dist(i,j)$ is the great-circle distance between node j and node i.

Local Clustering Coefficient (LC): The Local Clustering Coefficient of node i is defined as the fraction of the number of triangles and the number of pairs of links attached to it. It is thus an estimate of the probability that two nodes connected to node i are connected as well.

$$LC_i := \frac{\sum_{j<k} A_{ij} A_{jk} A_{ki}}{\sum_{j<k} A_{ij} A_{ki}} \,. \tag{3.7}$$

Closeness Centrality (CC): Closeness Centrality is a shortest-path-based network measure. For a given node i it is defined as the inverse mean topological distance of that node to all other nodes of the network:

$$CC_i := \frac{N}{\sum_{j=1}^{N} d_{ji}} \,, \tag{3.8}$$

where d_{ji} is the topological distance (i.e., the number of links) of a shortest path between node j and node i.

The node-weighted Closeness Centrality (CC^*) for a node i is defined as:

$$CC_i^* := \frac{\sum_{j=1}^{N} u_j}{u_i + \sum_{j=1}^{N} u_j d_{ji}}\,, \tag{3.9}$$

with node weights u. If all node weights are equal to one CC^* is given by

$$CC_i^* = CC_i \frac{\sum_{j=1}^{N} d_{ji}}{1 + \sum_{j=1}^{N} d_{ji}}\,.$$

Thus, in this case CC_i^* converges to CC_i for $N \to \inf$.

Shortest-path Betweenness (BC): The Shortest-path Betweenness of node i is defined as the number shortest paths in the network that go through node i:

$$BC_i := \sum_{l<k\neq i} \frac{\sigma_{lk}(i)}{\sigma_{lk}}\,, \tag{3.10}$$

where σ_{lk} is the total number of shortest paths between node l and node k, and $\sigma_{lk}(i)$ the number of shortest paths between node l and node k that go through node i. For larger networks this measure has a highly skewed distribution of scores and is often visualized by its decadic logarithm.

3.2.2. Combined Measures

Especially for more complex network measures there is often not a unique explanation for their values. For instance, a high Shortest-path Betweenness score can occur for topological bottle necks or for hubs in a network. Hubs, nodes with a high Degree, tend to be topologically well connected, and hence, many shortest paths go through hubs. Degree and Shortest-path Betweenness fields are therefore often very similar. However, bottle necks have a high BC score regardless of their DG score. Accordingly, bottle necks can be detected by a combination of the network measures Degree and Shortest-path Betweenness. In order to combine network measures we first calculate the normalized ranks of their scores. These normalized ranks of network measures have the prefix NR. For instance, the normalized ranks of Degree are given by

$$NRDG := \frac{1}{N}\mathrm{rank}(DG) \in (0, 1]\,. \tag{3.11}$$

A good combination for the estimation of bottle necks would then be the measure Bottleneckness.

Bottleneckness (BN): The Bottleneckness of a node i is given by the combination of normalized ranks of Degree ($NRDG$) and Shortest-path Betweenness ($NRBC$):

$$BN_i := NRBC_i - NRDG_i \in [0, 1)\,. \tag{3.12}$$

Long-ranged Directedness (LD): Another, climatologically more interesting combination of network measures is the Long-ranged Directedness introduced in C3. It is calculated from the normalized ranks of Shortest-path Betweenness ($NRBC$), Mean Geographical Distance ($NRMD$) and Local Clustering Coefficient ($NRLC$). The Long-ranged Directedness of a node i is given by

$$LD_i := \frac{1}{2}NRBC_i + \frac{1}{2}NRMD_i - NRLC_i \in [0, 1)\,. \tag{3.13}$$

3.2.3. Directionality

Directionality (DR) is a novel node-based network measure introduced first in P3 as a measure for link-weighted networks including spatial effect estimation. A modified version for unweighted networks was applied in P2. It is a non-scalar network measure as it reveals the dominant link direction for each node i as well as the dominance of that direction as a vector in polar coordinates $DR_i = (\rho_i, \varphi_i)$. These vectors DR_i are vectors in the two dimensional embedding space, which makes spatial effect estimation for this network measure fairly interesting.

The dominant link direction of a node i is determined by the mode of the frequency distribution $P_i(\varphi)$ of link directions in the embedding space. This distribution is determined from all links of node i. In the case of directed networks one has to distinguish between links pointing away from node i and links pointing toward node i. Link directions are represented by angles, i.e., the direction of a link between node i and j is defined as

$$\phi(i, j) := \arctan \frac{Y_j - Y_i}{X_j - X_i}\,, \tag{3.14}$$

given the geographical coordinates (X_i, Y_i) and (X_j, Y_j) of nodes i and j. For undirected networks two directions are identical if they point in opposite directions, therefore angles are projected onto the semicircle so that $\phi(i, j) = \phi(j, i)$. In practice $P_i(\varphi)$ is calculated using a fuzzy angle definition instead of, for instance, ordinary binning. In that definition two angles are considered to be identical if they do not differ by more than ϵ. For applications we commonly used $\epsilon = 0.02$ rad. The Directionality of node i is then defined as:

$$\rho_i := \max_{\varphi} P_i(\varphi) \tag{3.15}$$

$$\varphi_i := \arg\max_{\varphi} P_i(\varphi)\,, \tag{3.16}$$

with $P_i(\varphi)$ being estimated by:

$$P_i(\varphi) = \sum_{j:\phi(i,j)\in(\varphi-\epsilon,\varphi+\epsilon)} A_{ij} \,. \tag{3.17}$$

The link-weighted version is retrieved by replacing the entries of the adjacency matrix A by the link weights w_{ij}. The frequency distribution of link directions is then adapted accordingly:

$$P_i(\varphi) = \sum_{j:\phi(i,j)\in(\varphi-\epsilon,\varphi+\epsilon)} w_{ij} \,. \tag{3.18}$$

In Figure 3.1 an example of this measure is shown on an Event Synchronization network from artificial event time series. These time series are short (length is 250) and noisy. Only a single front of events is inherent in this set of time series and additionally this set has a random event at 1% of the times. The network is constructed using Event Synchronization without delay and hence, weighted links occur only for synchronous events and are undirected. Accordingly, the dominant orientation of links estimated by the Directionality feature *isochrones*. *Isochrones* are lines in the embedding space along which events synchronize statistically more often than along other directions.

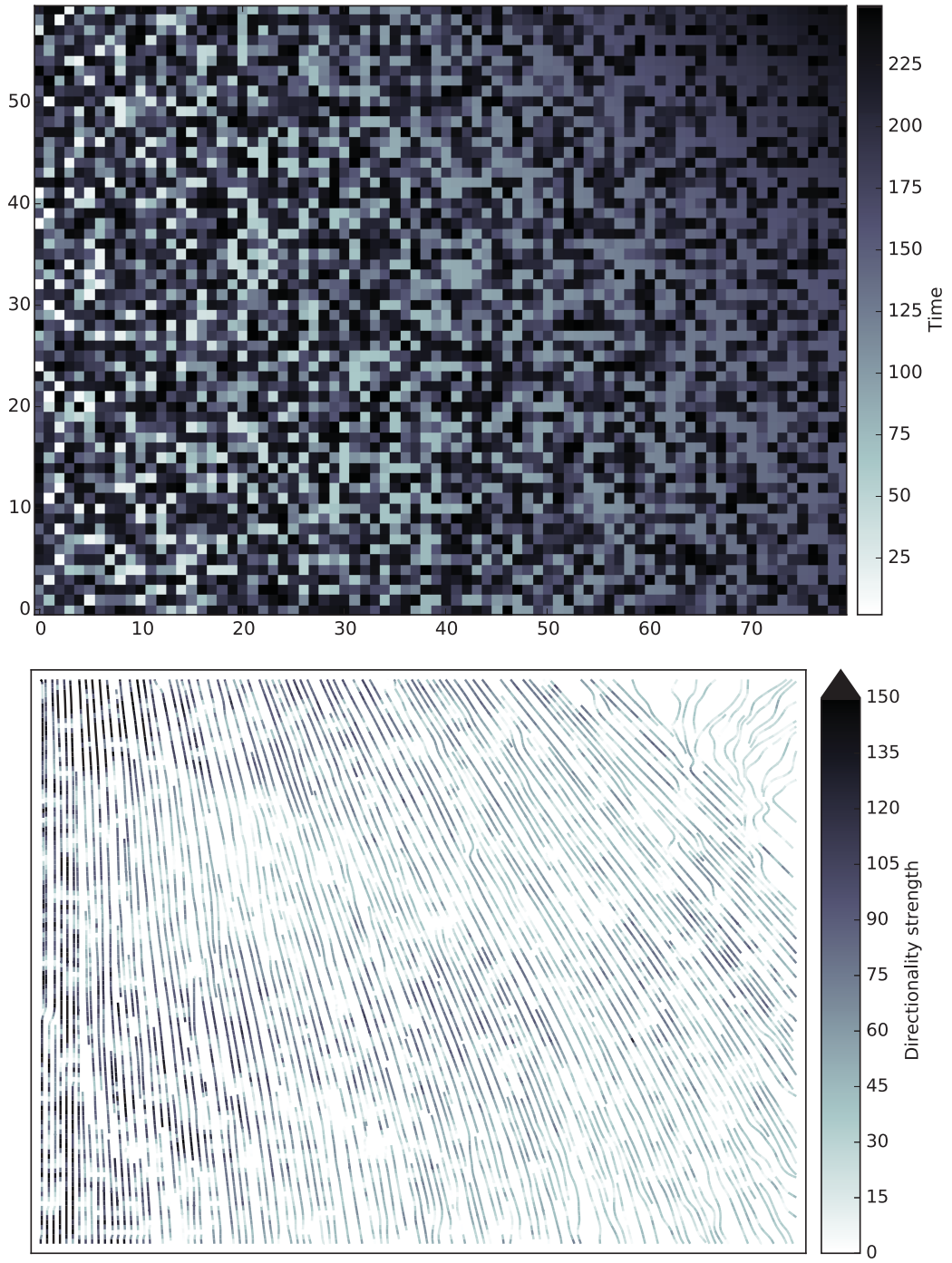

Figure 3.1.: Performance example for the network measure Directionality on Event Synchronization networks from short and noisy artificial event time series. *Top*: Spatial visualization of the set of 4800 (80x60 grid) short and noisy artificial event time series of length 250. Each grid point is colored according to the time of the last occurrence of an event in the corresponding time series. Observe the single event front moving from the left to the top-right. The event front is blurred due to noise. *Bottom*: The network measure Directionality for the Event Synchronization network without delay constructed from the set of event time series shown above. Observe the close resemblance between event front orientation and Directionality *isochrones*.

3.3. Spatial Effects

3.3.1. Introduction

Many networks are spatial networks. However, the network structure is often influenced by the spatial embedding due to distance-based costs of links, i.e., the link probability depends on the spatial length of links (Gastner and Newman, 2006; Barnett et al., 2007; Henderson and Robinson, 2011; Barthelemy, 2011). Although this effect is usually isotropic, it becomes anisotropic if boundaries in space are introduced to the network, as this is the case with the spatial confinement of brain networks which are embedded in three dimensional space and confined by the area of placed electrodes (Henderson and Robinson, 2011). Climate Networks might be bounded if only a smaller region is considered (Yamasaki et al., 2008a; Malik et al., 2011; Boers et al., 2013, e.g.); similarly, power grids are confined by the economic region (e.g., by the boundary of the European Network of Transmission System Operators for Electricity). We also call spatially confined networks *regional* networks.

Boundaries cut links which would connect the region under consideration with the outside region. Obviously, this artificially reduces node degrees and the amount of longer links in the remaining network, and hence influences corresponding network measures. The effect of cut links due to boundaries is larger when the network consists of many long links, as the probability is high that such links connect the inside and the outside regions. How strong boundaries affect network measures depends on the distribution of link lengths and on the network measures themselves.

Based on the network of interest and its spatial confinement, boundary effects might be negligible, of interest, or distracting from network structure not imposed by boundaries. Neglecting boundary effects can lead to spurious conclusions, e.g., for the identification of hubs in brain networks (Henderson and Robinson, 2011). In many applications, resulting boundary effects are often not negligible and, consequently, network measures should be corrected in order to exclude them.

The dependence of the link probability on the spatial length of links has also anisotropic effects on the network structure if the node sampling in the embedding space is not homogeneous. For instance Climate Networks from regular latitude longitude gridded data sets suffer from a bias in the topological difference between regions where nodes are close to each other and regions where nodes are far away from each other. This leads to statistically more links in polar regions than in equatorial regions. This particular spatial effect has also been described by Heitzig et al., 2012, who propose an analytic approach that introduces node weights to network measures and makes network measures node splitting invariant. Although this approach removes the bias of this specific spatial effect (see Fig. 3.4) it does not remove the bias caused by boundaries.

Here we propose a correction procedure for network measures that consistently estimates all spatial effects (i.e., effects on network measures due to the spatial embedding). We will use a specific random network construction with properties similar to those in the original network, i.e., it shares a similar link probability $p(\Delta_{ik})$

that two nodes, i and k that have the distance Δ_{ik} in space, are linked. This $p(\Delta)$ is not the probability to find a link of length Δ among all, but rather the probability conditioned on the number of possible links of that length due to the embedding of nodes in space.

3.3.2. Spatial Effect Estimation

In spatially embedded random networks (SERN), influences of spatial embedding on network structure are quantified by a link probability that depends on the spatial length of a link in the embedding metric space (Barnett et al., 2007). We propose this as a model for spatial effects and generate SERN for the same node positions in space as the original network and with the same link probability depending on spatial link lengths as the network of interest. Thus for a spatial network, we consider the result of a certain network measure on such a SERN as an estimate of spatial effects in that measure. Hence, the SERN we use is a surrogate in the sense that it mimics the same length dependency in the link probability as in the original network:

- Nodes are embedded in a metric space S with the metric $\Delta\colon S \times S \to \mathbb{R}^+$; thus Δ_{ik} is the spatial distance between node i and k.
- Nodes have given positions (X, Y) in S. These positions are the same as in the original network.
- Nodes i and k are connected with the link probability $p(\Delta_{ik})$, i.e., the probability of finding a link of length Δ_{ik} in the original network with respect to how many links of that length could occur.

Depending on the positioning of nodes in space, a binning of spatial link length might be necessary in order to improve the link probability estimate of the original network. This can be achieved by rounding spatial link lengths to appropriate integers so that similar lengths fall into one integer length. A measurement using such a procedure can be done with the following algorithm. Here A_d is the number of possible links with integer length d and B_d is the number of actually present links with integer length d. The fraction p_d of both is an estimate for the underlying link probability $p(\Delta)$.

$A_d = B_d = p_d = 0 \quad \forall\, d \in \text{rounded } \Delta$
for $i \in nodes$ **do**
  for $k < i$ **do**
    $d \leftarrow$ rounded Δ_{ik}
    increase A_d by one
    if node i and k are linked in the original network **then**
      increase B_d by one
    end if
  end for
end for

$p \leftarrow \frac{B}{A}$

Link probabilities p for real data derived by this algorithm can be seen in the applications part in Fig. **??**.

Unweighted Networks The SERN are a statistical model for spatial effects for unweighted networks. Prior to modeling spatial effects with SERN, the link probability $p(\Delta)$ for these random networks has to be estimated from the network of interest. In contrast to the spatial effects estimation for link-weighted networks, an unweighted surrogate network is then constructed by use of $p(\Delta)$. However, one realization of such a surrogate is not enough. In order to improve the estimation of spatial effects in a certain network measure, such as Closeness Centrality, the result of that measure has to be averaged over an ensemble of surrogates. The reliability of such an estimate is then derived from the distribution of ensemble measures.

If a node-based network measure that returns a value for each node is used, an estimate of spatial effects can be found for each node in space. An average estimate such as the node-wise median from 1000 surrogates is shown for the Closeness Centrality field in Fig. 3.2 (bottom-left). As a measure of reliability for this node-wise median we take the interquartile range divided by the median of each node. Thus for a fixed number of surrogates we get a distribution over all nodes of values given by the interquartile ranges divided by the medians. The evolution of that distribution with increasing number of surrogates is visualized in Fig. 3.3. As one can see we could have used only half as many surrogates and would have gotten a very similar reliability of our correction. The network on which the regional Closeness is computed (bottom-right) is a ripped-off part of the global network (top-left). Thus, nodes in the regional network are connected if they are connected in the global network. The global network is a SERN with the link probability $p(\Delta) \propto \Delta^{-3.5}$. The scales in colorbars go from the minimum to the maximum value in all figures since we are only interested in relative quantities.

Link-weighted Networks For link-weighted networks an average over an ensemble of surrogates is not necessary. Similar to the proposed algorithm for unweighted networks in the previous paragraph, the following algorithm is proposed:

$A_d = B_d = W_d = 0 \quad \forall\, d \in$ rounded Δ
for $i \in nodes$ **do**
  for $k < i$ **do**
    $d \leftarrow$ rounded Δ_{ik}
    increase A_d by one
    increase B_d by w_{ik}
  end for
end for
$W \leftarrow \frac{B}{A}$.

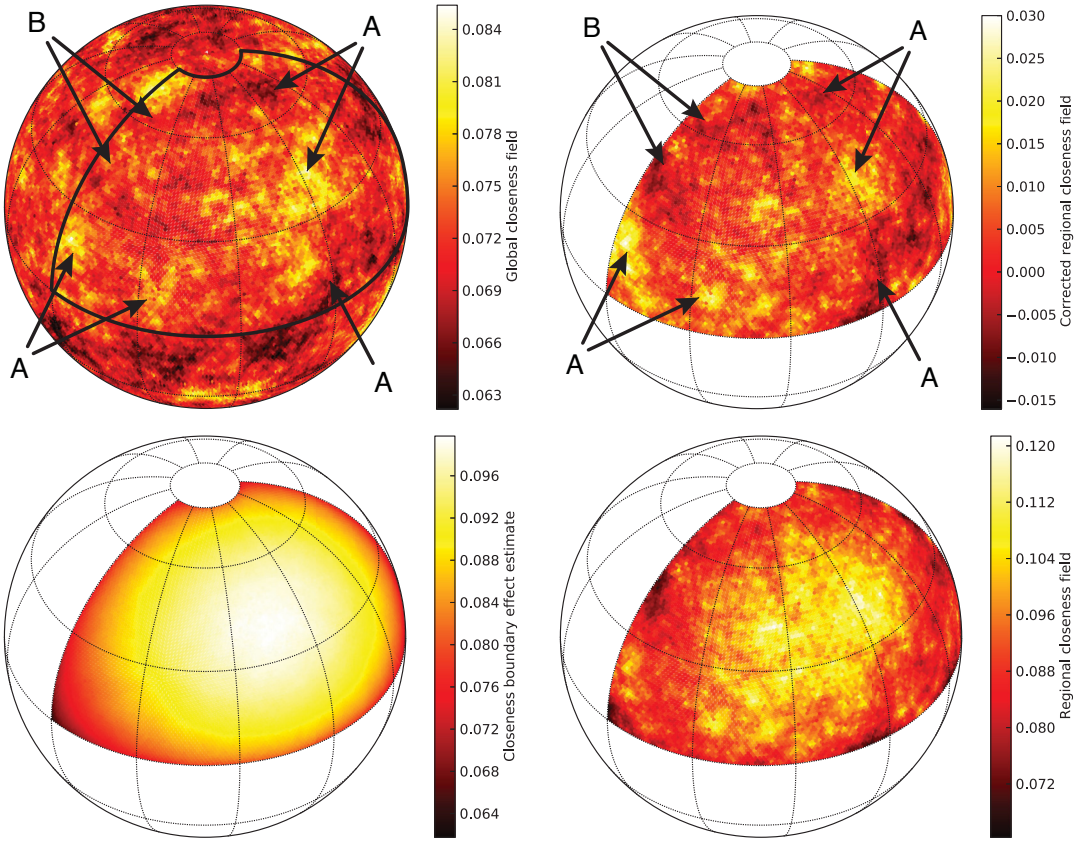

Figure 3.2.: *Top-left*: Global Closeness - Closeness Centrality of a random network on a sphere. The connection probability depends only on the spatial link length and follows a power-law with the exponent -3.5. *Top-right*: Corrected regional Closeness. Arrows point out areas of strong similarity (A) and dissimilarity (B) in the spatial patterns in the considered region. *Bottom-left*: Closeness spatial effects estimate, taken as the median from 1000 surrogates. *Bottom-right*: Regional Closeness - Closeness Centrality on a part of the same network as on the whole globe (*top-left*). Nodes in the depicted region are connected if they are connected in the global network.

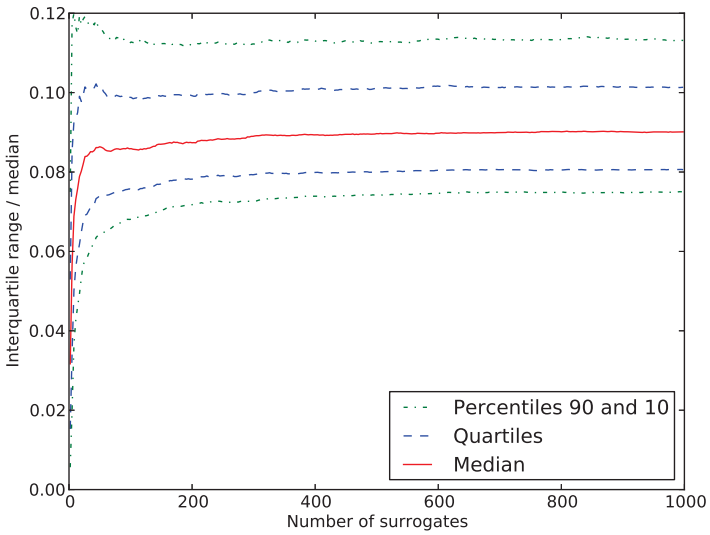

Figure 3.3.: Evolution of surrogate reliability with increasing number of surrogates for the example shown in Fig. 3.2. Shown are important quantiles of the distribution of node-wise interquartile ranges per median. After 400 surrogates reliability does not improve much further.

Instead of counting links corresponding to a certain spatial link length bin, corresponding link weights are added up. The result $W(\Delta)$ is not a probability, but the average link weight depending on the spatial link length. The single surrogate is then a weighted network with link weights according to $W(\Delta)$, and spatial effects of a given network measure are estimated by that measure on that single surrogate.

Network Measure Correction A *corrected* network measure is now calculated by subtracting the estimate of spatial effects for a certain network measure from the measure on the original network of interest. For instance, the corrected Closeness Centrality CC^- would be given by:

$$CC_i^- = CC_i^{orig} - CC_i^{surr} \,. \tag{3.19}$$

Illustratively, a comparison between CC, CC^* and CC^- is shown in Figure 3.4 for a random network on the widely used longitude-latitude grid. By construction, the link probability $p(\Delta)$ has an exponential decay and the network has 7080 nodes (3° grid). The network is embedded in a closed sphere and therefore network measures are not boundary affected. However, the common Closeness Centrality CC is biased by spatial effects due to the inhomogeneous spatial sampling of nodes (The grid is more dense close to poles). This bias is not inherent in node-splitting invariant measures (e.g. CC^* as introduced by Heitzig et al., 2012) or spatially corrected measures (e.g.

CC^- as introduced in P1). Spatially corrected network measures have the advantage that they also correct for spatial effects due to boundaries and that the measure itself is not modified (algorithms are unaffected, e.g.). However, node-splitting invariant measures are beneficial for networks with unweighted links and without boundaries if the network measure is numerically expensive. It might be numerically too costly to estimate spatial effects for an already costly link-unweighted measure like Arenas' Random Walk Betweenness (Pastor-Satorras et al., 2003) or Newman's Random Walk Betweenness (Newman, 2005) by use of an ensemble of SERNs.

In Figure 3.2 nodes are approximately homogeneously distributed and therefore network measures are biased by spatial effects only if boundaries are introduced. The common Closeness Centrality CC and the node-splitting invariant Closeness Centrality CC^* can be regarded as similar on such a grid. However, if boundaries are introduced to the network the common Closeness Centrality CC and the spatially corrected Closeness Centrality CC^- vary (compare top-right with bottom-right in Fig. 3.2). The intrinsic spatial bias of the measure due to the artificial boundary in the regional case (Fig. 3.2, bottom-right) is obviously removed in the corrected measure (Fig. 3.2, top-right).

The spatial structure of the corrected Closeness Centrality field resembles the one of the global network in the corresponding region (Fig. 3.2). Strong similarities are denoted by A and dissimilarities by B. The similarity between corrected Closeness Centrality and the uncorrected Closeness Centrality values in the corresponding region of the global network can be expressed quantitatively by a Spearman's rank correlation coefficient of 0.661, compared to a coefficient of 0.575 if the uncorrected measure is used. Figure 3.2 is just a visual example, thus if we generate an ensemble of 1000 such examples we get the distributions of correlation coefficients as shown in in Fig. 3.5. The distribution of coefficients corresponding to the corrected Closeness is not only shifted to higher similarity, but is also more narrow. The difference as well as the absolute value in similarity vary strongly with measures and link probabilities used. However similarities for corrected measures are typically higher than for regional measures.

The similarity in network measures is due to the removed spatial effects as well as to the similarity in network structure. All links that connect nodes within the specified region are the same in both networks. The global network has more links as well as links that connect nodes in the region with nodes that are not in the region; in particular, links that reach deep into the region are rare due to the power-law dependency in the link probability $p(\Delta)$. Note that the Degree is not as strongly affected by these additional links in the global in comparison to the regional case as is the path-based measure Closeness Centrality. For instance, correlation coefficients for Degree are higher.

However, due to the lack of information in the regional in comparison to the global network (border-crossing links cannot be resolved) the corrected values of the network measures of the regional network can still differ from those derived from ripped-off parts of a global network. Additional links in the global in comparison to the regional

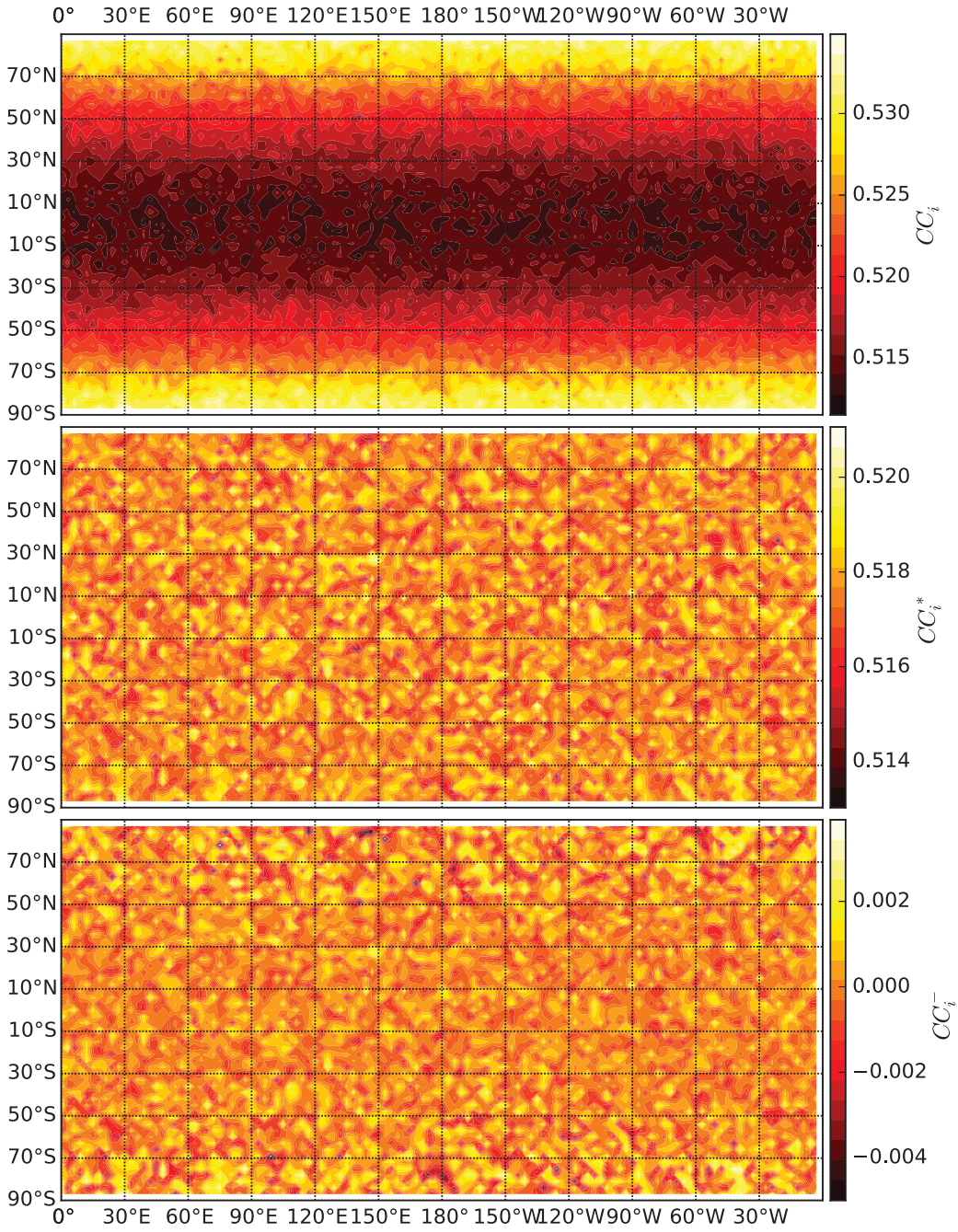

Figure 3.4.: Three Closeness Centrality versions on a spatially embedded random network (SERN) on the widely used longitude-latitude grid without boundaries (closed sphere). The link probability $p(\Delta)$ has an exponential decay and the network has 7080 nodes (3° grid). Observe that the common Closeness Centrality CC is affected by the spatial embedding, whereas CC^* and CC^- are not.

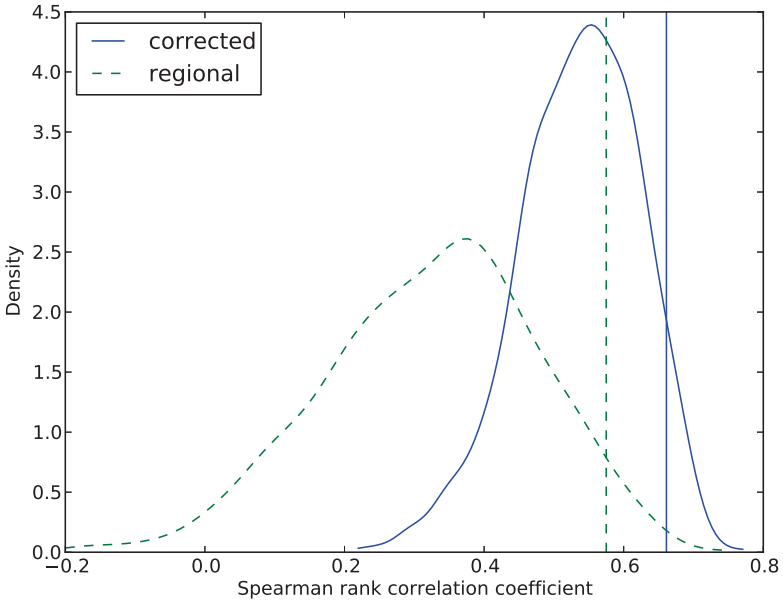

Figure 3.5.: Gaussian kernel density estimate from 1000 samples for the distribution of Spearman's rank correlation coefficients between regional and global Closenesses (green) and between corrected and global Closenesses (blue). Vertical lines correspond to the example shown in Fig. 3.2.

network can have an additional effect on network measures – especially if they are long.

These examples show the potential of the method. The described method removes spatial effects in network measures, but can of course not predict how these measures would change if adjacent regions were incorporated in the network construction. However corrected measures for different boundaries are comparable, whereas the uncorrected measures are not.

Directionality The network measure Directionality serves also as an example measure with a more sophisticated entanglement between network topology and spatial embedding. This makes spatial effects estimation more challenging, but has the advantage that such a measure also provides information regarding the orientation of links in the embedding space.

In order to correct for spatial effects in the measure Directionality, we propose the following: The frequency distribution of links $P_i(\varphi)$ for all nodes i is not only estimated on the original network, but also on the spatial surrogate. The corrected Directionality (ρ_i^-, φ_i^-) of node i is then estimated by:

$$\rho_i^- = \max_{\varphi} P_i^{orig}(\varphi) - c_i P_i^{surr}(\varphi) \tag{3.20}$$

$$\varphi_i^- = \arg\max_{\varphi} P_i^{orig}(\varphi) - c_i P_i^{surr}(\varphi)\,. \tag{3.21}$$

The constant c_i is defined as the quotient of the two corresponding average frequencies of links: $c_i = \langle P_i^{orig} \rangle / \langle P_i^{surr} \rangle$. This local re-scaling of the spatial surrogate frequency distribution of links $P_i^{surr}(\varphi)$ ensures that the Directionality (ρ_i^-, φ_i^-) is approximately the null vector if links are isotropic. ρ_i^- is therefore also a measure of anisotropy that is independent of the spatial embedding of the network.

Due to the irregular sampling of nodes in space and boundaries in the network, even an isotropic link probability can lead to preferred directions of links. In other words, in such a case the frequency distribution of links $P_i^{orig}(\varphi)$ is not necessarily flat. However, if the surrogate frequency distribution of links $P_i^{surr}(\varphi)$ is subtracted, it does become flat since the spatial surrogate is constructed with an isotropic link probability. On the other hand, if the difference of the frequency distributions has large positive maxima at certain angles, then these are due to more links into the direction of these angles than what is expected from spatial effects.

Since Directionality is, in contrast to for instance path-based measures like Closeness Centrality, a topologically local measure, this spatial effects estimation scheme is valid for unweighted as well as for link-weighted spatial networks. The difference stems from the different estimation of frequency distributions $P(\varphi)$: For unweighted networks $P(\varphi)$ is estimated by Equation (3.17) and for link-weighted networks by Equation (3.18).

Two examples of the link-weighted Directionality with spatial effects estimation are given in Figure 3.6 and 3.7. Both examples are take for a node i that has 1000 links, but the spatial embedding as well as link probabilities are different each time. Example one (see Fig. 3.6) has 1000 homogeneously sampled nodes in a quadratic region of size 1 by 1. The node i for which the Directionality is estimated is in the center at $(0.5, 0.5)$. The Voronoi cell for each node's position is colored according to the link weight for the link to node i (Top-left). Similarly, the surrogate is depicted (Top-center), as well as the difference between these two fields (Top-right). The corresponding frequency distributions $P_i^{orig}(\varphi)$ (Bottom-left), $P_i^{surr}(\varphi)$ (Bottom-center) and the adjusted difference $P_i^{orig}(\varphi) - c_i P_i^{surr}(\varphi)$ (Bottom-right) are shown in the bottom row. The underlying anisotropy in the link probability of the original network (Left-column) is almost unperturbed by the node sampling or boundaries in the network. Hence, already the uncorrected frequency distribution $P_i^{orig}(\varphi)$ (Bottom-left) features the more or less correct direction of $\frac{\pi}{2}$. However, this is not necessarily the case as one can observe in example two (see Fig. 3.7). This example is constructed similarly as the first, but the node i is now in the top-right corner at $(0.9, 0.9)$, the link probability is decreasing slower[1] and the node sampling in the embedding space is inhomogeneous. This leads to strong spatial effects that severely bias the uncorrected Directionality. Instead of $\frac{\pi}{2}$, φ_i would be less than $\frac{\pi}{4}$ (see the location of the maximum in the Bottom-left panel of Fig. 3.7). Nevertheless, the corrected Directionality reveals the correct orientation; $\varphi_i^- \approx \frac{\pi}{2}$ (see the red line in the Bottom-right panel of Fig. 3.7).

[1] This leads to an increase in *boundary effects* due to more longer links, or as in this case, higher link weights also for longer links (see P1).

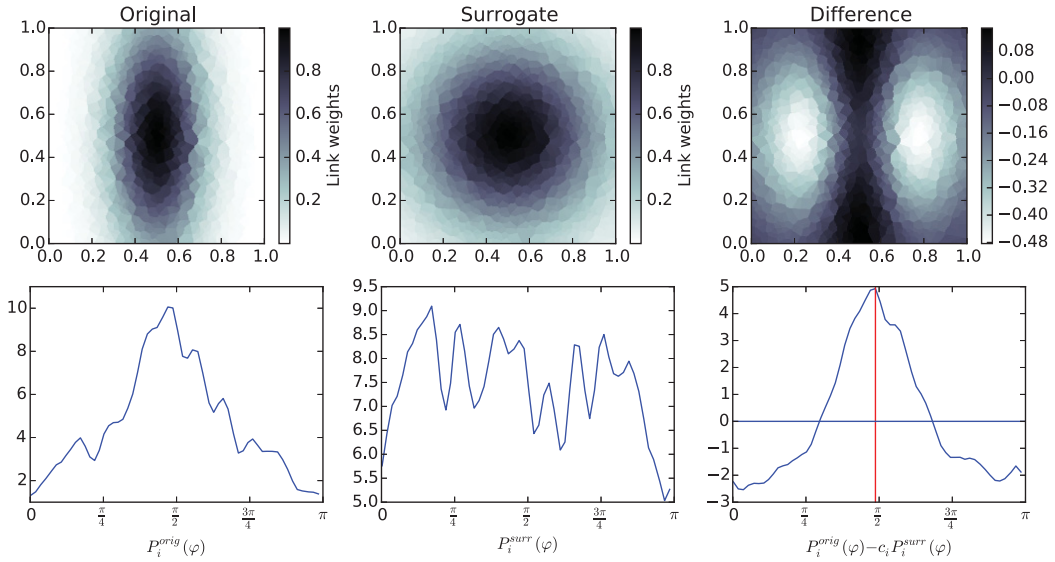

Figure 3.6.: Link-weighted Directionality example for a node i in the center of a quadratic region with 1000 nodes. Voronoi cells for each node are colored according to the link weight for the link to node i (Top-left). Similarly the surrogate is depicted (Top-center) and the difference between these two fields (Top-right). The corresponding frequency distributions $P_i^{orig}(\varphi)$ (Bottom-left), $P_i^{surr}(\varphi)$ (Bottom-center) and the adjusted difference $P_i^{orig}(\varphi) - c_i P_i^{surr}(\varphi)$ (Bottom-right) are shown in the bottom row. The corrected Directionality features the correct orientation; $\varphi_i^- \approx \frac{\pi}{2}$ (see the red line in the Bottom-right panel). So does the uncorrected orientation $\varphi_i \approx \frac{\pi}{2}$ since spatial effects are negligible.

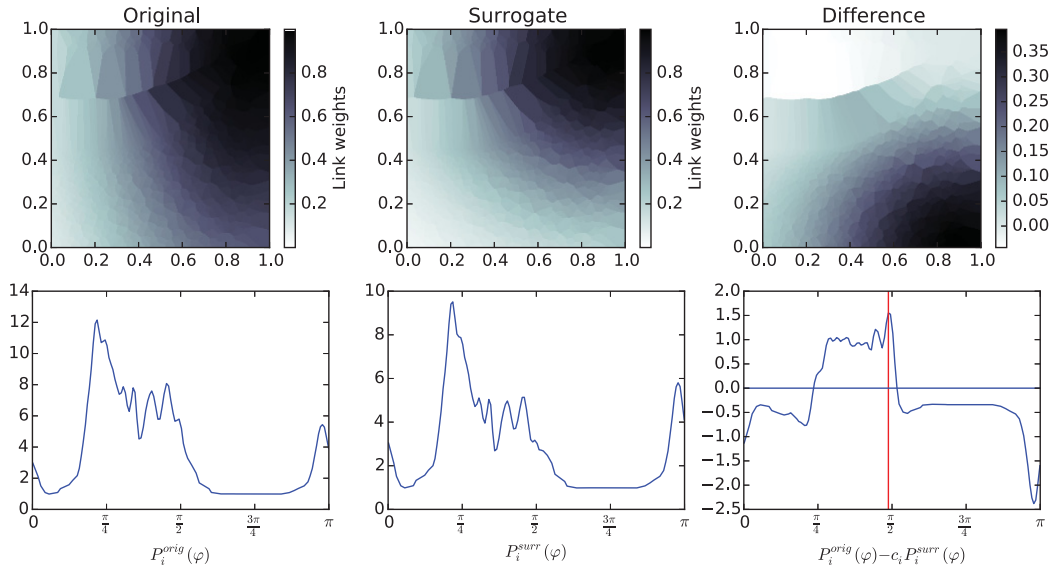

Figure 3.7.: Same as Fig. 3.6 but with the node i not in the center of the region but in the top-right corner $(0.9, 0.9)$. Additionally, the link probability is more long-ranged and the spatial sampling of nodes is inhomogeneous. The corrected Directionality features the correct orientation; $\varphi_i^- \approx \frac{\pi}{2}$ (see the red line in the Bottom-right panel), while the uncorrected Directionality is biased by spatial effects; $\varphi_i < \frac{\pi}{4}$ (see the location of the maximum in the Bottom-left panel).

3.4. Visual Analytics

3.4.1. Introduction

Visual Analytics is an emerging approach to study large complex systems in various fields of research, for instance climate science (Graves et al., 2012; Santos et al., 2013, e.g.), molecular biology (Bajaj et al., 2004, e.g.), genetics (Xia et al., 2014, e.g.), computer science (Sarma et al., 2009; Hawes et al., 2013, e.g.), or even for Climate Networks (Tominski et al., 2009, e.g.). Its aim is to provide an interactive visualization of complex systems for knowledge discovery. A key idea is to have a visualization system that extracts interactively certain features of a complex system in order to make it visually interpretable. In the following a new feature extraction technique is presented. It displays teleconnections in Climate Networks.

3.4.2. Coarse Graining and Teleconnections

So far, empirical orthogonal functions (EOFs), which are derived from principal component analysis of covariance matrices, are commonly used for a spatial analysis of teleconnections in climatological data (Kawamura, 1994; Deser, 2000; Ding and Wang, 2005, e.g.). While certainly very useful in many situations, they carry certain caveats in such analyses: First, if the data are not normally distributed the corresponding EOFs will in general, while uncorrelated, not be statistically independent (Monahan et al., 2009). Second, even if they are independent, EOFs do not necessarily uniquely correspond to climatological mechanisms (Dommenget and Latif, 2002). Third, and maybe most importantly, analyses based on the covariance matrix will only be able to capture linear dependencies. This might be considered insufficient in view of the strong nonlinearities involved in climatic interactions. Climate Networks can be considered as a complementary approach to study spatial patterns of climatic interrelations, which do not suffer from these statistical problems if derived from a nonlinear similarity measure. Furthermore, since teleconnections are not directly represented as links in EOFs they have to be deduced from the spatial patterns. Although this might be possible for simple teleconnection structures, it becomes challenging for more complicated ones.

Nonetheless, the common way of Climate Network construction is not suitable for the investigation of teleconnections as well. There, traditionally a pair-wise similarity analysis between all pairs of time series is performed, for instance by use of Pearson's correlation coefficient (Donges et al., 2009a; Tsonis et al., 2006b, e.g.). However, Climate Networks are spatially embedded networks and the similarity between time series is strongly dependent on their spatial distance (P1): Two time series that are spatially close to each other are likely to be more similar than two time series which are far away from each other in space. By focusing only on strong similarities as in most Climate Network studies, networks have essentially only short links, which led to the investigation of paths in Climate Networks (Donges et al., 2009b, e.g.).

Here an approach is proposed that groups all time series by similarity into clusters. A related idea was also pursued by Hlinka et al., 2014. The specific clustering scheme

employed here, typically provides spatially connected clusters due to the distance dependence of the similarities in climate systems. In other words, these clusters are localized regions of high resemblance according to the dynamics of the corresponding time series. Each cluster will in our approach be represented by a single time series and only the similarity structure between these representatives will be explored. By doing so we do not only reduce the dimensionality of the network, but we more importantly constructed a Climate Network that is reduced to its teleconnections. We will refer to these networks as *Teleconnection Climate Networks* (TCN).

3.4.3. Teleconnection Climate Networks

In order to group time series by similarity the standard fast greedy hierarchical agglomerative *complete linkage clustering* (Defays, 1977) is performed. This clustering is done in a metric space with dissimilarities between time series as distances. Depending on the measure of similarity used, different distances between time series are possible. The right choice depends on the research question. In general the distance d_{ij} between two time series x_i and x_j is given by the distance function D that converts the similarity s_{ij} between x_i and x_j into a metric distance:

$$d_{ij} = D(s_{ij})\,. \tag{3.22}$$

The intention is to group time series into clusters in such a way that all similarities between time series within a given cluster are statistically significant. This is achieved by use of the *complete linkage clustering* scheme that is also known as *farthest neighbor clustering*. The distance measure between two clusters U and V is in this scheme defined as

$$D^{clust}(U,V) := \max_{x_i \in U, x_j \in V} D(s_{ij}) = \max_{x_i \in U, x_j \in V} d_{ij}\,. \tag{3.23}$$

The resulting dendrogram then has to be cut at the distance d_{crit} that corresponds to the significance threshold of all pair-wise similarities. For instance, for a 98% significance level d_{crit} is given by

$$d_{crit} = D(\max_{ij} {}_{up}T^{0.98}(s_{ij}))\,. \tag{3.24}$$

This yields the maximum number of partitions of the set of time series such that for any two clusters U and V holds $D(U,V) \geq d_{crit}$, which is the same as the minimum number of partitions such that for any two time series $x_i, x_k \in U$ in any given cluster U we have $d_{ik} < d_{crit}$. This clustering method does not only assure that all time series within a cluster are significantly correlated when cutting the dendrogram at d_{crit}, but also avoids the *chaining phenomenon* of the *single linkage clustering* where a set of time series might form a cluster although only a few time series are actually close to each other in the embedding space (Everitt et al., 2001). The clustering reduces the dimensionality of the problem by merging dynamically similar time series

into clusters, which will serve as nodes for the *Teleconnection Climate Networks* TCN that will be constructed in the following.

More specifically, a TCN node is represented by a single time series from the corresponding cluster. Although there are clustering schemes, such as the *k-means clustering* (MacQueen et al., 1967), that suggest a certain member of a cluster as a representative, the in this study anticipated *complete linkage clustering* does not. Also, since cluster sizes vary, special care has to be taken when choosing a representative time series for a cluster. For instance, the point-wise mean of all time series within a cluster would be influenced by the size of the cluster. Instead the time series with the highest average similarity to all other time series within that cluster is sensible choice for a representative of that cluster. This also has the advantage that the representative time series retain the original variabilities.

The TCN is then constructed by computing the similarities s_{ij} for all pairs of representative time series. With M being the number of representative time series, where $M \ll N$, the subset of nodes $i \in 1, \ldots, M$ denotes the set of TCN nodes. The TCN adjacency matrix A is then a MxM matrix. For instance, with the lower significance threshold ${}_{lo}T^{0.98}(s_{ij})$, and the upper threshold ${}_{up}T^{0.98}(s_{ij})$, the link-weighted adjacency matrix A is given by

$$A_{ij} = \begin{cases} 0 & \text{if} \quad {}_{lo}T^{0.98}(s_{ij}) \leq s_{ij} \leq {}_{up}T^{0.98}(s_{ij})\,, \\ s_{ij} & \text{otherwise}\,. \end{cases} \tag{3.25}$$

Part II.

Applications

This part presents several applications of previously presented methodologies and is based on the associated publications P1 to C3.

Chapter 4.

Event Synchronization Climate Networks in Studies of Extreme Rainfall

Abstract

The analysis of spatial patterns of co-variability of extreme rainfall is challenging because traditional techniques based on principal component analysis of the covariance matrix only capture the first two statistical moments of the data distribution, and are thus not suitable to analyze the behavior in the tails of the respective distributions. Here, we describe an alternative to these techniques which is based on the combination of a non-linear synchronization measure and complex network theory. This approach allows to derive spatial patterns encoding the co-variability of extreme rainfall at different locations. By introducing suitable network measures, the methodology can be used to perform climatological analysis, but also for statistical prediction of extreme rainfall events. We introduce the methodological framework and present applications to high-spatiotemporal resolution rainfall data (TRMM 3B42) over South America.

4.1. Introduction

The analysis of the spatial structure of co-variability of climatic time series at different locations forms an integral part of meteorological and climatological research. Traditional techniques in this context are based on principal component analysis (PCA) of the covariance matrix of the dataset under consideration. By construction, such approaches only capture the first two statistical moments of the distributions of the individual time series, and the resulting empirical orthogonal functions (EOFs) do thus not describe the behavior of extreme events. By combining a non-linear synchronization measure with complex network theory, we introduce a methodology that can fill this gap and show how it can be applied for climatological analysis, but also for statistical prediction of extreme rainfall events.

In the recent past, so-called *climate networks* have attracted great attention as tools to analyze spatial patterns of climatic co-variability, complementarily to traditional PCA-based techniques (e.g. Tsonis and Roebber, 2004c; Tsonis and Swanson, 2008; Donges et al., 2009b; Donges et al., 2009a; Donges et al., 2011; Gozolchiani et

al., 2011; Steinhaeuser et al., 2012a; Van Der Mheen et al., 2013; Ludescher et al., 2013). Here, we show how these approaches can be extended to capture the dynamical characteristics of extreme events. The key idea of the methodology that shall be presented in the following sections is to identify rainfall time series measured at different locations with network nodes, and represent strong synchronizations of extreme events in these time series by network links connecting the respective nodes. The climatological mechanisms driving the synchronization and propagation of extreme rainfall events are assumed to be encoded in the topology of the resulting climate network. Different aspects of this topology can be quantified by means of suitable network measures, and upon providing climatological interpretations of these network measures, we will show that the spatial patterns they exhibit reveal the underlying climatological mechanisms (Boers et al., 2013). Furthermore, using directed and weighted networks, we will show how this approach can be used for statistical prediction of extreme events (Boers et al., 2014a), given that the synchronization patterns are sufficiently pronounced.

While we restrict ourselves to present its application to satellite-derived rainfall data, the methodological framework is more general and can in principle be applied to analyze collective synchronization patterns of extreme events in many types of complex systems. The methodology should be considered as a general data exploration tool that can provide the basis for building scientific hypotheses on the mechanisms underlying the synchronization of extreme events in large, interactive systems.

4.2. Climatic Setting

The monsoon season in South America from December to February (DJF) is characterized by a southward shift of the Intertropical Convergence Zone (ITCZ) and by an amplification of the trade winds due to the differential heating between ocean and land (Zhou and Lau, 1998) (Fig. 4.1). These low-level winds transport moist air from the tropical Atlantic ocean toward the tropical parts of the continent, where they cause abundant rainfall. Substantial fractions of this precipitation are recycled back to the atmosphere by evapotranspiration, and the winds carry the water vapor farther west across the Amazon Basin towards the Andes. There, the shape of the mountain range forces the winds southward towards the subtropics (Vera et al., 2006; Marengo et al., 2012). The specific exit regions of this moisture flow vary considerably from the central Argentinean plains to southeastern Brazil. These variations are associated with frontal systems approaching from the South, which are triggered by Rossby waves of the polar jet streams (Siqueira and Machado, 2004; Carvalho et al., 2010). A dominant southward component of the flow leads to the South American Low-Level Jet (SALLJ) east of the Andes (Marengo et al., 2004), which conveys large amounts of moisture from the tropics to southeastern South America (SESA). The occurrence of this wind system is associated with huge thunderstorms (so-called Mesoscale Convective Systems (Durkee et al., 2009)) in this region (Salio et al., 2007a). On the other hand, if the flow to the subtropics is directed mainly eastwards, it leads

to the establishment of the South Atlantic Convergence Zone (SACZ), a convective band that extends from the central Amazon Basin to southeastern Brazil (SEBRA) (Carvalho et al., 2004b). The oscillation between these two circulation regimes leads to the so-called South American rainfall dipole (cf. P2), and constitutes the dominant mode of intraseasonal variability of the monsoon (Nogués-Paegle and Mo, 1997).

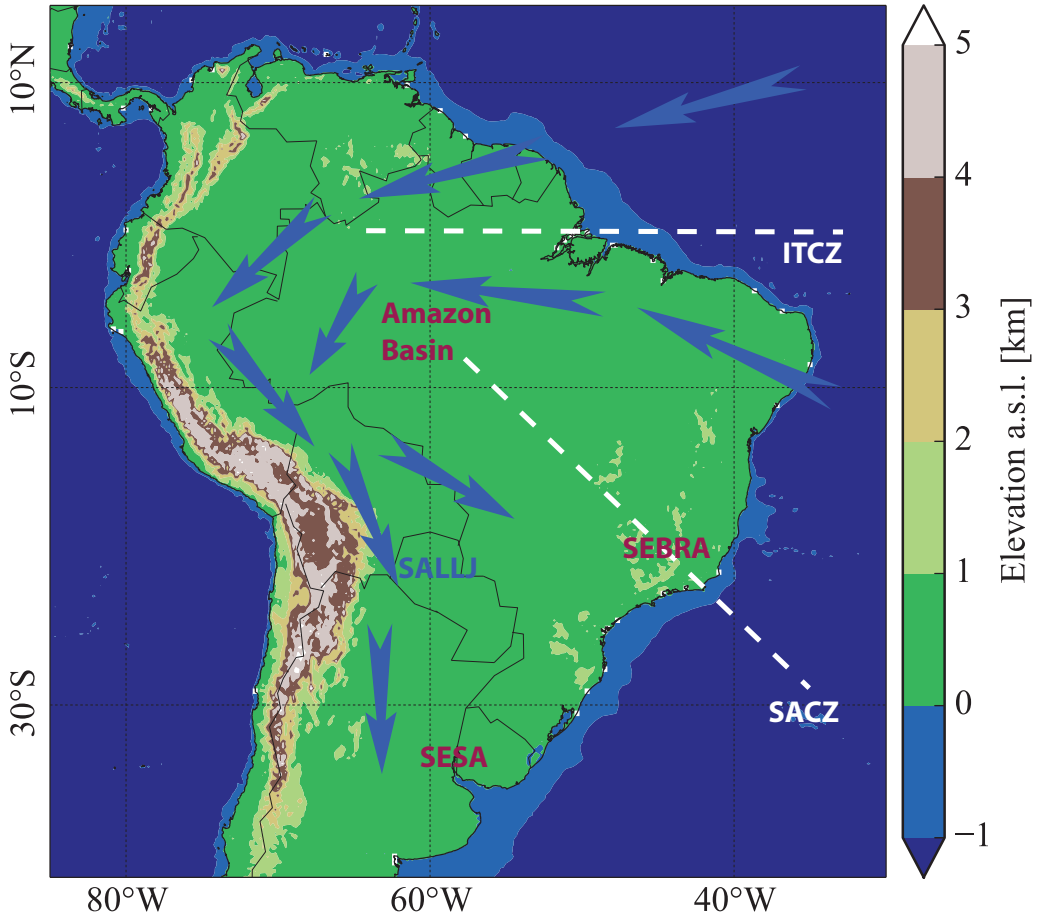

Figure 4.1.: Topography of South America and key features of the South American monsoon system, including the main low-level wind directions, the Intertropical Convergence Zone (ITCZ), the South Atlantic Convergence Zone (SACZ), and the South American Low-Level Jet (SALLJ). The geographical regions southeastern South America (SESA), southeastern Brazil (SEBRA), and Amazon Basin are referred to in the main text.

4.3. Data and Methods

Data We employ satellite-derived rainfall data from the Tropical Rainfall Measurement Mission (TRMM 3B42 V7, (Huffman et al., 2007)) with 3-hourly temporal and $0.25° \times 0.25°$ spatial resolutions, resulting in $N = 48,400$ time series with values measures in mmh^{-1}. Daily (3-hourly) extreme events are defined *locally* as points in time for which the corresponding rainfall rate is above the 90th (99th) percentile for the corresponding time series, confined to the monsoon seasons (DJF) from 1998 to 2012.

Event Synchronization The non-linear synchronization measure we employ is called Event Synchronization and was first introduced by Quiroga et al., 2002. It quantifies the synchronicity between events in two given time series x_i and x_j by counting the number of events that can be uniquely associated with each other within a prescribed maximum delay, while taking into account their temporal ordering: Consider two event series $\{e_i^\mu\}_{1\leq\mu\leq l}$ and $\{e_j^\nu\}_{1\leq\nu\leq l}$ containing l events, where e_i^μ denotes the time index of the μ-th event observed at grid point i. In order to decide if two events e_i^μ and e_j^ν with $e_i^\mu > e_j^\nu$ can be assigned to each other uniquely, we first compute the waiting time $d_{ij}^{\mu,\nu} := e_i^\mu - e_j^\nu$, and then define the *dynamical delay*:

$$\tau_{ij}^{\mu\nu} = \frac{1}{2}\min\{d_{ii}^{\mu,\mu-1}, d_{ii}^{\mu,\mu+1}, d_{jj}^{\nu,\nu-1}, d_{jj}^{\nu,\nu+1}\} \tag{4.1}$$

We further introduce a maximum delay τ_{max} which shall serve as an upper bound for the dynamical delay. If then $0 < d_{ij}^{\mu,\nu} \leq \tau_{ij}^{\mu\nu}$ and $d_{ij}^{\mu,\nu} \leq \tau_{max}$, we count this as a directed synchronization from j to i:

$$S_{ij}^{\mu\nu} = \begin{cases} 1 & \text{if} \quad 0 < d_{ij}^{\mu,\nu} \leq \tau_{ij}^{\mu\nu} \quad \text{and} \quad d_{ij}^{\mu,\nu} \leq \tau_{max}, \\ 0 & \text{else.} \end{cases} \tag{4.2}$$

Directed Event Synchronization (Boers et al., 2014a) from j to i is then given by

$$ES_{ij}^{dir} := \sum_{\mu\nu} S_{ij}^{\mu\nu}\,. \tag{4.3}$$

A symmetric version of this measure can be obtained by also counting events at the very same time as synchronous and taking the absolute value of the dynamical delay in equation (4.2),

$$\overline{S}_{ij}^{\mu\nu} = \begin{cases} 1 & \text{if} \quad |d_{ij}^{\mu,\nu}| \leq \tau_{ij}^{\mu\nu} \quad \text{and} \quad d_{ij}^{\mu,\nu} \leq \tau_{max}, \\ 0 & \text{else,} \end{cases} \tag{4.4}$$

and computing the corresponding sum:

$$ES_{ij}^{sym} := \sum_{\mu\nu} \overline{S}_{ij}^{\mu\nu}\,. \tag{4.5}$$

A major advantage of this measure is that it allows for a *dynamical delay* between events in the original time series x_i and x_j. In classical lead-lag analysis (using, e.g., Pearson's Correlation Coefficient) this is not the case, since it only provides one single delay between the two time series, namely the time window by which the time series x_i and x_j are shifted against each other. Since the various climatological mechanisms underlying the interrelations between time series measured at different locations cannot be assumed to operate on one single time scale, the temporal homogeneity assumed by a classical lead-lag analysis is not justified. Furthermore, the identification of the correct lead (or lag) is not a well-defined problem, as there may be several maxima of the correlation value over the range of leads or lags.

Network construction In the following, the notations ES for the measure or **ES** for the corresponding similarity matrix will be used if a statement applies to both versions of Event Synchronization. From the matrix **ES**, we derive networks by representing its strongest entries by network links. It has to be assured that these values are statistically significant. For this purpose, we construct $10,000$ surrogates of event time series preserving the block structure of subsequent events by uniformly randomly distributing the original blocks of subsequent events, and compute ES for all possible pairs. From the resulting histogram of values, we obtain the threshold $T^{0.95}$ corresponding to the 5% confidence level. The link density of the network is then chosen such that the smallest entry of **ES** that is represented by a network link is above $T^{0.95}$. In terms of the adjacency matrix **A**, this is captured by

$$A_{ij} = \begin{cases} ES_{ij} & \text{if} \quad ES_{ij} > T^{0.95}\,, \\ 0 & \text{else}\,. \end{cases} \tag{4.6}$$

Note that the values of ES have been assigned to the links as weights. Of course, one can also set the corresponding entries of **A** to 1 in order to obtain an unweighted network. In case of ES^{sym}, the corresponding network will be undirected, while for ES^{dir}, it will be directed.

Network measures On undirected and unweighted networks, we will consider four different network measures: First, we consider betweenness centrality (BC), which is defined on the basis of shortest network paths, i.e. the shortest sequences of links connecting two nodes:

$$BC_i := \frac{\sum_{l<k\neq i} \sigma_{kl}(i)}{\sum_{l<k\neq i} \sigma_{kl}}\,, \tag{4.7}$$

where σ_{kl} denotes the total number of shortest network paths between nodes k and l and $\sigma_{kl}(i)$ the number of shortest network paths between k and l which pass through node i. Since BC is a non-local centrality measure, we expect BC to exhibit high values in regions which are important for the long-ranged, directed propagation of extreme events.

Second, we are interested in the mean geographical distance (MD, Boers et al., 2013) of links at each node:

$$MD_i := \frac{1}{DG_i} \sum_{j=1}^{N} A_{ij} dist(i,j) \tag{4.8}$$

where $dist(i,j)$ denotes the great-circle distance between the grid points corresponding to the nodes i and j. MD should show high values in regions where extreme events occur synchronously with extreme events at remote locations, and thus quantifies similar aspects of the topology as BC, although not based on network paths. Therefore, to confirm our interpretation of BC, we would expect this measure to have a similar spatial distribution as BC.

Third, we employ the clustering coefficient, defined as the fraction of neighbors of a given node that are themselves connected:

$$LC_i := \frac{\sum_{j<k} A_{ij} A_{jk} A_{ik}}{\sum_{j<k} A_{ij} A_{ik}} \tag{4.9}$$

LC measures complementary aspects of the topology as compared to the previous two measures, and should be high in regions where extreme events exhibit large spatial coherence, as for example due to large thunderstorms.

Furthermore, we introduce a combination of these measures, called long-ranged directedness (LD, Boers et al., 2013). For this purpose, we calculate the normalized ranks of BC, LC, and MD, denoted by NRBC, NRLC, and NRMD, respectively, and put

$$LD_i := \frac{1}{2} NRBC_i + \frac{1}{2} NRMD_i - NRLC_i \,. \tag{4.10}$$

The prefactors in this definition are motivated by the fact that BC and MD are expected to quantify similar aspects of the network topology, while LC was introduced to estimate complementary properties of the network. We thus take the mean of the normalized ranks of BC and MD, and subtract the normalized rank of LC. High values of LD should indicate regions which are important for the long-ranged propagation of extreme events, while low values should indicate regions where extreme events strongly cluster, but do not propagate over long spatial distances.

On directed and weighted networks, we will consider the well-known in- and out-strength, defined as

$$SG_i^{in} := \sum_{j=1}^{N} A_{ij} \quad \text{and} \quad SG_i^{out} := \sum_{j=1}^{N} A_{ji} \tag{4.11}$$

On the basis of these measures, we define the measure *network divergence* (ΔSG, Boers et al., 2014a) as the difference of in-strength and out-strength at each grid cell:

$$\Delta SG_i := SG_i^{in} - SG_i^{out} \,. \tag{4.12}$$

This measure can be used to identify source and sink regions of extreme events on a continental scale. In order to investigate where extreme events originating from a given source region go to, we define the strength out of a geographical region R into a node i as

$$SG_i^{in}(R) := \frac{1}{|R|} \sum_{j \in R} A_{ij} \,, \tag{4.13}$$

where $|R|$ denotes the number of grid cells contained in R.

4.4. Results and discussion

We will first use undirected and unweighted networks to show that the methodology introduced above reveals climatic features which are consistent with the scientific understanding of the South American monsoon system. This is mainly intended as a proof of concept. Thereafter we will show that, using directed and weighted networks, the approach can in certain situations be used to predict extreme events.

Climatic analysis of extreme rainfall We compute the measures BC, MD, LC, and LD for undirected and unweighted networks with a prescribed link density of 2%. These networks are derived from ES^{sym} computed for daily events above the 90th percentile.

BC and MD show a very similar spatial distribution, with high values over the ITCZ, the Amazon Basin, as well as at the eastern slopes of the Andes along the entire mountain range (Fig. 4.2A,B). These regions are in fact crucial for the large-scale distribution of extreme events over the South American continent: The low-level trade winds drive them from the tropical Atlantic towards the continent (Zhou and Lau, 1998), and upon a cascade of rainfall and evapotranspiration over the Amazon Basin (Eltahir and Bras Rafael L., 1993), the winds force the moist air against the Andean slopes, leading to so-called orographic rainfall (Bookhagen and Strecker, 2008). The positioning of the branch of high BC and MD values from the western Amazon Basin along the Andean slopes towards the subtropics corresponds to the climatological location of the SALLJ, which provides the moisture necessary for extreme rainfall events (Marengo et al., 2004).

In contrast, the only regions over the mainland that exhibit high values of LC (Fig. 4.2C) are SESA, where some of the largest thunderstorms on earth occur (Zipser et al., 2006), and the eastern coastal regions of the continent, which are exposed to the landfall of so-called squall lines (Cohen et al., 1995).

By construction, LD shows high values where BC and MD both show high values, and particularly low values in most parts of SESA, where LC is high. However,

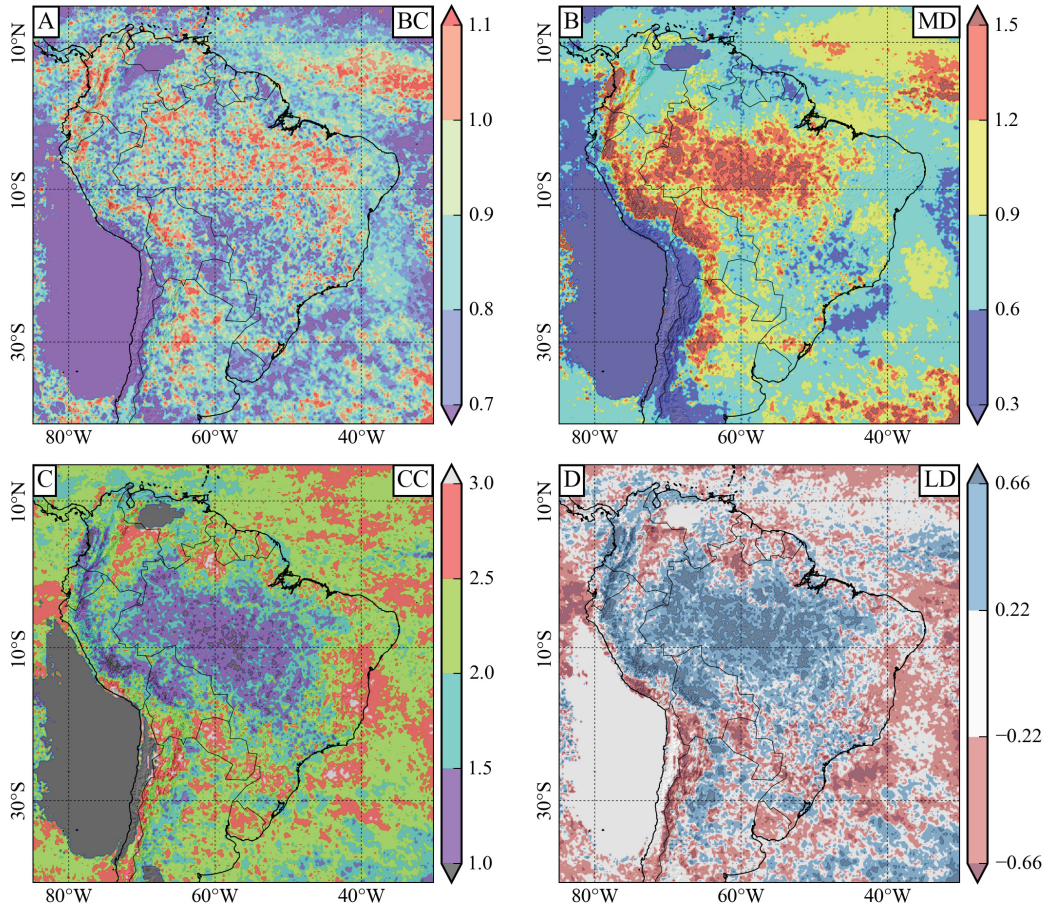

Figure 4.2.: Network measures for undirected and unweighted networks encoding the synchronization structure of daily rainfall events above the 90th percentile of the monsoon season (DJF). A. Betweenness centrality (BC) B. Mean geographical distance (MD) C. Clustering coefficient (LC) D. Long-ranged directedness

LD is also relatively high in SEBRA, concisely corresponding to the climatological position of the SACZ (Carvalho et al., 2002; Carvalho et al., 2004b). These high LD values indicate the highly dynamical character of extreme events associated with this convergence zone.

The spatial distributions of the four measures BC, MD, LC, and LD hence reveal these important climatological features, and our interpretation of these network measures is thus consistent with the understanding of the South American monsoon system (Boers et al., 2013).

Prediction of extreme rainfall We construct directed and weighted networks on the basis of ES^{dir} (cf. Eq. 4.6), computed for 3-hourly events above the 99th percentile. Network divergence ΔSG of the resulting network exhibits negative values (i.e. source regions for extreme events) over the ITCZ and the Amazon Basin, followed by pronounced positive values (i.e., sinks of extreme events) at the eastern slopes of the Andes (Fig. 4.3A). Surprisingly, SESA, which was described as one of the exit regions of the low-level flow from the tropics, is a pronounced source region of extreme rainfall. In order to reveal where these events subsequently propagate, we compute the strength out of the spatial box denoted by SESA in Fig. 4.3, and infer that, while some extreme events propagate northeastward, there also exits a concise signature of targets extending from SESA to the eastern slopes of the Central Andes in Bolivia. Thus, extreme rainfall in the Bolivian Andes should be predictable from preceding events in SESA. In (Boers et al., 2014a), the authors revealed the interplay of frontal systems approaching from the South, the Andean orography, and the low-level moisture flow from the tropics as responsible climatic mechanism. This interplay leads to the opening of a wind channel conveying warm and moist air from the western Amazon Basin to SESA. These air masses collide with cold air in the aftermath of the frontal system, leading to abundant precipitation. The typical propagation trajectory of the associated rainfall clusters is dictated by the northward movement of the frontal system and its alignment with respect to the Andean mountain range. Based on these insights, a simple forecast rule is formulated in (Boers et al., 2014a), which predicts 60% (90% during positive phases of the El Niño Southern Oscillation) of extreme rainfall events at the eastern slopes of the Central Andes.

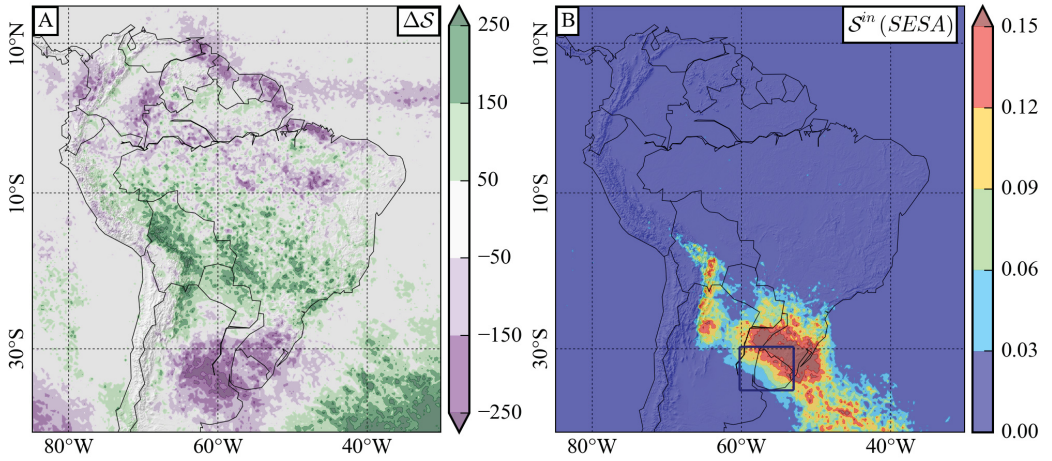

Figure 4.3.: Network measures for directed and weighted networks encoding the temporally resolved synchronization structure of 3-hourly rainfall events above the 99th percentile of the monsoon season (DJF). A. Network Divergence (ΔSG) B. Strength out of SESA ($SG^{in}(SESA)$), where SESA is defined as the spatial box extending from 35°S to 30°S and from 60°W to 53°W.

4.5. Conclusion

In this chapter, we showed how complex networks can be employed to reveal spatial patterns encoding the dynamical synchronization of extreme rainfall events, and how this can be used for climatic analysis as well as to estimate the predictability of extreme rainfall. We constructed networks on the basis of synchronization of extreme rainfall events in South America and showed that combining the network measures betweenness centrality, mean geographical distance, and clustering allowed to identify the main features of the South American monsoon system. Furthermore, we showed that a directed network approach can be applied to reveal typical propagation patterns of extreme rainfall events. Specifically, a pathway from southeastern South America to the Central Andes was revealed, which provides the basis for predicting extreme events in the Central Andes.

Chapter 5.

The South American Rainfall Dipole

A Complex Network Analysis of Extreme Events

Abstract

Intraseasonal rainfall variability of the South American monsoon system is characterized by a pronounced dipole between southeastern South America and southeastern Brazil. Here, we analyze the dynamical properties of extreme rainfall events associated with this dipole by combining a non-linear synchronization measure with complex networks. We make the following main observations: i) Our approach reveals the dominant synchronization pathways of extreme events for the two dipole phases. ii) While extreme rainfall synchronization in the tropics is directly driven by the trade winds and their deflection by the Andes mountains, extreme rainfall propagation in the subtropics is mainly dictated by frontal systems. iii) The well-known rainfall dipole is in fact only the most prominent mode of an oscillatory pattern that extends over the entire continent. This provides further evidence that the influence of Rossby waves, which cause frontal systems over South America and impact large-scale circulation patterns, extends beyond the equator.

5.1. Introduction

We aim at an improved understanding of the characteristics and origins of extreme rainfall (above the 90th percentile of wet days) in the two most densely populated areas in South America, namely southeastern South America (SESA) around Buenos Aires between 30°S and 35°S and southeastern Brazil (SEBRA) including Saõ Paulo and Rio de Janeiro between between 18°S and 23°S (see Fig. 5.1).

Large parts of the economies in these two regions depend on agriculture. They are thus heavily reliant on continuous water supply for irrigation and energy generation, but also particularly vulnerable to damages caused by extreme rainfall and associated floodings and landslides (Berbery et al., 2002; Carvalho et al., 2002; Barros et al., 2006; Marengo et al., 2013).

In large parts of South America, rainfall during the monsoon season (December to February, DJF) depends on atmospheric low-level moisture inflow from the tropical Atlantic Ocean to the Amazon Basin provided by the trade winds (Zhou and Lau, 1998). After crossing the Amazon Basin, the low-level winds are blocked by the Andes

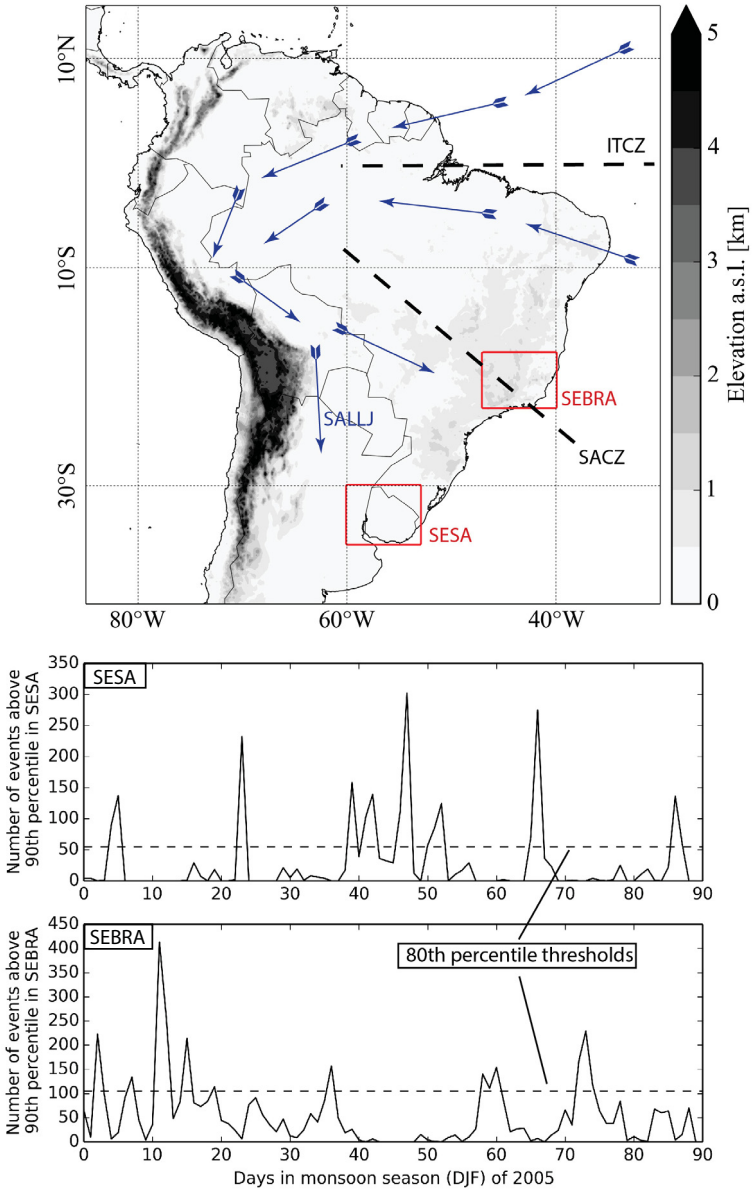

Figure 5.1.: *Top*: Topography of South America and key features of the region, including typical wind directions (blue arrows) and the South American Low Level Jet (SALLJ). The climatological positions of the Intertropical Convergence Zone (ITCZ) and the South Atlantic Convergence Zone (SACZ) are shown by dashed black lines, while the two study areas SESA (30°S to 35°S and 60°W to 53°W) and SEBRA (18°S to 23°S and 47°W to 40°W) are indicated by red boxes. *Bottom*: The time series of the number of extreme events in SESA (top) and SEBRA (bottom) for the monsoon season (DJF) of 2005. The 80th percentile thresholds used to define SESA and SEBRA phases are indicated by horizontal dashed lines.

mountains and channeled southwards, causing pronounced orographic rainfall peaks at the eastern slopes of the northern central Andes (Bookhagen and Strecker, 2008, e.g.). There exist two different regimes for the direction and strength of the subsequent low-level flow: i) If the flow has a strong southward component, it establishes the South American Low-Level Jet (SALLJ, (Marengo et al., 2004)), transporting large amounts of moisture to northern Argentina and SESA. This regime is associated with enhanced rainfall (Liebmann et al., 2004), and in particular with the formation of mesoscale convective systems (MCS, (Salio et al., 2007b; Durkee et al., 2009; Boers et al., 2013)) in SESA. ii) If the flow exhibits a pronounced eastward component, it transports moisture to the South Atlantic Convergence Zone (SACZ) and leads to enhanced rainfall in SEBRA (Liebmann et al., 2004; Carvalho et al., 2004a). The oscillation in rainfall amounts between SESA and SEBRA is the most important source of intraseasonal rainfall variability of the South American Monsoon System (Vera et al., 2006; Marengo et al., 2004; Jorgetti et al., 2013).

Several studies have investigated the dipolar rainfall pattern between SESA and SEBRA during the monsoon season in South America. They mostly rely on principal component analysis (PCA) on the basis of daily mean values of precipitation or outgoing long-wave radiation (as a proxy for convection) data (Nogués-Paegle and Mo, 1997; Liebmann et al., 2004; Marengo et al., 2004; Vera et al., 2006; Marengo et al., 2012). However, no corresponding characteristics of spatial co-variability have been analyzed for extreme values so far, since PCA is not applicable for this purpose: PCA only includes the first two moments of the data distribution, and is thus by construction not capable of capturing the characteristics of extreme events, which are located at the tail of the distribution. Furthermore, it is questionable in general to apply PCA to strongly non-Gaussian data distributions, since the resulting empirical orthogonal functions (EOFs) are – while uncorrelated – not independent and interpretation of their patterns is likely to be misleading (Monahan et al., 2009).

Here, we intend to fill this gap by employing a methodology which has recently been introduced to reveal and analyze the spatial characteristics of extreme rainfall co-variability (Malik et al., 2012b; Boers et al., 2013; Boers et al., 2014b). This method is based on the combination of a non-linear synchronization measure and complex network theory.

5.2. Data

We employ satellite-derived rainfall data from the Tropical Rainfall Measurement Mission (TRMM 3B42 V7, (Huffman et al., 2007)) with daily temporal and $0.25° \times 0.25°$ spatial resolutions. Geopotential height and wind fields at 850mb are obtained from NASA's Modern-Era Retrospective Analysis for Research and Applications (MERRA, (Rienecker et al., 2011)) on daily temporal and $1/2° \times 2/3°$ spatial resolutions. The spatial domain is confined to the coordinates 40°S to 15°N and 85°W to 30°W (Fig. 5.1). All datasets are analyzed for the monsoon seasons (December to February, DJF) between 1998 and 2012.

5.3. Methods

5.3.1. Extreme rainfall events

We define extreme rainfall events as days on which rainfall amounts exceed the 90th percentile of the rainfall distribution restricted to days with rainfall sums larger than 0.01mm/day. This percentile threshold is local in the sense that it depends on the respective grid cell's rainfall distribution.

5.3.2. Phases of the rainfall dipole

We define the two phases of the rainfall dipole between SESA and SEBRA on the basis of extreme rainfall event frequencies in the two regions. The SESA (SEBRA) phase of the dipole is defined as the set of days on which the sum of extreme events in the entire spatial domain of the SESA (SEBRA) box exceeds the 80 percentile (Fig. 5.1). Note that these definitions are by construction independent in the sense that the SESA (SEBRA) phase only depends on the number of events in the SESA (SEBRA) box. Nevertheless, there is no temporal overlap between the two phases, which can be explained by the dipolar rainfall pattern between the two regions.

This amounts to an average of 18 active days for both SESA and SEBRA phases per DJF season. Composites of rainfall, geopotential height, and wind, as well as complex networks will in the following be constructed separately for these two phases.

5.3.3. Event Synchronization

We consider an event-based measure of similarity to quantify the co-variability of extreme rainfall at different grid cells. For this purpose, we employ *event synchronization* (ES), modified on the basis of (Quiroga et al., 2002). Suppose we have two event series $\{e_i^\mu\}_{1\leq\mu\leq l_i}$ and $\{e_j^\nu\}_{1\leq\nu\leq l_j}$ with l_i (l_j) events at grid points i (j), where e_i^μ denotes the time index of the μ-th event observed at grid point i. In order to decide if two events e_i^μ and e_j^ν can be uniquely assigned to each other, we first compute the waiting times between events $d_{ij}^{\mu,\nu} := |e_i^\mu - e_j^\nu|$. On this basis, we define the *dynamical delay*:

$$\tau_{ij}^{\mu\nu} = \min \frac{\{d_{ii}^{\mu,\mu-1}, d_{ii}^{\mu,\mu+1}, d_{jj}^{\nu,\nu-1}, d_{jj}^{\nu,\nu+1}\}}{2} \tag{5.1}$$

To exclude unreasonably long delays between events at different locations, we introduce a maximum delay τ_{max}. If $d_{ij}^{\mu,\nu} \leq \tau_{ij}^{\mu\nu}$ and $d_{ij}^{\mu,\nu} < \tau_{max}$, we count this as synchronous events:

$$S_{ij}^{\mu\nu} = \begin{cases} 1 & \text{if} \quad d_{ij}^{\mu,\nu} \leq \tau_{ij}^{\mu\nu} \quad \text{and} \quad d_{ij}^{\mu,\nu} \leq \tau_{max}, \\ 0 & \text{else.} \end{cases} \tag{5.2}$$

ES between e_i and e_j is given as the sum of all $S_{ij}^{\mu\nu}$ (for fixed i and j):

$$ES_{ij} := \sum_{\mu\nu} S_{ij}^{\mu\nu}. \tag{5.3}$$

Each value ES_{ij} thus gives the number of events at grid points i and j which occurred synchronously (i.e., could be uniquely assigned to each other) within τ_{max} days. This procedure is performed for all combinations of grid points i and j, with $1 \leq i, j \leq N = 48400$. One advantage of this similarity measure is that it accounts for varying temporal delays between events at different grid cells within the prescribed maximum delay τ_{max}. In contrast, the classical lead-lag analysis on the basis of Pearson's Correlation Coefficient only yields one lead or lag for the entire time series.

5.3.4. Complex Networks: Construction

We will construct two separate networks for the two phases of the dipole in the following way: For the computation of the matrix $\{ES_{ij}\}_{1\leq i,j\leq N}$ for the SESA (SEBRA) phase, we will only consider those events in the time series $\{e_i^\mu\}_{1\leq\mu\leq l_i}$, which fall into the SESA (SEBRA) oscillation phase and ignore remaining events.

For each grid cell (i, j), we need to estimate the statistical significance of the value ES_{ij}. For this purpose, we construct a null model assuming that the l_i events at i and l_j events at j are placed independently according to a uniform distribution: We construct 1000 surrogate pairs of an event time series for each combination of event numbers (l_i, l_j). By computing ES for all these pairs, we obtain a relative frequency distribution of values of ES consistent with the assumptions of the null model and infer the score of the 95% significance level from this distribution. Denoting this score by T, a network link will be placed between grid cells i and j if the corresponding value ES_{ij} is above T. For two arbitrary grid cells i and j, this can be formally written as

$$A_{ij} = \Theta(ES_{ij} - T(l_i, l_j)) - \delta_{ij}, \tag{5.4}$$

where Θ denotes the Heaviside function and Kronecker's delta δ is added to exclude links from a grid cell to itself.

5.3.5. Complex Networks: Application

In this study, we consider two network measures. First, we compute the *degree* (DG). At a given network node i, DG_i is defined as the number of other nodes to which this node is connected to by a network link:

$$DG_i := \sum_{j=1}^{N} A_{ij} \tag{5.5}$$

Thus, DG at a given grid cell yields the number of other grid cells where extreme events occur synchronously with extreme events at that grid cell. A region with high DG will therefore be interpreted as a region, which is particularly important for spatial distribution and thus propagation of extreme rainfall.

In addition, we are interested in the directions along which extreme events occur synchronously. For this purpose, we employ the measure *directionality* (DR). At each node (i.e., grid cell) i, this measure yields two values: the dominant angle DR_i^ϕ among all network links at that node and the strength DR_i^s corresponding to that angle, quantified as the number of links pointing in that direction: Let ϕ_{ij} denote the angle between the meridian going through node i and the straight line between node i and j. Since the networks considered in this study are undirected, all angles ϕ_{ij} are taken as *modulo* π, thus $\phi_{ij} \in [0, \pi)$. In the following, we will therefore refer to DR_i^ϕ as an *orientation* rather than an angle. We first compute the frequency distribution of all orientations φ of links at i:

$$P_i(\varphi) = \sum_{j:\phi_{ij} \in (\varphi-\epsilon, \varphi+\epsilon)} A_{ij} \ , \tag{5.6}$$

where we consider all orientations differing by less than $\epsilon = 0.02$ as equal. DR is then defined by the maximum of this distribution

$$DR_i^s = \max_{\varphi \in [0,\pi)} P_i(\varphi) \tag{5.7}$$

together with the corresponding orientation

$$DR_i^\phi = \arg \max_{\varphi \in [0,\pi)} P_i(\varphi) \ . \tag{5.8}$$

This measure can be visualized by streamlines which are, at each grid point, directed along the orientation given by DR. The DR strength DR_i^s will in the following be indicated by the thickness of these streamlines. In order to be able to obtain a clear interpretation, we will compute DR only for networks constructed for *simultaneous* events ($\tau_{max} = 0$). By construction, extreme rainfall at grid points which lie on the same streamline occur typically at the same day. In this sense, they can be interpreted as *isochrones*. Under the assumption that the temporal resolution of 1 day is sufficiently high, we thus expect that rainfall clusters typically propagate perpendicular to these streamlines.

5.4. Results

5.4.1. Atmospheric Conditions

Composites of geopotential height and wind fields constructed separately for the two different dipole phases show distinctively different features (Fig. 5.2). As expected, we find anomalously high rainfall amounts over SESA for the SESA phase, but

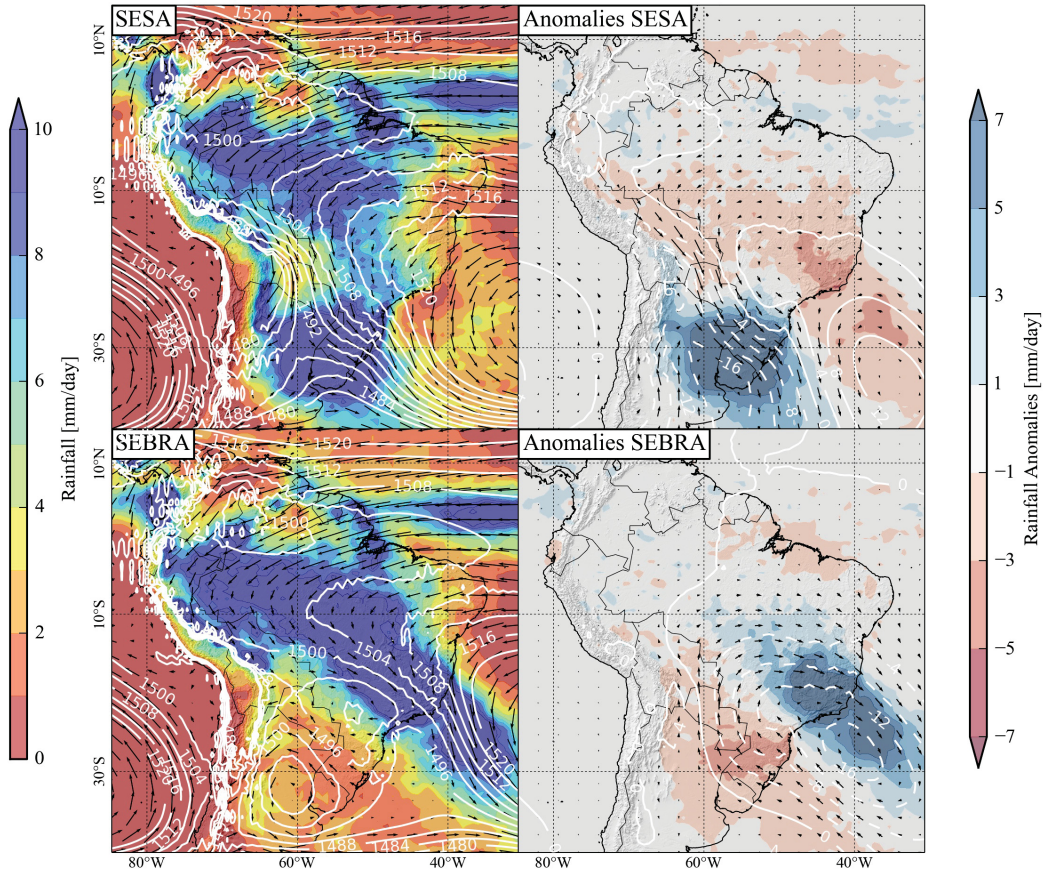

Figure 5.2.: *Left*: Composites for the SESA (top) and SEBRA (bottom) phase: daily rainfall (background), and geopotential height (white lines) and wind fields at 850mb (black lines). *Right*: Anomalies of the same variables with respect to DJF climatology.

negative anomalies over SEBRA (top row of Fig. 5.2). We further observe relatively low pressure values over SESA, which are associated with frontal systems initiated by Rossby wave-type patterns originating from the western Pacific Ocean (Siqueira and Machado, 2004; Liebmann et al., 2004; Seluchi and Garreaud, 2006). This low pressure system extends northwestwards along the eastern slopes of the southern Central Andes up to central Bolivia and forces the geostrophic low-level winds from the Amazon Basin southward along the eastern slopes of the Central Andes towards SESA (Nicolini et al., 2002).

For SEBRA phases (bottom row of Fig. 5.2), rainfall composites show pronounced positive rainfall anomalies over SEBRA, but negative anomalies over SESA. There is a high pressure system over northern Argentina and SESA, and relatively low pressure over SEBRA. This pressure configuration inhibits the southward flow from the Amazon and instead turns it eastwards towards the SACZ. There, we find clear indicators of an active convergence zone: northeasterly winds approaching from the

tropical Atlantic as well as northwesterly winds originating from the western Amazon Basin all converge over the SACZ. These results are consistent with earlier studies on intraseasonal rainfall variability over South America (Carvalho et al., 2002; Liebmann et al., 2004; Carvalho et al., 2004a, e.g.).

5.4.2. Complex network measures

For DG, we allow synchronizations of extreme events within a maximum delay of 3 days ($\tau_{max} = 3$), while for DR we only consider synchronizations at the very same day ($\tau_{max} = 0$). The choice of two different maximal temporal delays is justified by the different climatological interpretations we provide for the two network measures DG and DR. Note that computing DG for $\tau_{max} = 2$ or $\tau_{max} = 1$ does not substantially change the results.

5.4.3. Degree (DG)

For the SESA phase (top left in Fig. 5.3), we observe high values of DG over the ITCZ, the entire Amazon Basin, along the eastern slopes of the Andes from northern Peru to northern Argentina, as well as over SESA. In contrast, we find low DG values over the SEBRA.

For the SEBRA phase (bottom left in Fig. 5.3), we observe a substantially different spatial pattern of DG than for the SESA phase. High DG values in the vicinity of the ITCZ are located farther north than for the SESA phase. Over most parts of the Amazon Basin, values are even higher than for the SESA phase. Most notably, the highest values are located over the SACZ, extending from the central Amazon Basin to the subtropical Atlantic Ocean around 30°S and 30°W.

Note that the maximum delay $\tau_{max} = 3$ only serves as an upper bound for the dynamical delay τ (Eq. (5.1)) assuring the unique association of events in the computation of ES. Typically, extreme events synchronize on time scales shorter than 3 days, as is evident from comparing the results of Figure 5.3 with corresponding results for $\tau_{max} = 1$ and $\tau_{max} = 2$.

5.4.4. Directionality (DR)

For the reason explained in section 5.3.5, directionality is calculated for networks constructed from ES with $\tau_{max} = 0$. For the SESA phase, we observe N-S oriented streamlines over eastern Brazil, which turn anti-clockwise when moving farther west towards the Peruvian and Bolivian Andes, where they are approximately NW-SE oriented. The most pronounced streamlines can be found over SESA, where they are also oriented in NW-SE direction. This orientation continues towards the SACZ, however with reduced directionality strength $\mathbf{DR}^s$ (indicated by thinner streamlines).

For the SEBRA phase, the N-S oriented streamlines over northern Brazil rotate stronger than for the SESA phase when moving westward, with streamlines over central Brazil already oriented in NW-SE direction. We observe a clear pattern of almost straight, parallel streamlines extending from the central Amazon Basin

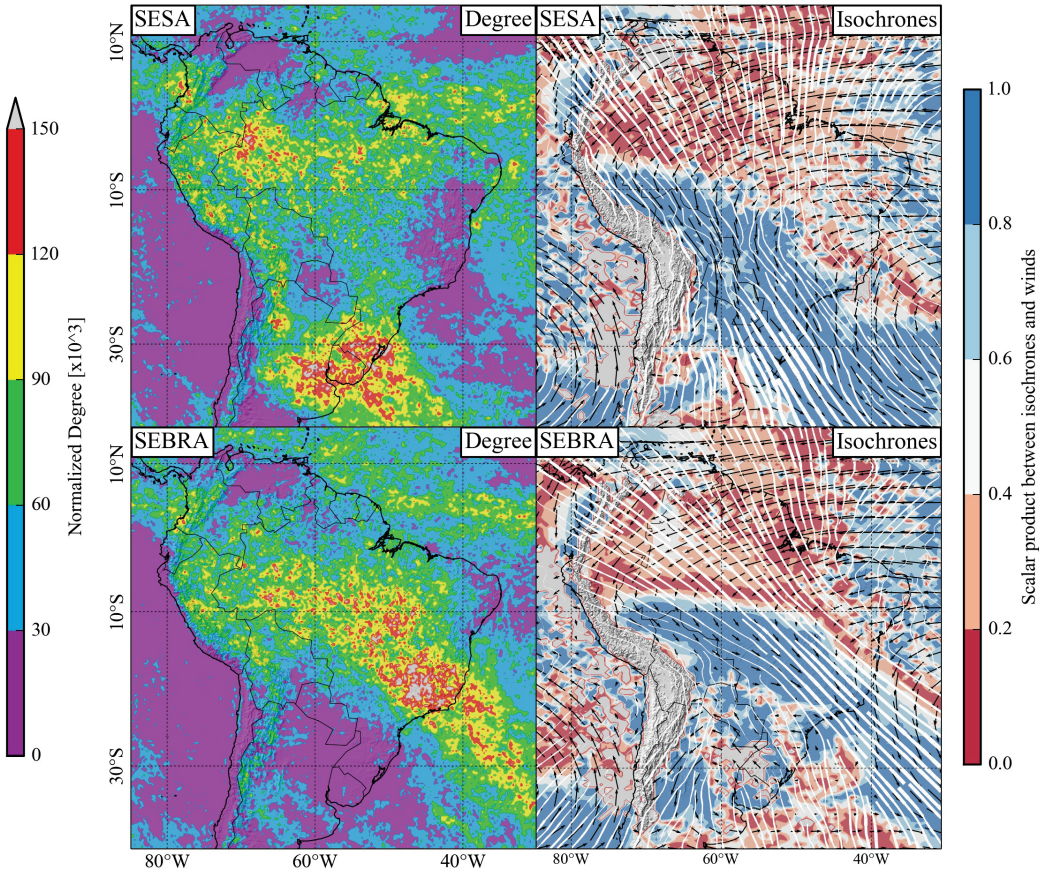

Figure 5.3.: *Left*: Degree (DG) normalized by the respective link density of the network for the SESA (top) and SEBRA (bottom) phase. *Right*: Isochrones, wind fields at 850mb as well as the absolute value of the scalar product between wind vectors and Isochrones for the SESA (top) and the SEBRA (bottom) phase.

southeastward across the southeastern Brazilian coast to the subtropical Atlantic Ocean. In contrast to the SESA phase, no streamlines can be observed over SESA.

For both dipole phases, we computed the scalar product between normalized wind vectors at 850mb (near surface) and normalized DR at each grid point in order to estimate the influence of the wind fields on the direction of extreme event propagation. We take the absolute value of the scalar product, since DR only yields an angle determining the orientation of network links rather then the actual direction. If the scalar product is close to 0, the orientation of extreme-event propagation is approximately parallel to the wind direction, while for values close to 1, it will be approximately perpendicular.

For the SESA phase (top right in Fig.5.3), we observe small values (between 0 and 0.4) over almost entire tropical South America between 10°N and 10°S, while the subtropics are characterized by values between 0.6 and 1.0. In particular, at the eastern slopes of the Central Andes in southern Peru and Bolivia as well as in SESA, the angles determined by DR are close to the wind angles.

For the SEBRA phase, the scalar product indicates that DR angles and wind vectors are perpendicular over the entire tropics between 10°N and 10°S. However, south of 10°, we find high values for the scalar product between wind vectors and DR angles extending from Bolivia east of the Andes to eastern Brazil. In particular over the climatological position of the SACZ, wind vectors and DR angles are typically parallel.

When subtracting the DG field for the SEBRA phase from the DG field obtained for the SESA phase (Fig. 5.4), the dipole between the two phases becomes clearly recognizable: Highest positive values are located over SESA, while highest negative values can be observed over SEBRA. However, it also becomes apparent that the oscillation is not confined to the dipole between SESA and SEBRA, but farther extends over the remaining parts of the South American continent, although with smaller amplitude: Southwest of SESA, we observe negative values, while northeast of the SACZ, around the equator, we find positive values for the DG difference. North of 5°N, we observe negative values again.

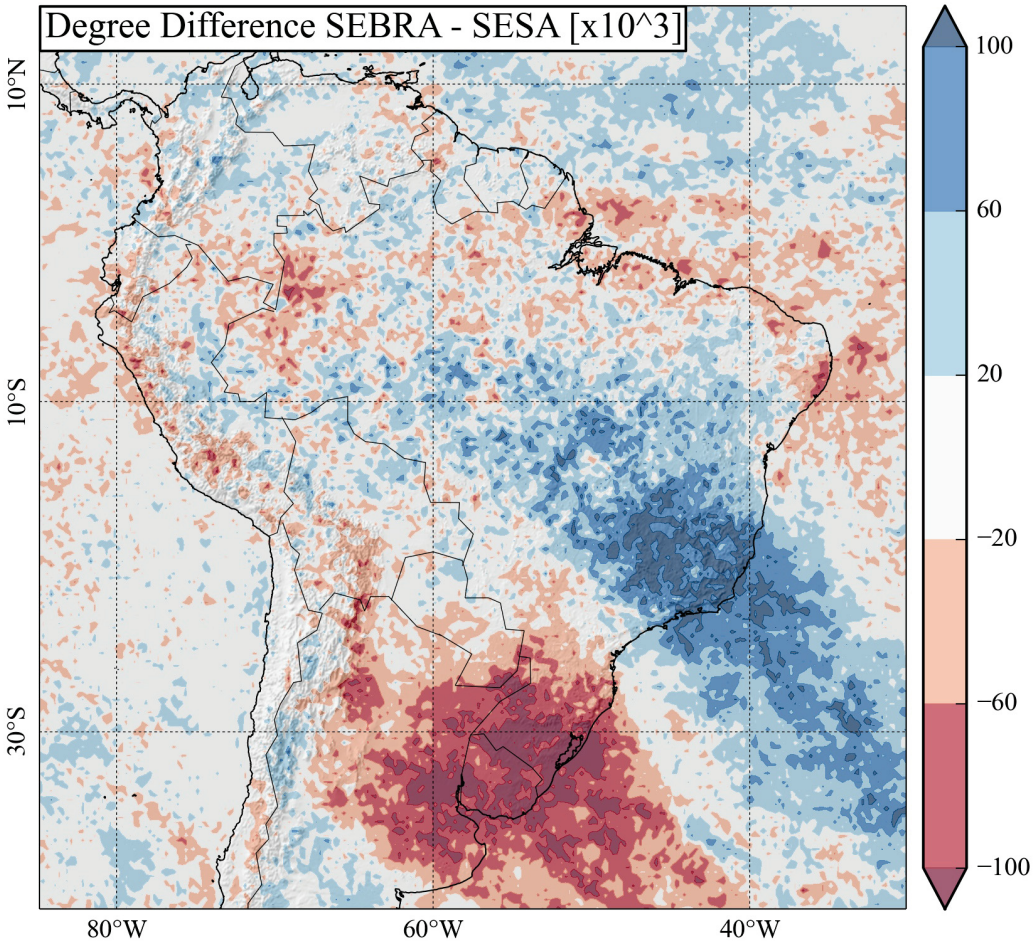

Figure 5.4.: Difference between Degree fields for the SEBRA and for the SESA phase. Note the oscillation between positive and negative values extending over the entire continent beyond the dipole between the SESA and SEBRA regions.

5.5. Discussion

DG yields an estimate of the importance of a given grid point for the synchronization paths of extreme events over the continent, as expressed by the number of other grid points where extreme events occur synchronously with extreme events at this grid point. It should be emphasized, however, that this does not imply that single rainfall clusters propagate along the entire signature of high DG values, but rather that possibly different extreme events synchronize along this signature.

Complementarily, DR provides the typical orientation along which extreme events synchronize. According to our interpretation that rainfall events propagate in direction perpendicular to the DR streamlines, and further assuming that rainfall events in tropical South America propagate from east to west rather than from west to east (Zhou and Lau, 1998; Vera et al., 2006; Marengo et al., 2012, e.g.), we infer the following main climatological propagation pathways:

1. For the SESA phase, rainfall events originating from the tropical and subtropical Atlantic Ocean enter the continent at the northern Brazilian coast, and propagate westward over the Amazon Basin (top row of Fig. 5.3). Extreme events synchronize according to the direction dictated by the low-level wind fields until they reach the western part of the Amazon Basin. As soon as they turn southward towards the Peruvian and Bolivian Andes, the streamlines of simultaneous occurrence of extreme events are aligned with the wind direction. This can be explained by the orographic impact of the Andes mountains (Bookhagen and Strecker, 2008): When the moist air is driven towards the mountains, it is lifted and causes extreme rainfall along the entire eastern slope of the Central Andes, extending from Peru to Bolivia. At the same time, the orography forces the low-level winds southward along the mountain slopes. South of 20°, we observe a pronounced propagation pattern over SESA, which implies synchronization of extreme rainfall events in a SW-NE direction. This is consistent with studies of extratropical cyclones and frontal systems, which move from southern Argentina northeastward, causing abundant rainfall over SESA (Siqueira and Machado, 2004; Seluchi and Garreaud, 2006). In SESA, wind directions are perpendicular to the direction along which extreme-rainfall events synchronize, which is typical for rainfall caused by these frontal systems, since the low-level winds from the north follow the isobars and interact with the frontal systems (top right in Fig. 5.2). Thus, we infer that the synchronization direction of extreme events in the tropics is determined by the low-level flow, while in the sub- and extra tropics, the influence of frontal systems is dominant.

2. For the SEBRA phase, extreme events also propagate from the tropical Atlantic Ocean westward to the Amazon Basin, but then occur simultaneously (i.e., at the same day) in a large area extending from the central Amazon Basin to the southeastern Brazilian coast and the adjacent subtropical Atlantic Ocean (Fig. 5.4). The orientation of isochrones suggests that they synchronize in NE-SW orientation, i.e., perpendicular to the wind direction, implying that frontal systems approaching from the south play the most pronounced role for driving extreme rainfall in SEBRA.

The difference between the DG fields obtained for the SEBRA and SESA phases suggest an oscillation over the entire continent rather then a single dipole between the regions SESA and SEBRA. While these two regions are clearly the most pronounced, the alternating pattern extends from central Argentina beyond the equator, with three maxima and two minima in total between 40°S and 15°N. During austral summer, large-scale circulation patterns in the form of Rossby waves, which emanate from the western Pacific Ocean, induce northward-propagating cold fronts in subtropical South America (Hoskins and Ambrizzi, 1993; Rodwell and Hoskins, 2001). The observed oscillation suggests that these Rossby waves control extreme-rainfall variability over the entire South American continent.

5.6. Conclusion

We studied the dynamical properties of extreme rainfall in the two most densely populated areas in South America: southeastern South America (SESA) including Buenos Aires and southeastern Brazil (SEBRA) around Saõ Paulo and Rio de Janeiro. A dipolar pattern of average rainfall between these two regions has previously been identified as the leading mode of intraseasonal variability in the South American Monsoon System. In order to study the dynamical properties of extreme-rainfall events associated with this dipole, we employed a combination of a non-linear synchronization measure and complex network theory. This approach allowed us to identify the pathways of extreme-rainfall synchronization and the network strength along these pathways. By constructing separate networks for the two phases of the rainfall dipole between SESA and SEBRA, we can distinguish the climatological synchronization routes of extreme rainfall for the two regimes: For the SESA phase, this route leads from the southern edge of the Intertropical Convergence Zone (ITCZ) across the Amazon Basin and subsequently southward along the Andes mountains to SESA. For the SEBRA phase, this path enters the continent north of the ITCZ and, after passing the Amazon Basin, turns southeastward to the SEBRA. By comparing climatological wind directions with the orientations of streamlines of synchronous extreme rainfall, we reveal a transition of driving mechanisms from the tropics to the subtropics: extreme-rainfall propagation in the tropics is driven directly by the (mainly easterly) low-level winds, but extreme-rainfall propagation in the subtropics is dominated by frontal systems approaching from the southern tip of the continent.

Our results indicate that the rainfall dipole between SESA and SEBRA is only the most prominent part of an oscillation which extends over the entire South American continent. This suggests that indirect influences of Rossby waves originating from the Pacific Ocean on extreme rainfall extend to tropical latitudes even beyond the equator.

Chapter 6.

Synchronous Occurrences of Heavy Rainfall Events in Germany

Abstract

Synchronous occurrences of heavy rainfall events and the study of their relation in time and space are of large socio-economical relevance, for instance for the agricultural and insurance sectors, but also for the general well-being of the population. In this study, the spatial synchronization structure is analyzed as a regional Climate Network constructed from precipitation event series. The similarity between event series is determined by the number of synchronous occurrences. We propose a novel standardization of this number that results in *synchronization scores* which are not biased by the number of events in the respective time series. Additionally, we introduce a new version of the network measure Directionality that measures the spatial Directionality of weighted links by also taking account of the effects of the spatial embedding of the network. This measure provides an estimate of heavy precipitation *isochrones* by pointing out directions along which rainfall events synchronize. We propose a climatological interpretation of this measure in terms of propagating fronts or *event traces* and confirm it for Germany by comparing our results to known atmospheric circulation patterns.

6.1. Introduction

In Germany, spatial structures of precipitation are mainly determined by the orography and its position in relation to the sea. This is not only the case for long-term means of precipitation sums (Klein and Menz, 2003), but holds true for heavy precipitation as well, as shown on the basis of the frequency of daily sums of 10mm and more (Gerstengarbe and Werner, 2009). As visualized in Fig. 6.1, large precipitation sums occur mainly in mountainous areas and in regions close to the coast of the North Sea. Small daily sums occur especially in the northeast.

In this study, our focus is exclusively on heavy precipitation in Germany, which we define as precipitation that leads to daily sums of at least 10mm. A weather station with a daily precipitation sum larger than or equal to this threshold is considered to have a precipitation event on the corresponding day.

Variations in the spatial distribution of precipitation are well-captured by precipitation events defined by said threshold: The spatial precipitation pattern, given by average daily precipitation sums, is very similar to the spatial pattern of average event rates (see Fig. 6.1). Furthermore, we consider 10mm as a good compromise between having a sufficient number of events at each location and a rather high threshold in order to study heavy precipitation. The average event rate for all event series of the entire period with a threshold of 10mm is around 0.064. On average, we have 1300 events per event series.

In order to study synchronous occurrences of heavy precipitation events, we specify *synchronization scores* between all 2337 meteorological stations and precipitation gauges in Germany. These *scores* are defined as the number of synchronous occurrences of events in the pairs of event series, standardized to the expected number of synchronous occurrences assuming uniform-randomly distributed events. The obtained spatial synchronization structure is analyzed in terms of a regional Climate Network. Recently, similar approaches have been taken for extreme rainfall events in the Indian Monsoon System (Malik et al., 2012a), as well as in the South American Monsoon System (Boers et al., 2013), but with a slightly different synchronization measure introduced by Quiroga et al., 2002 called Event Synchronization (ES). In general, applying complex network theory to climate time series has proven to deliver novel as well as established insights into climate dynamics (Paluš et al., 2011; Ebert-Uphoff and Deng, 2012b; Feng and Dijkstra, 2014).

We will introduce a new version of the measure Directionality, which provides *isochrones* along which heavy precipitation events typically occur synchronously. We are going to provide a climatological interpretation for this network measure, which will be confirmed by putting our results into relation with the atmospheric conditions of six important climatological circulation patterns over Germany, namely Bridge Central Europe (BM), High Central Europe (HM), Trough Central Europe (TRM), Trough Western Europe (TRW), Northwest Cyclonic (NWZ) and West Cyclonic (WZ).

6.2. Data

In this study, a precipitation gauge data set provided by the German Weather Service (Deutscher Wetterdienst) and the Potsdam Institute for Climate Impact Research is employed. It consists of 2337 daily time series for the time period 1951 to 2007 for Germany. Each time series consists of measurements of precipitation sums in mm/day. Since there are no missing measurements in the time period under consideration, the data is regularly sampled in time. However, it is irregularly sampled in space. In other words, the spatial coverage of rain gauges is not homogeneous. In order to derive spatial patterns which are independent of the spatial coverage of data, we use the method proposed by Rheinwalt et al., 2012, which will be explained in the next section.

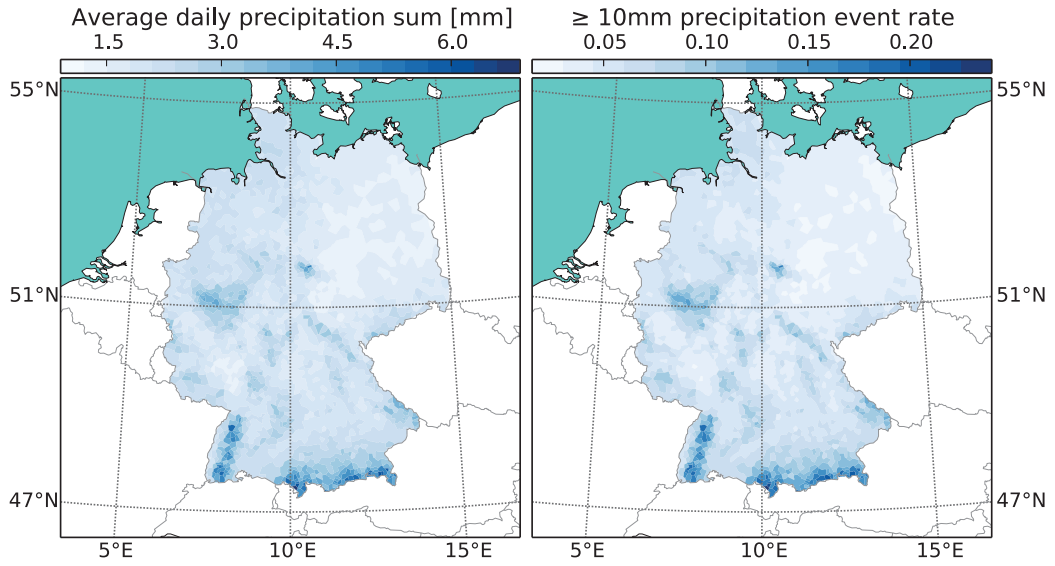

Figure 6.1.: Average daily precipitation sums (left) and rates of events with precipitation $\geq$ 10mm (right) for each weather station. The region of interest is decomposed into Voronoi cells with stations at cell centers. Each cell is colored according to the colorbar on top.

Concerning the comparison between results of this study to known atmospheric circulation patterns, we employ geopotential height at 850 hPa and corresponding wind field composites derived from the MERRA reanalysis product (Rienecker et al., 2011). This data product has a spatial resolution of $\frac{1}{2}^\circ$ on latitudes and $\frac{2}{3}^\circ$ on longitudes. This resolution does not resolve regional wind patterns, but still reveals large-scale atmospheric conditions. Although this daily data set starts only in 1981 it yields sufficiently accurate composites for the purpose of this study.

6.3. Methods

6.3.1. Climate Networks

To analyze the spatial structure of the temporal interrelations between climate time series, Climate Networks have proven to be particularly useful (Tsonis et al., 2006a; Donges et al., 2009b; Yamasaki et al., 2008a; Steinhaeuser et al., 2010; Malik et al., 2012a; Steinhaeuser et al., 2012b; Boers et al., 2013). Commonly, time series at different geographical locations are identified with network nodes and correlations between them are represented by network links.

Although there are studies using link-weighted Climate Networks (Gozolchiani et al., 2008; Steinhaeuser et al., 2010; Zemp et al., 2014; Boers et al., 2014a), most focus on unweighted Climate Networks; especially those concerning precipitation events (Malik et al., 2012a; Boers et al., 2013). In most Climate Networks studied

so far, only the strongest correlations between time series were represented by links. In addition, these links were typically unweighted. In contrast, we incorporate all possible links – not exclusively those corresponding to strong correlations – and weigh links according to their *synchronization score*.

The topological structure of such Climate Networks is assumed to encode properties of the underlying climate system. The topology is usually inspected in terms of network measures like Degree, Local Clustering Coefficient, Closeness Centrality, Shortest-path Betweenness, etc. (Donges et al., 2009b; Malik et al., 2012a; Boers et al., 2013, e.g.). These node-based measures determine values for each node in space. Such a scalar or vector field of network measure values can then be visualized on a geographical map with the same spatial embedding. Features of the underlying Climate Network can thereby be directly related to the corresponding geographical region.

However, said network measures are known to be influenced by the spatial embedding of the network itself (Gastner and Newman, 2006; Barnett et al., 2007; Henderson and Robinson, 2011; Rheinwalt et al., 2012; Berezin et al., 2012). Here, networks are confined by German national borders and thus network measures experience *boundary effects* (Rheinwalt et al., 2012). Additionally, network measures are also biased by the irregular sampling of nodes in space (Heitzig et al., 2012). In this study, all these effects of the spatial embedding on network measure scores are estimated and corrected for, using the method proposed by Rheinwalt et al., 2012, but adapted for weighted networks (see chapter 3).

6.3.2. Precipitation Event Series Analysis

Daily precipitation time series include many zeros and comparably few spikes. Such distributions are methodologically challenging for time series analysis, since commonly used similarity measures such as Pearson's correlation coefficient are debatable in such contexts. Here, we choose a novel event-based approach that uses a Poisson point process as a simple statistical model for expected synchronizations.

For each precipitation time series we have a series of precipitation events corresponding to days with precipitation sums above the threshold of 10mm. For each pair of such event time series, we then count how often events occur in both series at the very same time (see Fig. 6.2). These counts depend on the number of events. As the total number of events increases, the maximum possible number of counts also increases, as well as the number of counts that occur at random. In other words, these synchronization counts are biased by the number of events. This bias is called *event rate bias* in this study and is estimated by a Poisson point process. This point process provides the distribution of the expected number of synchronous events assuming a uniformly random distribution of events:

$$p(k) = \binom{m}{k} \frac{(l-m)!}{l!} \prod_{i=0}^{k-1} n - i \prod_{j=0}^{m-k-1} l - n - j \,, \tag{6.1}$$

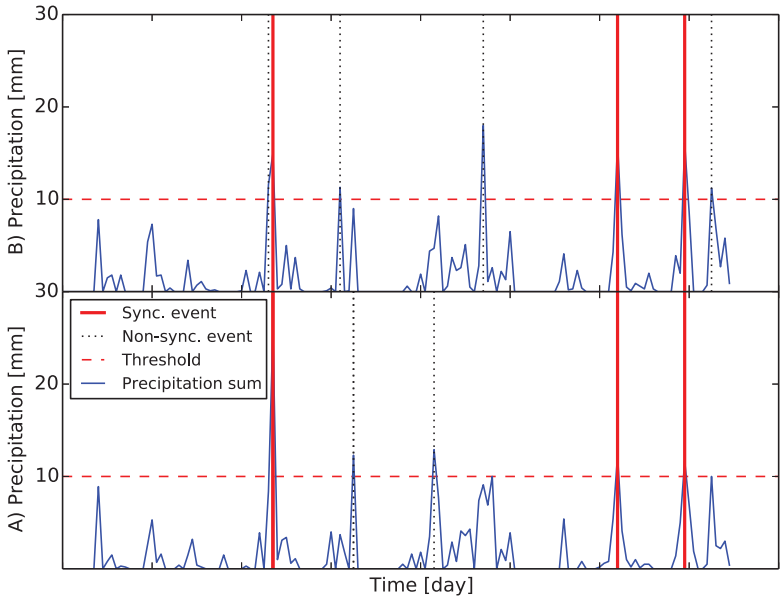

Figure 6.2.: Precipitation events (vertical red and dotted black lines) are defined as days with daily precipitation sums above the 10mm threshold (horizontal dashed red line). Events at weather station A synchronize with events at station B if they occur at the same time. These events are marked as vertical red lines.

where $p(k)$ is the probability to have k synchronizations between two event series of length l with n and m events, respectively, at uniformly random time positions ($n \geq m$). The *event rate bias* is estimated by the expected number of counts $\langle k \rangle_p$. The higher the event rates, the more events synchronize at random and hence, the higher is the *event rate bias*.

With the probability distribution $p(k)$ (Eq. 6.1) obtained for the Poisson point process as a statistical model, we could test the statistical significance of synchronizations. But due to the spatial proximity of weather stations, we would reject the null hypothesis that events at two weather stations are independent of each other in more than 97% cases with a confidence level of 99%. Hence, setting unweighted links by significance would lead to Climate Networks with very high link densities.

As an alternative, we standardize synchronization counts to the distribution $p(k)$. Instead of using observed counts, we use the difference between observed counts and expected counts $\langle k \rangle_p$ in units of the standard deviation of the corresponding $p(k)$. Hereby, the *event rate bias* in synchronization counts is eliminated and *synchronization scores* that are independent of the number of events are obtained. A *synchronization score* Z_{ij} between two time series i and j is regarded as an estimate of similarity between i and j:

$$Z_{ij} = \frac{k_{ij} - \langle k \rangle_p}{\sigma_p}\,, \tag{6.2}$$

where k_{ij} is the actual number of synchronous events for station i and j, $\langle k \rangle_p$ is the corresponding expectation value with respect to $p(k)$, and σ_p is standard deviation of the distribution $p(k)$. A comparison between the commonly used ES and our proposed similarity measure concerning the *event rate bias* is provided in the appendix A.3.

6.3.3. Weighted Climate Network Construction

Six weighted Climate Networks are examined in this study, constructed for the temporal subsets corresponding to the six dominant atmospheric circulation patterns over Germany. These networks thus represent the synchronization structure of precipitation events for the respective circulation pattern in the time period under consideration. The networks are constructed by using the corresponding matrix Z of *synchronization scores* computed for the respective temporal subsets as a link-weighted network adjacency matrix.

6.3.4. Network Measure

In this study, we analyze weighted Climate Networks with a novel version of the network measure Directionality (cf. P2), which accounts for biases in the measure due to spatial effects (cf. P1).

The Directionality of a node i specifies a vector (ρ_i, φ_i) that points in the dominant direction of links concerning their number and strength. The dominant direction is derived by the mode of the frequency distribution $P_i(\varphi)$ of synchronizations depending on the direction φ. The calculation of the mode is performed by using a fuzzy angle definition: A direction ϑ is regarded as identical to φ if it falls into the same angle interval $(\varphi - \epsilon, \varphi + \epsilon)$. In this study ϵ is chosen as 0.02 rad. Thus, $P_i(\varphi)$ is defined as:

$$P_i(\varphi) = \sum_{j:\phi(i,j)\in(\varphi-\epsilon,\varphi+\epsilon)} Z_{ij}\,,$$

where $\phi(i,j)$ denotes the angle of the link from node i to node j. Here, networks are undirected; therefore angles are projected onto the semicircle so that $\phi(i,j) = \phi(j,i)$ (see Fig. 6.3 for an example). Thus, φ_i specifies a tangent to node i, and not a vector, along which synchronizations occur most often. The Directionality of node i is defined as:

$$\rho_i = \max_{\varphi\in[0,\pi)} P_i(\varphi) \tag{6.3}$$

$$\varphi_i = \arg\max_{\varphi\in[0,\pi)} P_i(\varphi)\,. \tag{6.4}$$

Since $P_i(\varphi)$ is not a count of links, but a count of standardized synchronizations for a given fuzzy angle φ, it is called a frequency distribution of synchronizations and not of links. ρ_i is regarded as the *Directionality strength*.

In this study of synchronously occurring heavy precipitation events, the tangent to a node i defined by φ_i specifies *isochrones* for these events, i.e., lines along which events

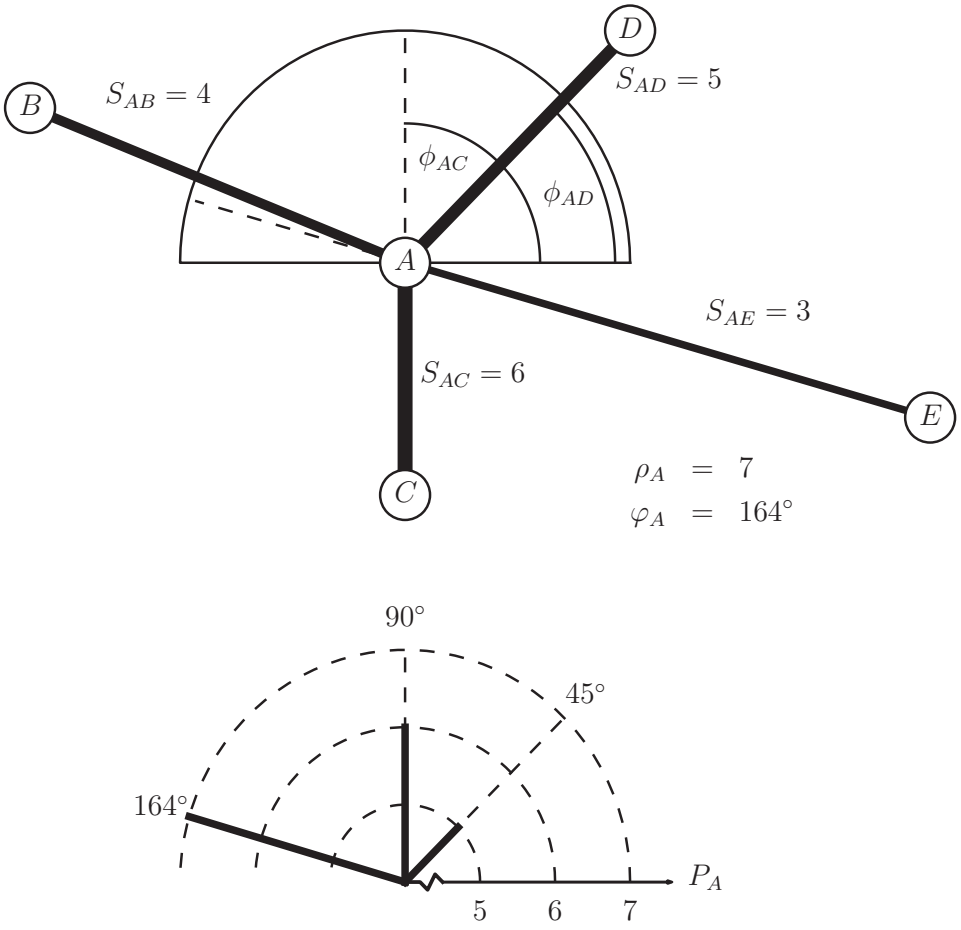

Figure 6.3.: *Directionality* for an example node A with four neighbors B, C, D and E. These four neighbors result in three directions due to the fuzzy definition of angles and the network being undirected. $\phi_{AD} = 45°$ and $\phi_{AC} = 270°$ but is flipped back into the first semicircle so that $\phi_{AC} = 90°$. The same is the case for ϕ_{AE}. Therefore, also due to the fuzziness of angles, we consider $\phi_{AB} \approx \phi_{AE} \approx 164°$. For this direction the distribution $P_A(164°) = S_{AB} + S_{AE} = 7$ which is the maximum of P_A. This results in a *Directionality strength* for node A of $\rho_A = 7$ and the tangent has an angle of $\varphi_A = 164°$.

occurred simultaneously. We thus expect the *isochrones* to be typically oriented along the low-level atmospheric flow direction as determined by the corresponding wind fields. However, this may depend on the propagation speed of precipitation systems, the temporal resolution of time series and the spatial extent of the region under consideration: In case of particularly low propagation speeds in relation to the daily temporal resolution of the data, the orientation of the *isochrones* may be dominated by the actual spatial extension of the precipitation system rather than by its direction of propagation. This may happen for a slowly moving frontal system, where the *isochrones* should be expected to align with the orientation of the frontal system, and not with its direction of propagation. In contrast, fast moving precipitation systems will leave *event traces* in the direction of the flow. If these traces are dominant, *isochrones* feature their orientation.

The measure Directionality is visualized by coloring for each node i a tangent with angle φ_i according to its strength ρ_i, i.e., *isochrones* are colored by their *Directionality strength* (see Fig. 6.4-6.6).

Spatial effects influencing the network measure are estimated by using a *spatial surrogate network*. Such a surrogate will be constructed such that it has the same number of nodes with the same spatial embedding as well as the same dependence of the link probability on the spatial length of links. Accordingly, the average link weight for links of a certain spatial length is the same in the original network as well as in the surrogate. This is in contrast to (Rheinwalt et al., 2012), where the *spatial surrogate* was sampled by many unweighted networks in order to estimate spatial effects for unweighted network measures.

In order to correct for spatial effects in the measure Directionality, we propose the following: The frequency distribution of synchronizations $P_i(\varphi)$ for all nodes i is not only estimated on the original network, but also on the spatial surrogate. The Directionality (ρ_i^-, φ_i^-) of node i that is independent of the spatial embedding of the network is then estimated by:

$$\rho_i^- = \max_{\varphi\in[0,\pi)} P_i^{orig}(\varphi) - c_i P_i^{surr}(\varphi) \tag{6.5}$$

$$\varphi_i^- = \arg\max_{\varphi\in[0,\pi)} P_i^{orig}(\varphi) - c_i P_i^{surr}(\varphi)\,. \tag{6.6}$$

The constant c_i is defined as the quotient of the two corresponding average frequencies of synchronizations: $c_i = \langle P_i^{orig}\rangle / \langle P_i^{surr}\rangle$. This re-scaling of the spatial surrogate frequency distribution of synchronizations $P_i^{surr}(\varphi)$ ensures that the Directionality (ρ_i^-, φ_i^-) is approximately the null vector if links are isotropic. However, due to the irregular sampling of nodes in space and boundaries in the network, even an isotropic link probability can lead to preferred directions of synchronizations. In other words, in such a case the frequency distribution of synchronizations $P_i^{orig}(\varphi)$ is not necessarily flat. However, if the surrogate frequency distribution of synchronizations $P_i^{surr}(\varphi)$ is subtracted, it does become flat since the spatial surrogate is constructed with an isotropic link probability. On the other hand, if the difference in frequency distributions has large positive maxima at certain angles, then these are due to more

synchronizations into the direction of these angles. A positive maximum at a certain fuzzy angle is therefore due to more synchronizations into that direction than what is expected from spatial effects.

In principle, a node can have multiple directions of preferred synchronizations, and our measure picks the dominant one. In order to estimate the uncertainty σ_i^- of this dominant direction, the weighted circular variance of all angles with positive differences in the frequency distributions of synchronizations is computed. It is defined as

$$(\sigma_i^-)^2 = \sum_{\varphi\in\Omega} w(\varphi)(\varphi - \varphi_i^-)^2 / \sum_{\varphi\in\Omega} w(\varphi)\,, \tag{6.7}$$

with $\Omega = \{\varphi : w(\varphi) > 0\}$ and weights

$w(\varphi) = P_i^{orig}(\varphi) - c_i P_i^{surr}(\varphi)$.

6.4. Results and Discussion

As mentioned in the methods section above, it is expected that *isochrone* patterns for precipitation are strongly determined by the flow direction of air masses. Here, the influence of atmospheric circulation patterns on the orientation of *isochrones* and values of the *Directionality strength* is assessed by reference to six of the most frequent circulation patterns according to Hess and Brezowsky (Werner and Gerstengarbe, 2010): Two rather dry patterns with low wind speeds (BM and HM, Fig. 6.4), two with intermediate wind speeds but high importance for long lasting rainfall that is likely to lead to floods (Mudelsee et al., 2004) (TRM and TRW, Fig. 6.5), and the most frequent circulation patterns with high wind speeds (NWZ and WZ, Fig. 6.6). For each circulation pattern the novel network measure is compared to a composite of geopotential height and wind at 850 hPa. Note that the geopotential height fields are on the same scale in Figs. 6.4, 6.5, and 6.6, while the length of the wind arrows are not comparable among figures. Since wind speeds are proportional to the margin between the isobars, they can thus still be compared qualitatively.

Low wind speeds (Fig.6.4): The circulation pattern BM with bridge-like highs over Central Europe leads to a relatively slow eastward movement of northwest-to-southwest fronts over Germany (Werner and Gerstengarbe, 2010). If such fronts cause precipitation events, these events propagate slowly and are hence expected to lead to *isochrones* that are parallel to the orientation of fronts.

The circulation pattern HM is even drier than BM, especially in the northeast of Germany, where precipitation events do not synchronize significantly enough along a dominant direction. For this pattern, *isochrones* exhibit rather high uncertainty in some geographical regions, and are hence very thin or even non-existent in these areas. With anticyclones located over the middle of Germany, the circulation pattern HM favors clockwise, slow circulations of frontal systems (Werner and Gerstengarbe,

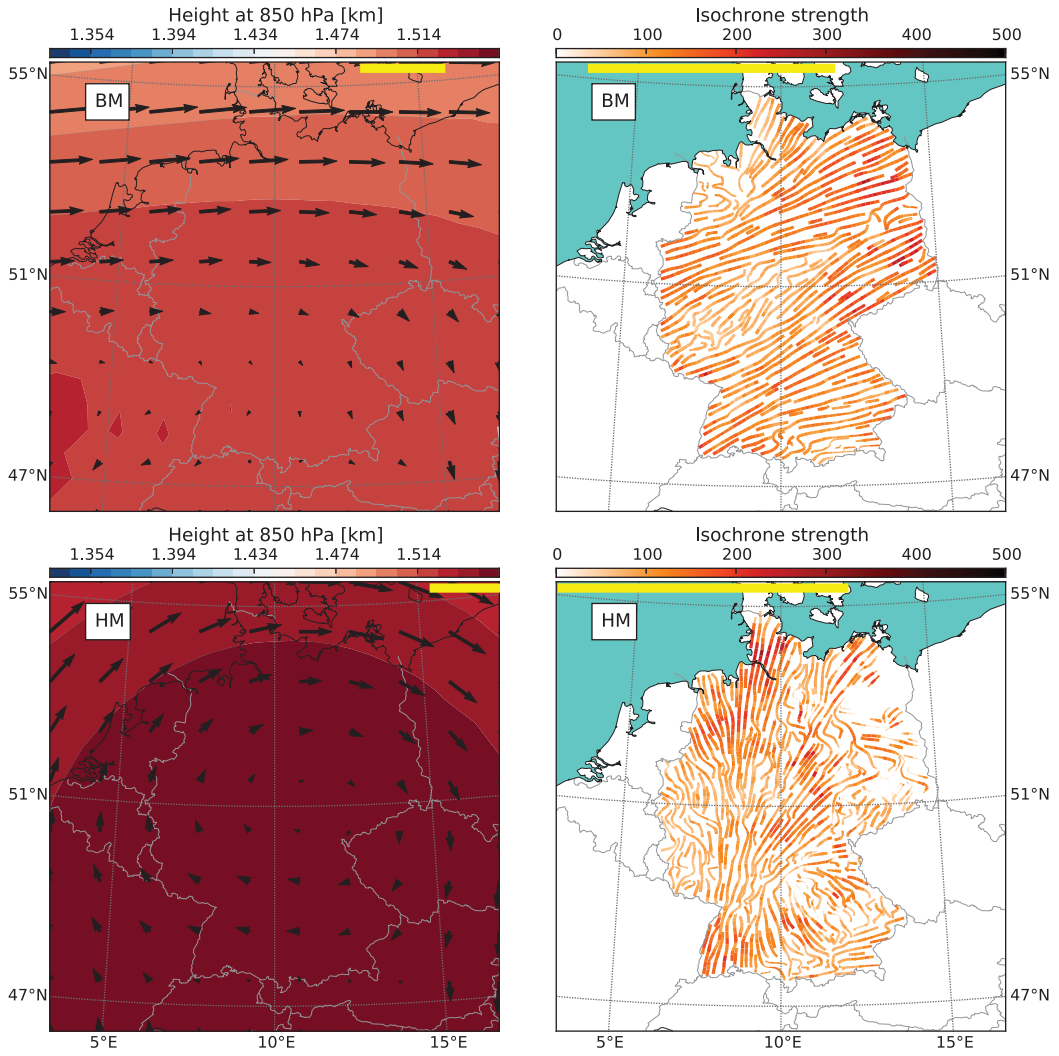

Figure 6.4.: *Left column*: Geopotential height at 850 hPa and corresponding wind field composites for the low wind speed circulation patterns BM (top) and HM (bottom). *Right column*: Network measure Directionality, visualized by *isochrones* for simultaneous events with color-coded values corresponding to the *Directionality strength*. The line width of *isochrones* is proportional to $1 - (\sigma^{-})^2$. Hence, more uncertain *isochrones* are thinner than more certain ones. Observe that *isochrones* are parallel to the expected orientation of fronts (as can be inferred from the composites in the left column). Both columns share a common colorbar and the range of values in a panel is marked by the corresponding yellow bar for that panel.

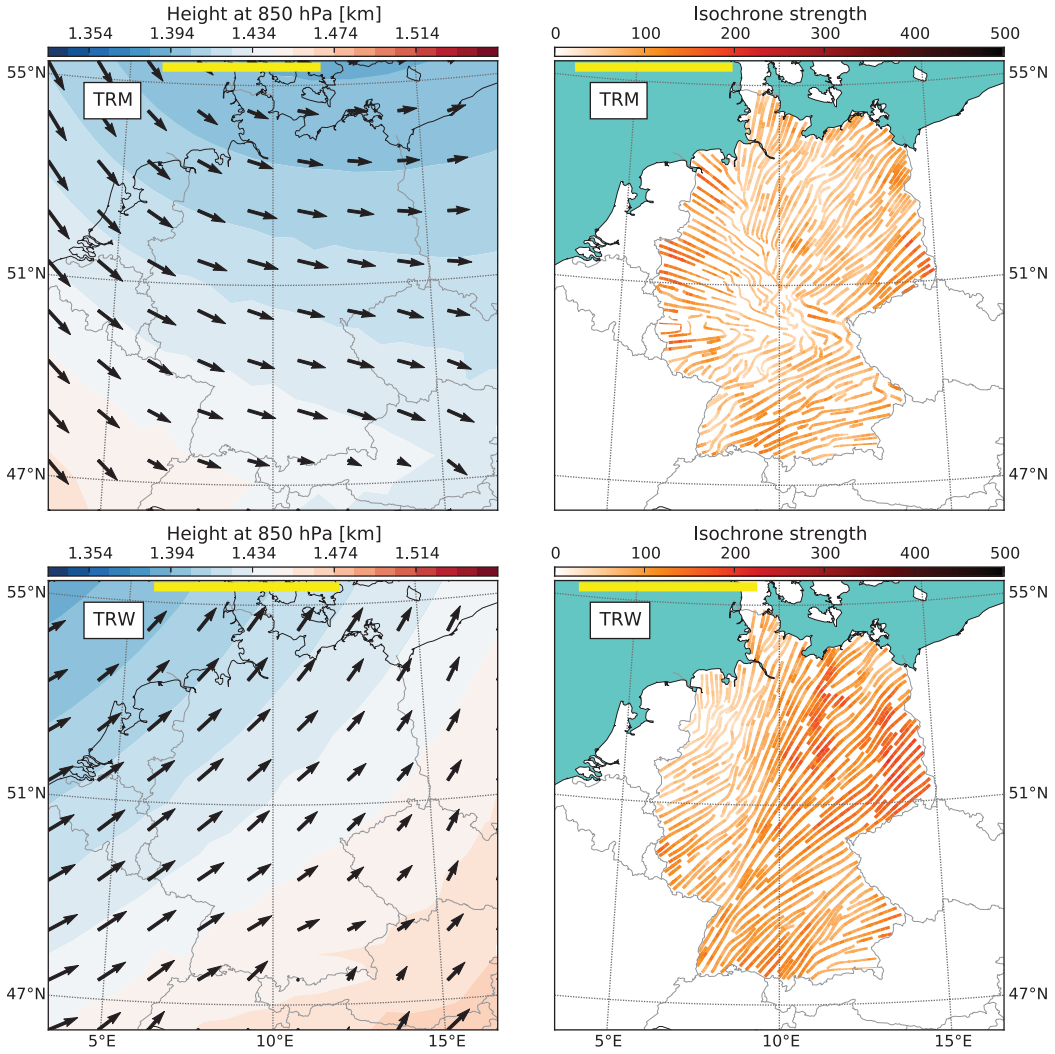

Figure 6.5.: Same as Fig. 6.4, but for the intermediate wind speed circulation patterns TRM and TRW. Observe that for TRM *isochrones* are typically oriented in accordance with corresponding frontal systems except for the northwestern part of Germany. For the circulation pattern TRW with even faster wind speeds, *isochrones* are mostly parallel to the flow.

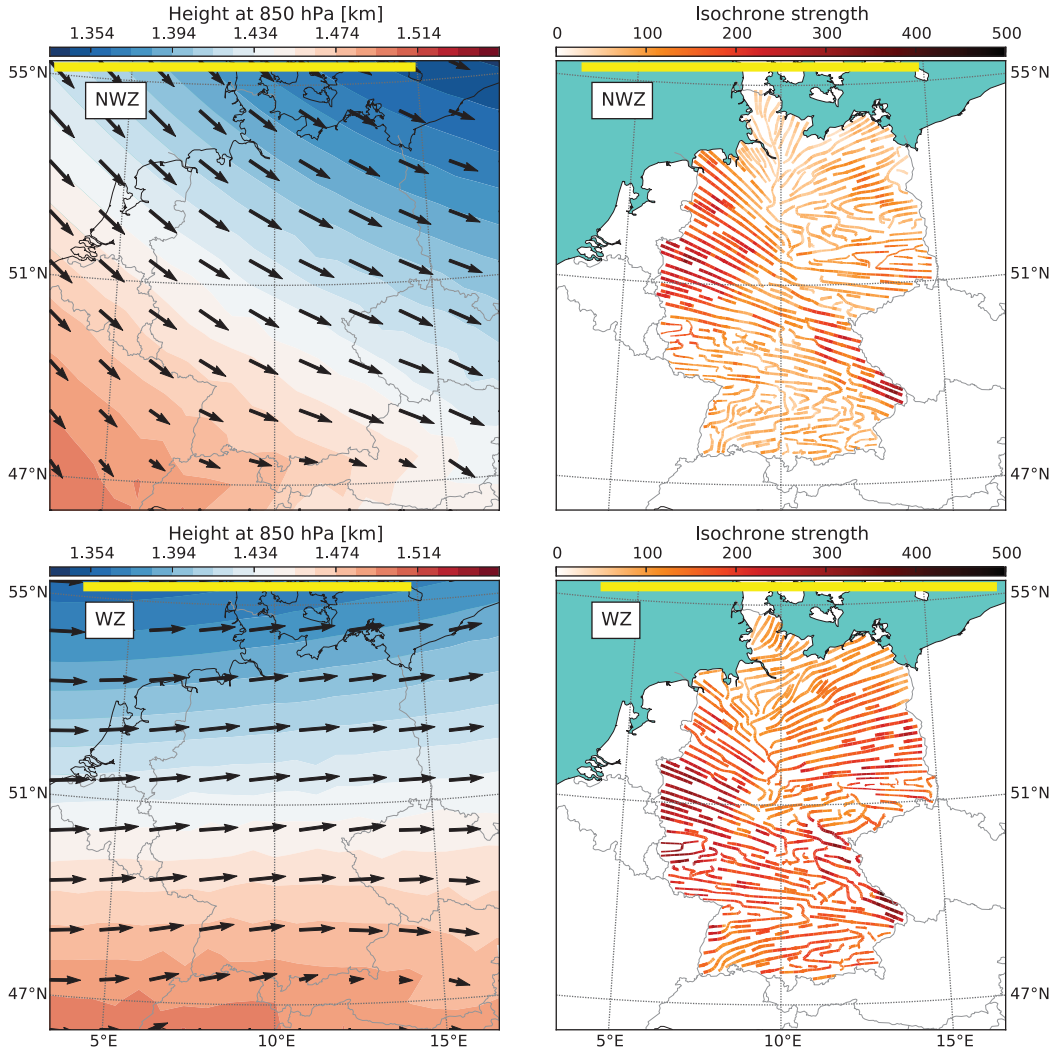

Figure 6.6.: Same as Fig. 6.4, but for the intermediate wind speed circulation patterns NWZ and WZ. Note that except for orographic perturbations *isochrones* are overall parallel to the wind flow at 850 hPa.

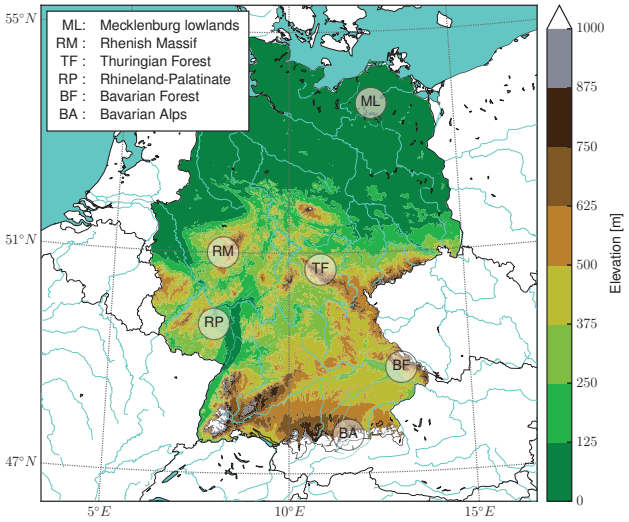

Figure 6.7: Orographic map of Germany showing regions to which this study refers to by name.

2010). In accordance with our climatological interpretation, this leads to an *isochrone* pattern with isochrones typically oriented perpendicularly to the circulation direction. However, this pattern is rather perturbed. This may be explained by the strong influence of orography on precipitation: For instance in the Bavarian Alps (BA in Fig. 6.7), *isochrones* are parallel to the mountain range of the Alps. Also the Rhenish Massif (RM in Fig. 6.7) and the Rhineland-Palatinate (RP in Fig. 6.7) perturb the *isochrone* pattern.

Intermediate wind speeds (Fig.6.5): From the geopotential height and wind composites obtained for the circulation pattern TRM, a southwest-northeast orientation of frontal systems can be inferred. In most parts of Germany, the corresponding *isochrones* are aligned with this orientation, suggesting that they are caused by rather slowly moving fronts. The exception to this is the northwestern part of Germany, where *isochrones* are rather parallel to the flow. This might be explainable by the fact that wind speeds over this region are higher than over the remaining parts of Germany, resulting in *isochrones* along the propagation direction of precipitation systems. This change in the orientation of the *isochrones* is very abrupt and can be expected to associated with the influence of orography. For instance, the influence of the Thuringian Forest (TF in Fig. 6.7) on *isochrones* is very pronounced in this circulation pattern.

For the circulation pattern TRW, *isochrones* are oriented along *event traces* and therefore parallel to the orientation of the flow, in accordance with our interpretation given in Section 6.3.4.

High wind speeds (Fig.6.6): For the circulation pattern NWZ, relatively strong *isochrones* are observed. Note the dark red *isochrones* in the northwest of Germany and along the mountain range from the Thuringian Forest (TF in Fig. 6.7) to the

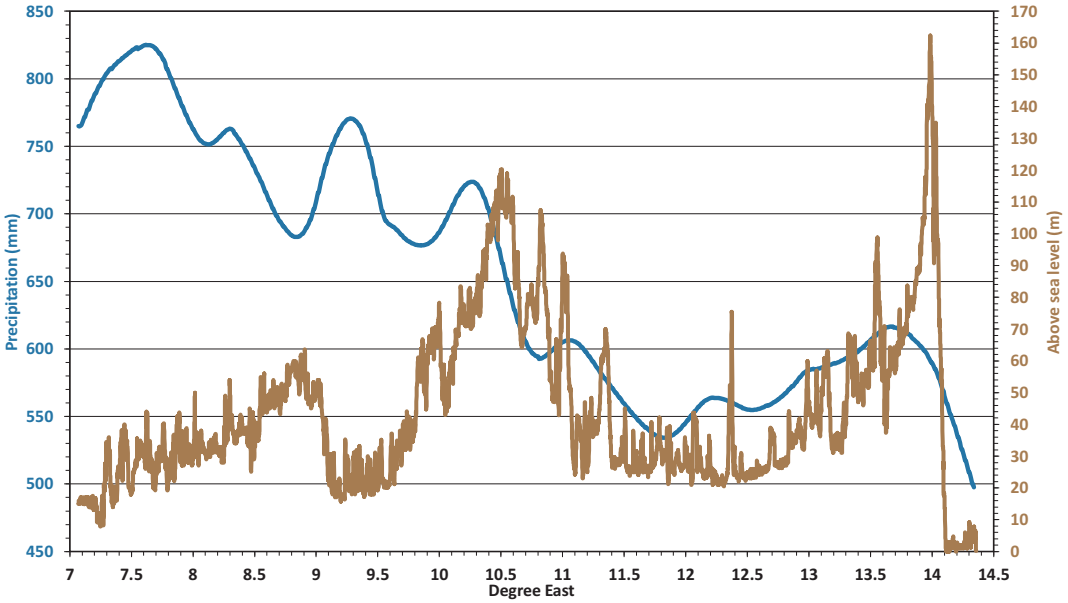

Figure 6.8.: Yearly precipitation in blue and above-sea-level altitude in brown along the $52.75°N$ line of latitude.

Bavarian Forest (BF in Fig. 6.7). All strong *isochrones* are parallel to the wind flow at 850 hPa. Especially in the northeast of Germany, *isochrones* are rather weak, and as for TRM, the orientation of *isochrones* changes abruptly approximately along the 10°E line of longitude. The reason for this abrupt change in orientation might be an orographic barrier at ≈ 10°E (see Fig. 6.8). The influence of this barrier on yearly precipitation sums is seen in Fig. 6.8, which depicts the orography and precipitation distribution alongside the 52.75°N line of latitude. One can see that the terrain ascends from approximately 9.5°E to 10.5°E from 20m to 120m above sea level, before descending again towards the Mecklenburg lowlands (ML in Fig. 6.7). Parallel to this, precipitation drops from 740mm/year at 10.5°E to only 550mm/year at 11.5°E.

This sudden change in *isochrone* orientation in the north of Germany can also be observed for the circulation pattern WZ. However, in this case the *Directionality strength* in the northeast of Germany is higher than for NWZ and more parallel to the wind flow. Also, *isochrones* in the mountainous South of Germany are less perturbed by orography and thus more in accordance with the flow direction for WZ than for NWZ.

6.5. Conclusion

Based on long-term gauge station data with high spatial resolution, we investigated the spatial characteristics of daily heavy precipitation synchronicity in Germany by means of complex networks. We introduced a new version of the network measure

Directionality that accounts for spatial effects in weighted Climate Networks, discussed its applicability on regional scales and provided climatological interpretations. This measure provides insights into the climatological orientation of the propagation of precipitation clusters and frontal systems in terms of *isochrones* which, while being consistent with known circulation patterns, go beyond the hitherto known. The presented methodology also provides a promising framework for evaluating climate models with respect to their implementation regarding heavy precipitation. Furthermore, we are convinced that our findings can be helpful for the assessment of hazard risks in form of floodings and, if combined with climate model projections, the development of these risks under ongoing climate change.

Chapter 7.

Teleconnections in Climate Networks

A Network of Networks Approach to Investigate the Influence of Sea Surface Temperature Variability on Monsoon Systems

Abstract

We analyze large-scale interdependencies between sea surface temperature (SST) and rainfall variability. We propose a novel Climate Network construction scheme which we call *Teleconnection Climate Networks* (TCN). On account of this analysis, gridded SST and rainfall data sets are coarse-grained by merging grid points that are dynamically similar to each other. The resulting clusters of time series are taken as the nodes of the TCN. The SST and rainfall systems are investigated as two separate Climate Networks, and teleconnections within the individual Climate Networks are studied with special focus on dipolar patterns. Our analysis reveals a pronounced rainfall dipole between southeastern Asia and the Afghanistan-Pakistan region, and we discuss the influences of Pacific SST anomalies on this dipole.

7.1. Introduction

Precipitation on the Asian continent is known to be influenced by large scale atmospheric processes like the Hadley and Walker circulation. However the intricate interplay of different atmospheric processes and how they influence precipitation variability are still not completely understood. Here, we study long-range interrelations within the precipitation system as well as between precipitation and sea surface temperature (SST) dynamics. Our aim is to shed light on the spatial structure of such teleconnections, with a special focus on precipitation dipoles and how they are influenced by SST variability.

For this purpose, we employ the Climate Networks approach by representing the interrelations between climatic time series as complex networks (Tsonis and Roebber, 2004b; Tsonis et al., 2006b; Ebert-Uphoff and Deng, 2012a; Donges et al., 2009b; Yamasaki et al., 2008b; Donges et al., 2009a; Malik et al., 2012c; Boers et al., 2013; Boers et al., 2014a). The SST as well as the precipitation system are studied as two separate networks, and the interrelations between them by their cross-topology.

So far, empirical orthogonal functions (EOFs), which are derived from principal component analysis of covariance matrices, are commonly used for a spatial analysis of teleconnections in climatological data (Kawamura, 1994; Deser, 2000; Ding and Wang, 2005, e.g.). While certainly very useful in many situations, they carry certain caveats in such analyses: First, if the data are not normally distributed the corresponding EOFs will in general, while uncorrelated, not be statistically independent (Monahan et al., 2009). Second, even if they are independent, EOFs do not necessarily uniquely correspond to climatological mechanisms (Dommenget and Latif, 2002). Third, and maybe most importantly, analyses based on the covariance matrix will only be able to capture linear dependencies. This might be considered insufficient in view of the strong nonlinearities involved in climatic interactions. Climate Networks can be considered as a complementary approach to study spatial patterns of climatic interrelations, which do not suffer from these statistical problems if derived from a nonlinear similarity measure. Furthermore, since teleconnections are not directly represented as links in EOFs they have to be deduced from the spatial patterns. Although this might be possible for simple teleconnection structures, it becomes challenging for more complicated ones.

Nonetheless, the common way of Climate Network construction is not suitable for the investigation of teleconnections as well. There, traditionally a pair-wise similarity analysis between all pairs of time series is performed, for instance by use of Pearson's correlation coefficient (Donges et al., 2009a; Tsonis et al., 2006b). However, Climate Networks are spatially embedded networks and the similarity between time series is strongly dependent on their spatial distance (P1): Two time series that are spatially close to each other are likely to be more similar than two time series which are far away from each other in space. By focusing only on strong similarities as in most Climate Network studies, networks have essentially only short links, which led to the investigation of paths in Climate Networks (Donges et al., 2009b).

Here we propose an approach that groups all time series by similarity into clusters. A related idea was also pursued in (Hlinka et al., 2014). We use a specific clustering scheme that typically provides spatially connected clusters due to the distance dependence of the similarities in climate systems. In other words, these clusters are localized regions of high resemblance according to the dynamics of the corresponding time series. Each cluster will in our approach be represented by a single time series and only the similarity structure between these representatives will be explored. By doing so we do not only reduce the dimensionality of the network, but we more importantly constructed a Climate Network that is reduced to its teleconnections. We will refer to these networks as *Teleconnection Climate Networks* (TCN).

7.2. Method

In order to group time series by similarity we use the standard fast greedy hierarchical agglomerative *complete linkage clustering* (Defays, 1977). This clustering is performed in a metric space with dissimilarities between time series as distances. In this study

we focus on the Spearman's rho correlation coefficient as the similarity measure in order to capture not only linear but also other monotonic relationships and in order to avoid problems of skewed distributions in precipitation data. In our case of standardized anomalies that have zero mean and unit variance this coefficient is proportional to the dot product between the ranked variables and can be interpreted as the cosine of the angle θ between these two ranked variables. More precisely, the Spearman's rho $\varrho_{X,Y}$ between two ranked time series X and Y is given by

$$\varrho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \equiv \frac{X \cdot Y}{\|X\| \, \|Y\|} = \cos(\theta_{X,Y}) \,. \tag{7.1}$$

This angle θ in radians between two time series is a distance that we use as the dissimilarity measure for the clustering.

Statistical significance of Spearman's rho values is estimated using twin surrogates[2]. These carry the advantage of preserving dynamical features of the original time series in contrast to bootstrapping methods (Thiel et al., 2006; Marwan et al., 2007b; Thiel et al., 2008; Romano et al., 2009). For each pair of time series we test against the null hypothesis that they are independent realizations of the same dynamical system. Upon repeating this for all pairs of time series we pick the maximum threshold corresponding to the 98% confidence level as a global significance threshold $T^{0.98}(\varrho)$.

We intend to group time series into clusters in such a way that all correlation values between time series within a given cluster are statistically significant. This is achieved by use of the *complete linkage clustering* scheme that is also known as *farthest neighbor clustering*. The distance measure between two clusters U and V is in this scheme defined as

$$D(U,V) = \max_{X \in U, Y \in V} d(X,Y) = \max_{X \in U, Y \in V} \theta_{X,Y} \,. \tag{7.2}$$

We cut the resulting dendrogram at the distance d_{crit} that corresponds to the significance threshold of all pair-wise correlation values, i.e., $d_{crit} = \arccos(T^{0.98}(\rho))$. This yields the maximum number of partitions of the set of time series such that for any two clusters U and V holds $D(U,V) \geq d_{crit}$, which is the same as the minimum number of partitions such that for any two time series $X, Y \in U$ in any given cluster U we have $\theta_{X,Y} < d_{crit}$. This clustering method does not only assure that all time series within a cluster are significantly correlated when cutting the dendrogram at d_{crit}, but also avoids the *chaining phenomenon* of the *single linkage clustering* where a set of time series might form a cluster although only a few time series are actually close to each other in the embedding space (Everitt et al., 2001). The clustering reduces the dimensionality of the problem by merging dynamically similar time series into clusters, which will serve as nodes for the *Teleconnection Climate Networks* TCN that will be constructed in the following.

More specifically, a TCN node is represented by a single time series from the corresponding cluster. Although there are clustering schemes, such as the *k-means*

[2] Due to the short length of time series we obtain the twin surrogates without embedding.

clustering (MacQueen et al., 1967), that suggest a certain member of a cluster as a representative, the in this study anticipated *complete linkage clustering* does not. Also, since cluster sizes vary, special care has to be taken when choosing a representative time series for a cluster. For instance, the point-wise mean of all time series within a cluster would be influenced by the size of the cluster. Instead we pick the time series with the highest average correlation to all other time series within that cluster as a representative for that cluster. This also has the advantage that the representative time series retain the original variabilities.

The TCN is now constructed by computing ϱ for all pairs of representative time series and assigning the corresponding values as link weights. We remove all links from the TCN that have a weight equal or below $T^{0.98}(\varrho)$.

We note that TCN could as well be studied using node-weighted network measures (Heitzig et al., 2012; Wiedermann et al., 2013). Although not focus of this study, this is an interesting topic of future research.

7.3. Application

We apply the proposed methodology to precipitation data for the Asian continent together with a global SST data set. We will in the following investigate dipole structures in the precipitation system and how these dipoles are influenced by SST variability.

7.3.1. Data

We use monthly time series for the years 1982 to 2008: SST data is obtained from the NOAA Optimum Interpolation SST V2 on a one by one degree grid (Reynolds et al., 2002), and precipitation data over land is taken from the APHRODITE V1101 daily precipitation data product on a 0.25×0.25 degree grid (Yatagai et al., 2012). In the latter data set, monthly mean values were calculated from daily values in a pre-processing step. We study monthly anomalies, in contrast to the monthly mean values itself, where the seasonal cycle would dominate correlation coefficients. Anomalies are calculated by subtracting from each value the long-term mean for that month and dividing by the corresponding long-term standard deviation.

7.3.2. Coarse-graining

Based on the significance tests explained above, we obtain significance thresholds $T^{0.98}(\varrho) = 0.199$ for the precipitation data set and $T^{0.98}(\varrho) = 0.494$ for the SST data set. Hence, we cut the Asian precipitation dendrogram at $\varrho = 0.2$. This leads to 111 precipitation clusters which are shown in Fig. 7.1. The geographical location of representative time series are depicted as black dots. With an initial number of 31624 time series the coarse-graining reduces the number of time series by a factor of $\approx$ 285. While the minimum correlation within clusters is 0.2, the average correlation within a cluster has a much higher value of 0.7.

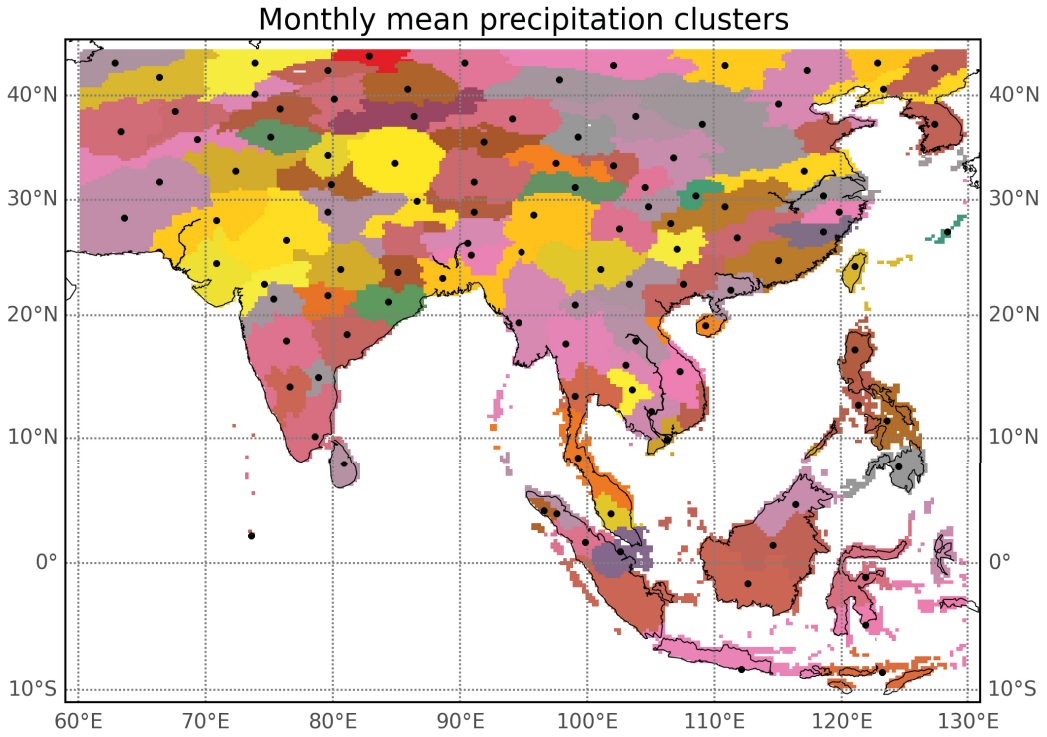

Figure 7.1.: Clustering of the precipitation data using the arccosine of the Spearman's rank correlation as a distance metric. All time series within a cluster are significantly correlated to each other. This corresponds to a minimum correlation of 0.2 between time series within a cluster. However, the average correlation within a cluster is on average 0.7. Geographical locations of representative time series for clusters are depicted as black dots.

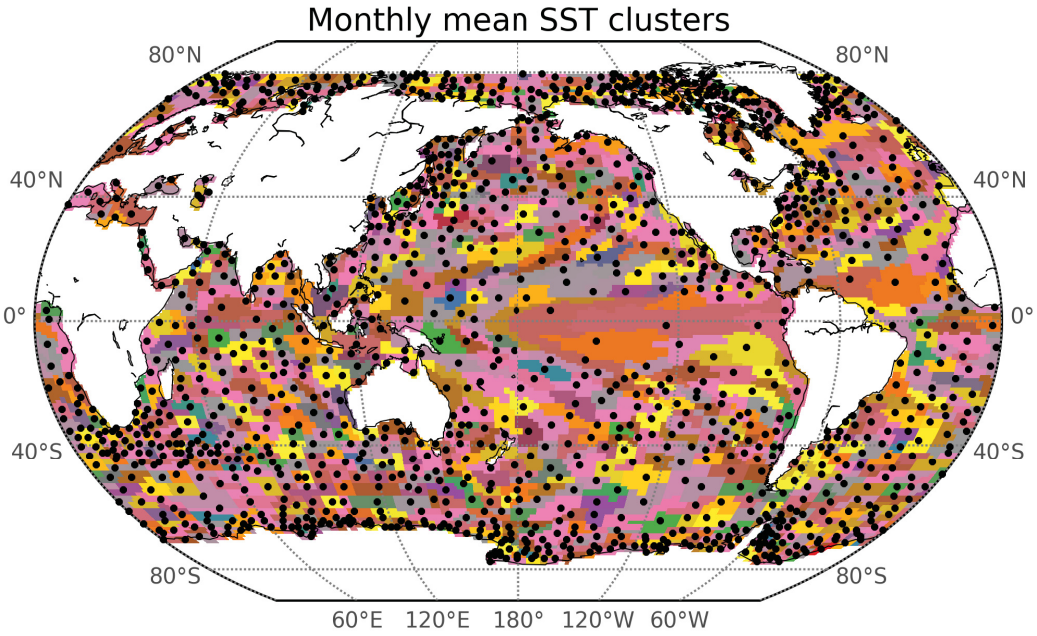

Figure 7.2.: Clustering of the SST data using the arccosine of the Spearman's rank correlation as a distance metric. All time series within a cluster are significantly correlated to each other, which corresponds to a minimum correlation of 0.5 between time series within a cluster. The average correlation within a cluster is on average 0.8. Geographical locations of representative time series for clusters are depicted as black dots.

We cut the global SST dendrogram at a threshold of $\varrho = 0.5$. This leads to 1419 SST clusters as shown in Fig. 7.2. With an initially number of 40780 SST time series the coarse-graining reduces the number of time series only by a factor of ≈ 29. This lower reduction is due to the relatively coarser spatial resolution of the SST data set. The correlation coefficient between SST time series within a cluster is, with an average value of 0.8, even higher than for the precipitation clustering.

7.3.3. Dipoles

In order to focus on precipitation dipoles we reduce the precipitation TCN by removing all nodes that do not even have a single significant link with a negative link weight. Note that we understand dipoles as anti-correlations between representative time series. The resulting network reflects the dipole structure that is captured from the APHRODITE data set for the considered time period. It consists of only 36 anti-correlation links (red) and 83 correlation links (blue) (see Fig. 7.3).

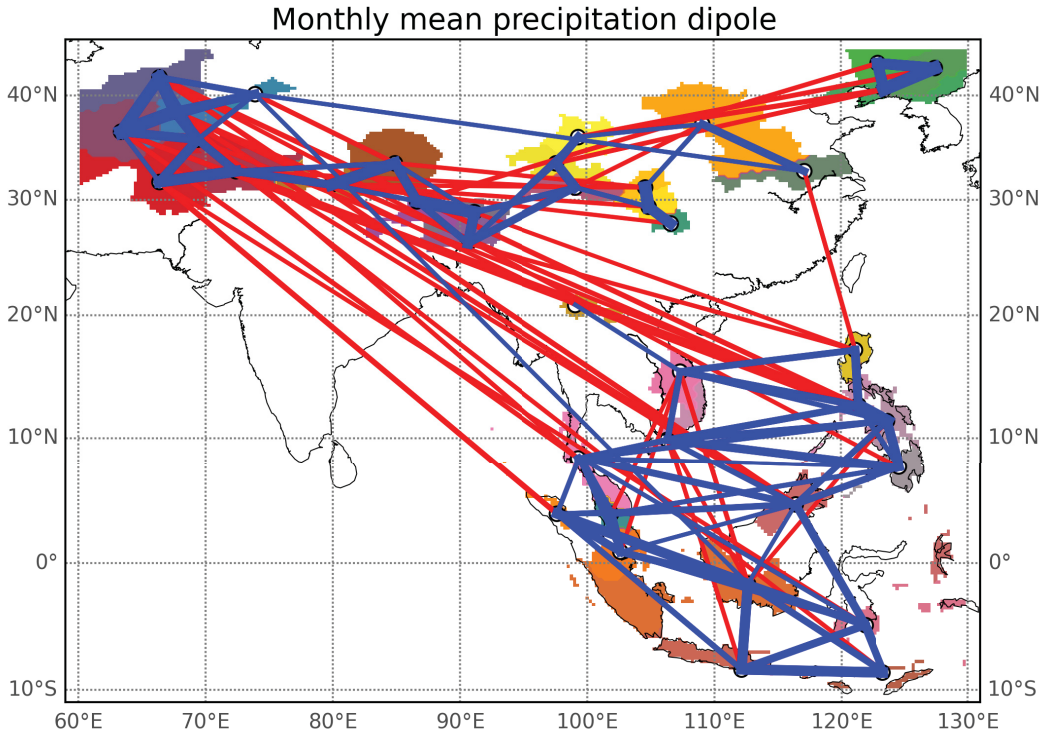

Figure 7.3.: The precipitation TCN reduced to nodes that have significant anti-correlations (red links) and correlations (blue links) to other representative precipitation time series. Link thickness is proportional to absolute link weight. Links are drawn between geographical positions of representative time series and the corresponding clusters are colored. Observe the pronounced precipitation dipole between southeastern Asia and the Afghanistan-Pakistan region.

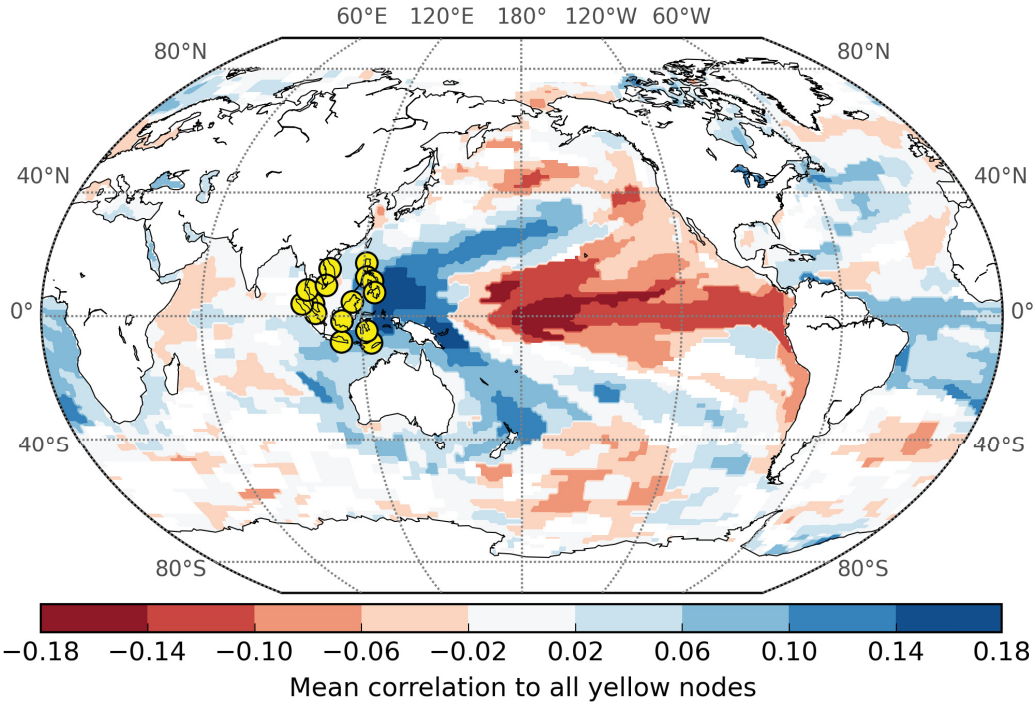

Figure 7.4.: Mean correlation between monthly precipitation anomalies in the southeast Asian pole of the dipole (yellow dots) to the global SST field. Observe the negative (red) mean correlation values between this pole and the SSTs in the tropical central and eastern Pacific, as well as the positive (blue) mean correlation pattern extending from the pole to the subtropics.

7.3.4. Networks of Climate Networks

Given the two sets of representative time series for the precipitation data set as well as for the SST data set, we estimate all pair-wise lagged correlation coefficients between these two sets. We consider possibly lagged correlation, because teleconnections between Asian precipitation and the global SST field can in general occur with a delay even on monthly scales. We employ a simple maximum correlation approach as follows. We focus on the influence of SST variability on precipitation, and thus only consider lags that correspond to SST dynamics preceding precipitation dynamics. Where we consider only lags up to 12 month. As link weights we take the first local maximum of Spearman's rho over this range of lags. A similar approach was taken for example by Yamasaki et al., 2008b.

In order to understand the influence of SST variability on the obtained Asian precipitation dipole, we examine cross links of nodes from the southeast Asian pole (see Fig. 7.3). All the nodes in this region, marked as yellow dots in Fig. 7.4, experience a spatially very similar influence from the SST network (not shown). Thus, we show the mean correlation from the SST network to these precipitation nodes (see Fig. 7.4).

7.4. Results and Discussion

Using the proposed method of TCN construction we find a strikingly pronounced precipitation dipole between the southeast Asian region and the Afghanistan-Pakistan region. This dipole has for example been described in Barlow et al., 2005. In that study, the authors partly explain its occurrence by an interplay of the Madden-Julian oscillation and the African-Arabian jet stream. Furthermore, this dipolar pattern is most likely related to the lateral component of the Asian monsoon system (Webster et al., 1999; Webster et al., 1998; Trenberth et al., 2000).

The southeast Asian region, in the precipitation network represented by nodes marked as yellow dots in Fig. 7.4, is a major deep convection area of the considered precipitation network. Convection is forced by solar heating, and forms a rising branch of the Hadley cell in this area, but is also modulated by the Walker circulation (Gill, 1980). This modulating effect explains the negative correlation values between precipitation in the southeast Asian region and SST anomalies in the eastern central tropical Pacific observed in Fig. 7.4: The Walker circulation causes upward atmospheric motion at the western boundary of the tropical Pacific, and downward motion at the eastern boundary. If the Walker circulation weakens as under El Niño conditions, convection is suppressed in the southeast Asian region, resulting in reduced precipitation. At the same time, upwelling of cold water in the eastern Pacific ocean is reduced, which causes positive SST anomalies in the eastern and central tropical Pacific. Correspondingly, a strengthened Walker circulation causes stronger convection in the southeast Asian region and negative SST anomalies in the eastern and central tropical Pacific.

On the other hand, we also observe a V-shaped pattern of positive correlation values in Fig. 7.4, with two branches extending to the subtropics. These two branches follow the climatological orientation of the trade winds in this region, and we suggest the following explanation for this pattern: Since the specific humidity of the low-level atmosphere rises with temperature, and the air temperature is in turn coupled to the SSTs, air parcels arriving at the southeast Asian region will carry the more (less) moisture the warmer (cooler) the SSTs are along the trajectory of the trade winds from the subtropics. This modulates the water vapor content of the air that rises in the southeast Asian region due to the convection discussed in the last paragraph, and hence the amount of precipitation. We note that this mechanism should also apply to the tropical Pacific, but there, its influence is strongly overprinted by the Walker circulation.

7.5. Conclusion

We proposed a new framework to construct multivariate Climate Networks from observational data. This framework is designed to study long-range interrelations, i.e., teleconnections, by first merging dynamically similar time series into clusters, and then investigating connections between these clusters. We applied our approach

to SST data as well as precipitation data over the Asian continent, and coupled the two separate networks obtained for each variable to a network of Climate Networks in order to study the impacts of SST variability on teleconnections in the precipitation network. Our analysis reveals a pronounced precipitation dipole between southeastern Asia and the Afghanistan-Pakistan region, which may be controlled by an interplay of the Madden-Julian Oscillation, and the African-Arabian jet stream. Results obtained from the coupled network of networks analysis further suggest that trade winds from the subtropics, as well as the Walker Circulation over the tropical Pacific in turn modulate this dipole.

Chapter 8.

Conclusion

8.1. Contributions

In conclusion, this thesis provides four main advances to the Climate Network approach, as well as to complex network theory in general. First, an improved estimation of co-variabilities of events by employing the similarity measure Event Synchronization in an unbiased way was introduced and applied, including the development of a novel continuous version of this measure called Kernel-based Event Synchronization. Second, a methodology to estimate spatial effects in complex networks was proposed and applied to artificial and observational data. Third, the novel network measure Directionality for spatially embedded networks was invented and applied. Fourth, a framework for the identification of teleconnections in climate systems using complex networks of networks was introduced and applied to Asian precipitation and global sea surface temperature data.

Event Synchronization We improved existing Climate Network approaches for studies of event synchronicity by introducing an Event Synchronization measure that is unbiased by the event rate in event time series. Although in general the *event rate bias* has to be estimated numerically, we could present an analytic model for the case of simultaneous events. Rainfall event Climate Networks constructed by use of this unbiased Event Synchronization measure were discussed and climatically interpreted on regional as well as continental scales. The introduced continuous Event Synchronization provides opportunities to extend the range of application to continuous time time series and point cloud data.

Associated publications: P2, P3, P4 and C3

Spatial Effects Depending on the spatial dependence of the link probability in complex networks, network measures are biased by the spatial embedding. Although this effect is known, the first approach to estimate these spatial effects in order to retrieve unbiased network measures was proposed in this thesis. The potential of the approach was demonstrated on artificial random networks, link weighted and unweighted, and on Climate Networks constructed from rainfall data, also link weighted as well as unweighted.

Associated publications: P1, P3 and C1

Directionality We introduced a novel network measure called Directionality for link-weighted as well as unweighted spatial networks that is unbiased by the spatial embedding. Its applicability to Climate Networks on regional as well as continental scales was discussed and climatological interpretations were provided. This measure provides insights into the climatological orientation of the propagation of precipitation clusters and frontal systems in terms of *isochrones* which, while being consistent with known circulation patterns, go beyond the hitherto known. The presented methodology also provides a promising framework for evaluating climate models with respect to their implementation regarding heavy precipitation. Furthermore, we are convinced that this network measure can be helpful for the assessment of hazard risks in form of floodings and, if combined with climate model projections, the development of these risks under ongoing climate change.

Associated publications: P2, P3 and P4

Teleconnections A novel framework for the construction of multivariate Climate Networks was proposed that is designed to study long-range interrelations between time series. By first merging dynamically similar time series into clusters, and then constructing a network of links between clusters, a coarse-grained teleconnection Climate Network is retrieved. This approach was applied to sea surface temperature data as well as precipitation data over the Asian continent. Discovered climatological insights proved the concept but went also beyond the hitherto known.

Associated publication: C2

8.2. Outlook

The introduced and applied Climate Network approach considering synchronizations of precipitation events does not depend on a specific type of event. Accordingly, one could apply this framework also in studies of precipitation synchronization considering different types of precipitation events, for instance hail events or events of sudden changes in precipitation rates. Or one could apply this framework also to completely different observables. Promising examples for that are lightning strokes, rapid soil moisture changes, earth quakes, volcano eruptions or neuronal spike trains.

First applications of the novel Kernel-based Event Synchronization to lightning strokes data also revealed a new concept of temporal networks. In contrast to a sliding window analysis, the time evolving network is estimated for a specific time point given by the temporal position of the kernel. Considering lightning activity this could, by use of the novel network measure Directionality, lead to interesting storm tracking applications in order to analyze origins of hurricanes. Directionality is a good example of a network measure that was specifically designed for Climate Networks with a certain task in mind. Unfortunately, this is rarely happening, in other words, little research is done on the development of new Climate Network measures.

Commonly, existing network measures from other disciplines like sociometry are applied to Climate Networks and therefore sometimes hard to reinterpret.

Moreover, could a network of Climate Network analysis of heavy rainfall events and rapid soil moisture changes provide deeper insights into the spatiotemporal coupling between these, and hence, serve as a basis for better flood prediction.

Especially in networks of Teleconnection Climate Networks with even more observables like precipitation, sea surface temperature and sea level pressure for instance, a combination with the causal networks approach would be instructive. A detailed topology of long-ranged causal dependencies would most likely reveal a deeper insight into the entanglement between monsoon variability, ENSO and planetary wave dynamics. This could further be extended to a higher temporal resolution. With daily precipitation time series of 56 years as in the APHRODITE product, time series are long enough for a sliding window analysis. This makes it possible to study the evolution of found teleconnections under ongoing climate change.

Appendix

Appendix A.

Regarding Chapter 6

A.1. Spatial effects on network measures of regional Climate Networks

From a link-weighted network constructed as in the corresponding chapter but for the entire time span of 56 years, an unweighted Climate Network is constructed in order to test the performance of the spatial effects estimation on unweighted networks. The unweighted network is retrieved by thresholding the link-weighted adjacency matrix with a threshold corresponding to a link density of 15%. In other words, the unweighted test network comprises the 15% strongest links, but considers them as equally strong. The performance of the method proposed (see Sec. 3.3, P1) is tested with three unweighted network measures: Degree, Closeness Centrality and Shortest-path Betweenness. In order to get a quantitative comparison, results of the whole regional network spanning all of Germany are compared to results of two smaller subnetworks. The two smaller subnetworks are given by dividing the whole regional network along the magenta colored line. All estimations of spatial effects are done using 1000 surrogates.

In summary, huge differences between network measures of the whole regional network and them of the two subnetworks are observable (cf. Fig. A.1.A with Fig. A.2.A). However, corresponding spatially corrected network measures are very similar regarding the spatial pattern (cf. Fig. A.1.C with Fig. A.2.C). This holds true even for complex path-based network measures like Shortest-path Betweenness.

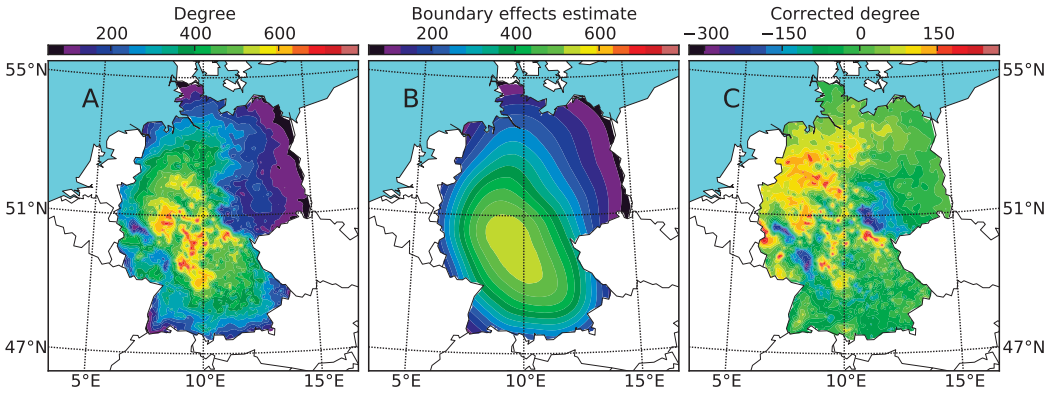

Figure A.1.: Degree example for the regional network with link density 15% of the whole region of Germany. Shown are the degree field (A), the corresponding boundary effects estimate (B) and the corresponding corrected degree field (C).

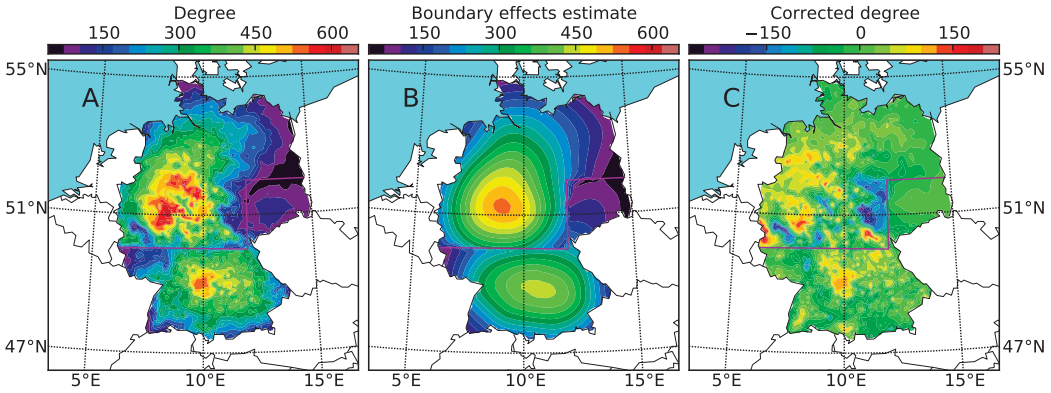

Figure A.2.: Degree example for two separate regional networks of two subregions of Germany that have the same links within their regions as the network of Fig. A.1. The border between these subregions is marked by the magenta colored line. Shown are the degree fields (A), the corresponding boundary effects estimates (B) and the corresponding corrected degree fields (C).

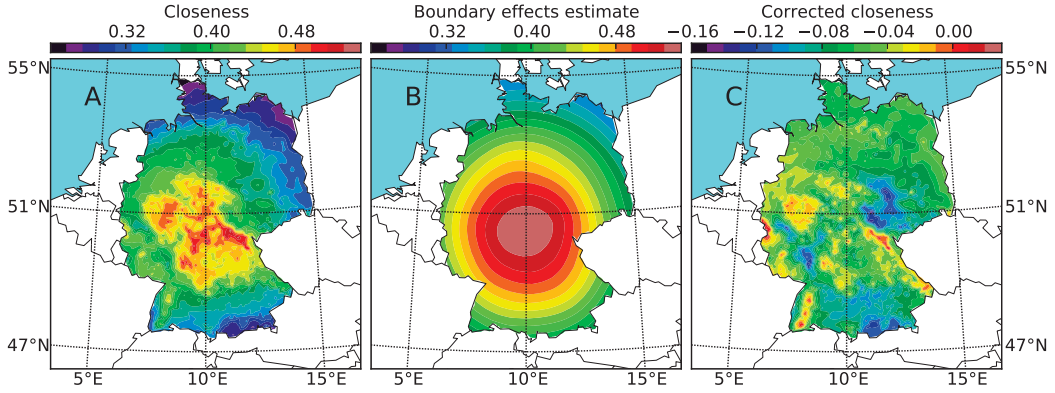

Figure A.3.: Closeness centrality example for the same regional network with link density 15% of the whole region of Germany as in Fig. A.1. Shown are the closeness field (A), the corresponding boundary effects estimate (B) and the corresponding corrected closeness field (C).

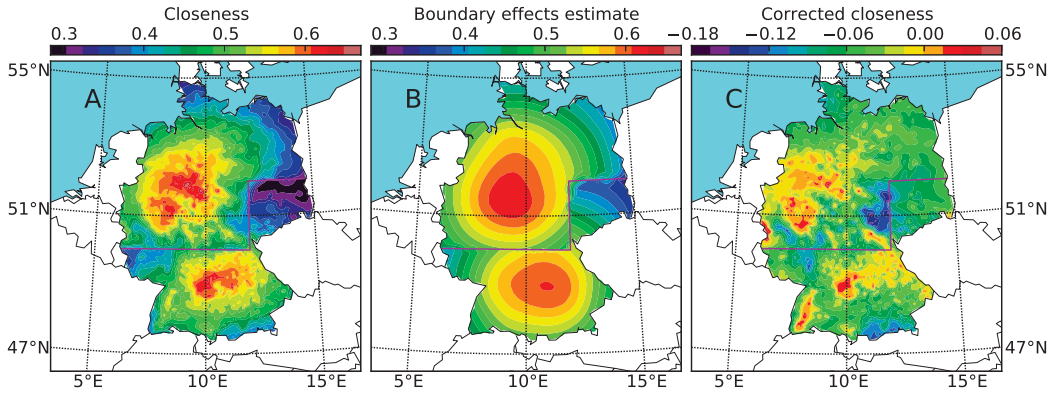

Figure A.4.: Closeness centrality example for two separate regional networks of two subregions of Germany that have the same links within their regions as the network of Fig. A.1. The border between these subregions is marked by the magenta colored line. Shown are the closeness fields (A), the corresponding boundary effects estimates (B) and the corresponding corrected closeness fields (C).

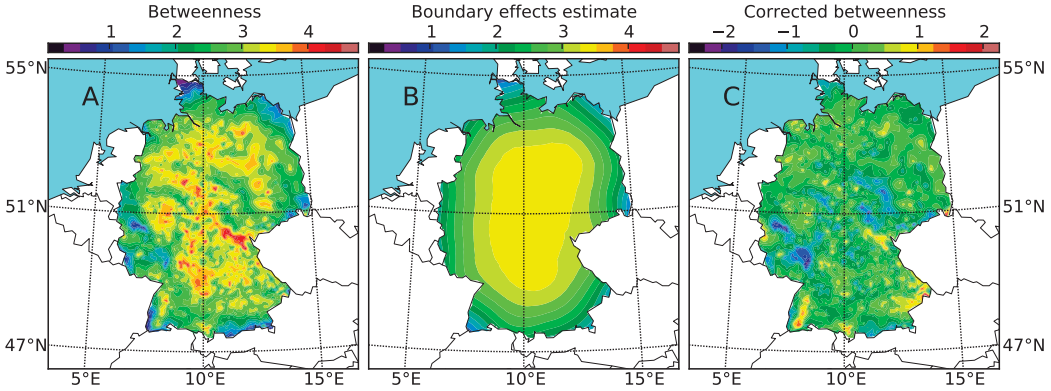

Figure A.5.: Shortest-path betweenness example for the same regional network with link density 15% of the whole region of Germany as in Fig. A.1. Shown are the betweenness field (A), the corresponding boundary effects estimate (B) and the corresponding corrected betweenness field (C).

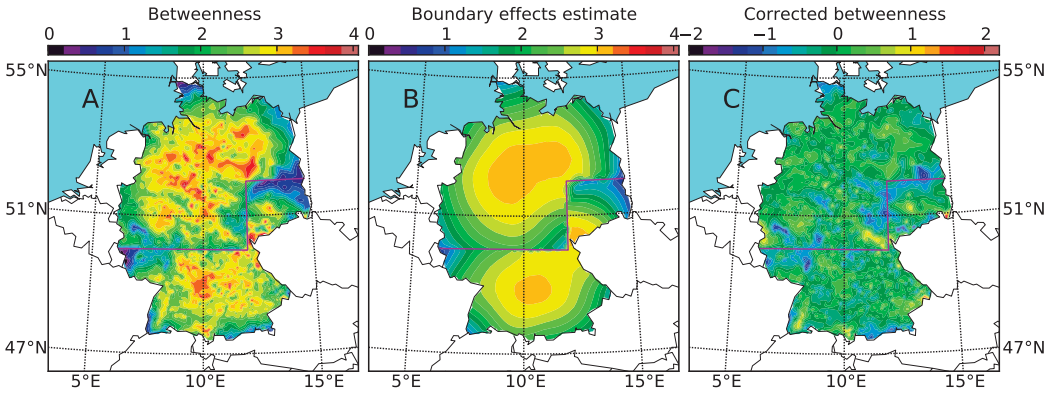

Figure A.6.: Shortest-path betweenness example for two separate regional networks of two subregions of Germany that have the same links within their regions as the network of Fig. A.1. The border between these subregions is marked by the magenta colored line. Shown are the betweenness fields (A), the corresponding boundary effects estimates (B) and the corresponding corrected betweenness fields (C).

A.2. Algorithm for the distribution of the expected number of counts

The distribution $p(k)$ of the expected number of synchronous events assuming a uniformly random distribution of events is given by

$$p(k) = \binom{m}{k} \frac{(l-m)!}{l!} \prod_{i=0}^{k-1} n-i \prod_{j=0}^{m-k-1} l-n-j\,, \tag{A.1}$$

and was computed in this thesis with the following algorithm:

$p(k) = 1.0$
$a = l - n + 2$
if $k \geq m - k$ **then**
  $s = m - k$
else
  $s = k$
end if
for $i \in 1, 2, \ldots, s$ **do**
Ensure: $p(k) \in [10^{-300}, 10^{300}]$
  $p(k) = p(k) * (n - i)$
  $p(k) = p(k)/l$
  $l = l - 1$
  $p(k) = p(k) * (a - i)$
  $p(k) = p(k)/l$
  $l = l - 1$
  $p(k) = p(k) * (m - s + i)$
  $p(k) = p(k)/i$
end for
for $i \in s + 1, s + 2, \ldots, k$ **do**
Ensure: $p(k) \in [10^{-300}, 10^{300}]$
  $p(k) = p(k) * (n - i)$
  $p(k) = p(k)/l$
  $l = l - 1$
end for
for $i \in s + 1, s + 2, \ldots, m - k$ **do**
Ensure: $p(k) \in [10^{-300}, 10^{300}]$
  $p(k) = p(k) * (a - i)$
  $p(k) = p(k)/l$
  $l = l - 1$
end for
return $p(k)$

At the end of the algorithm $p(k)$ holds the probability to have exactly k synchronizations if l is the length of the time series used and $n \geq m$ are the number of events in

the corresponding pair of series.

The ENSURE statement in the above algorithm is for our example of a double precision floating point variable for the probability $p(k)$. For small enough values of m, n and l such a statement is not necessary because $p(k)$ will not overflow. This is the case in applications of this algorithm in this thesis.

A.3. Event rate bias comparison

The initially by Quiroga et al., 2002 proposed Event Synchronization (ES) uses synchronization counts normalized by $\sqrt{mn}$, where m and n are the number of events in each time series. However, in order to obtain a notion of synchronicity that is independent of the number of events, this does not suffice.

The local node-based network measure Strength SG_i of a node i is defined as the sum of all its link weights (3.3). Hence, SG_i gives an estimate on how well heavy precipitation at a given weather station i is synchronized to heavy precipitation at all other weather stations.

In Fig. A.7 we compare this measure for two artificial networks. Both networks are constructed from the same set of random event series, but using a different similarity measure for each network. The left weighted network is constructed using the common ES and the right one using the proposed standardization of synchronization counts. Note the relevance of the bias by the event rate for this study concerning the spatial pattern of network measures.

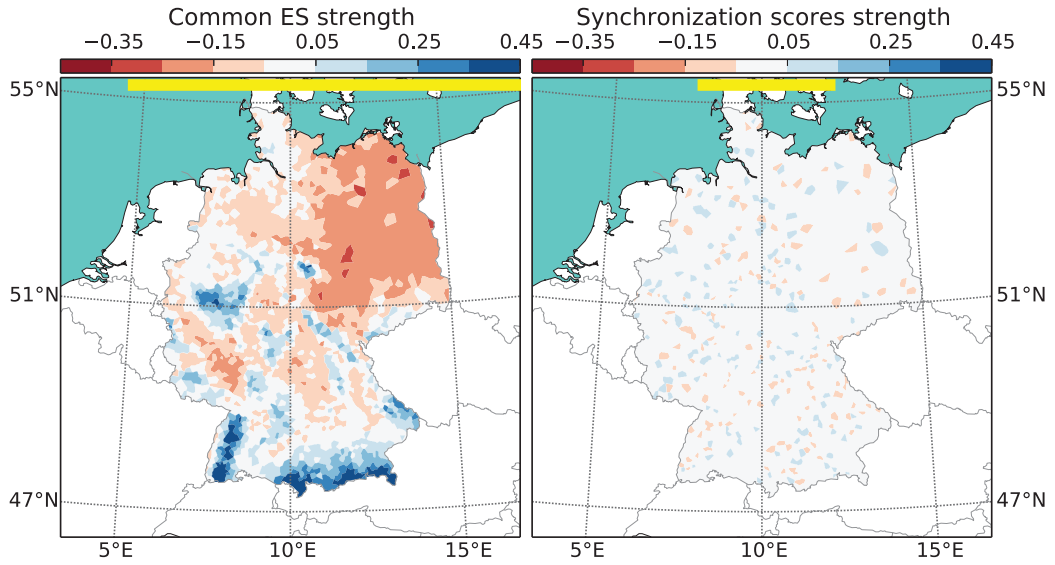

Figure A.7.: Network measure *Strength* for two weighted networks constructed from the same set of random event series of length 1000, but using a different similarity measure. The left weighted network is constructed using the common ES and the right one using the proposed standardization of synchronization counts. The number of random events in a time series corresponds to the actual event rate found in observations for these stations (see Fig. 6.1 (right)). The commonly used ES (left) is biased by the event rate and replicates the intensity pattern as seen in Fig. 6.1. The proposed *synchronization scores* (right) show a weak random strength field. The left and right panels have the same colorbar and the actual range of values for a panel is given by the corresponding yellow bar.

Bibliography

Agustí, P., V. J. Traver, M. J. Marin-Jimenez, and F. Pla (2011). "Exploring Alternative Spatial and Temporal Dense Representations for Action Recognition". In: *Lect. Notes Comp. Sc.* 6855, pp. 364–371. DOI: 10.1007/978-3-642-23678-5_43.

Aldecoa, Rodrigo and Ignacio Marín (2013). "Surprise maximization reveals the community structure of complex networks". In: *Scientific reports* 3.

Amelino-Camelia, Giovanni (2011). "Astrophysics: Shedding light on the fabric of space-time." en. In: *Nature* 478.7370, pp. 466–7. ISSN: 1476-4687. DOI: 10.1038/478466a.

Bajaj, Chandrajit, Peter Djeu, Vinay Siddavanahalli, and Anthony Thane (2004). "Texmol: Interactive visual exploration of large flexible multi-component molecular complexes". In: *Proceedings of the conference on Visualization'04*. IEEE Computer Society, pp. 243–250.

Barlow, Mathew, Matthew Wheeler, Bradfield Lyon, and Heidi Cullen (2005). "Modulation of daily precipitation over southwest Asia by the Madden-Julian oscillation". In: *Monthly weather review* 133.12, pp. 3579–3594.

Barnett, L., E. Di Paolo, and S. Bullock (2007). "Spatially embedded random networks". In: *Phys. Rev. E* 76 (5), p. 056115. DOI: 10.1103/PhysRevE.76.056115.

Barros, Vicente R., Robin Clarke, and Pedro Silva Días (2006). *Climate Change in the La Plata Basin*.

Barthelemy, Marc (2011). "Spatial networks". In: *Physics Reports* 499, pp. 1–101. ISSN: 03701573. DOI: 10.1016/j.physrep.2010.11.002. arXiv: 1010.0302.

Bashan, A., R.P. Bartsch, J.W. Kantelhardt, S. Havlin, and P.C. Ivanov (2012). "Network Physiology reveals relations between network topology and physiological function". In: *Nature Communications* 3, p. 702.

Berbery, E H, V R Barros, and V Olume (2002). "The hydrologic cycle of the La Plata basin in South America". In: *Journal of Hydrometeorology* 3.6, pp. 630–645.

Berezin, Yehiel, Avi Gozolchiani, Oded Guez, and Shlomo Havlin (2012). "Stability of Climate Networks with Time". In: *Scientific Reports* 2, pp. 1–8.

Blasius, Bernd, Amit Huppert, and Lewi Stone (1999). "Complex dynamics and phase synchronization in spatially extended ecological systems". In: *Nature* 399.6734, pp. 354–359.

Boers, N, B Bookhagen, HMJ Barbosa, N Marwan, J Kurths, and JA Marengo (2014a). "Prediction of extreme floods in the eastern Central Andes based on a complex networks approach". In: *Nature communications* 5. DOI: doi:10.1038/ncomms6199.

Boers, Niklas, Bodo Bookhagen, Norbert Marwan, Jürgen Kurths, and José Marengo (2013). "Complex networks identify spatial patterns of extreme rainfall events of

the South American Monsoon System". In: *Geophysical Research Letters* 40.16, pp. 4386–4392.

Boers, Niklas, Reik V. Donner, Bodo Bookhagen, and Jürgen Kurths (2014b). "Complex network analysis helps to identify impacts of the El Niño Southern Oscillation on moisture divergence in South America". In: *Climate Dynamics (accepted)*. DOI: 10.1007/s00382-014-2265-7.

Bookhagen, B and M R Strecker (2008). "Orographic barriers, high-resolution TRMM rainfall, and relief variations along the eastern Andes". In: *Geophysical Research Letters* 35.6, p. L06403.

Bullmore, E. and O. Sporns (2009). "Complex brain networks: graph theoretical analysis of structural and functional systems". In: *Nat. Rev. Neurosci.* 10.3, pp. 186–198.

Carvalho, L M V, C Jones, and B Liebmann (2004a). "The South Atlantic convergence zone: Intensity, form, persistence, and relationships with intraseasonal to interannual activity and extreme rainfall". In: *Journal of Climate* 17.1, pp. 88–108. URL: http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(2004)017<0088:TSACZI>2.0.CO;2.

Carvalho, Leila M. V., Charles Jones, and Brant Liebmann (2002). "Extreme precipitation events in southeastern South America and large-scale convective patterns in the South Atlantic convergence zone". In: *Journal of Climate* 15.17, pp. 2377–2394.

Carvalho, Leila M. V., Ana E. Silva, Charles Jones, Brant Liebmann, Pedro L. Silva Dias, and Humberto R. Rocha (2010). "Moisture transport and intraseasonal variability in the South America monsoon system". In: *Climate Dynamics* 36.9-10, pp. 1865–1880. ISSN: 0930-7575. DOI: 10.1007/s00382-010-0806-2. URL: http://www.springerlink.com/index/10.1007/s00382-010-0806-2.

Carvalho, LMV, C Jones, and B Liebmann (2004b). "The South Atlantic convergence zone: Intensity, form, persistence, and relationships with intraseasonal to interannual activity and extreme rainfall". In: *Journal of Climate* 17.1, pp. 88–108. URL: http://journals.ametsoc.org/doi/pdf/10.1175/1520-0442(2004)017<0088:TSACZI>2.0.CO;2.

Chen, Yu-Han, Bing-Hong Wang, Li-Chao Zhao, Changsong Zhou, and Tao Zhou (2010). "Optimal transport on supply-demand networks". In: *Phys. Rev. E* 81 (6), p. 066105. DOI: 10.1103/PhysRevE.81.066105.

Cohen, Júlia C. P., Maria A. F. Silva Silva Dias, and Carlos A. Nobre (1995). "Environmental Conditions Associated with Amazonian Squall Lines: A Case Study". In: *Monthly Weather Review* 123.11, pp. 3163–3174. URL: http://cat.inist.fr/?aModele=afficheN\&cpsidt=3697315.

Cohen, Reuven, Keren Erez, Daniel ben Avraham, and Shlomo Havlin (2000). "Resilience of the Internet to Random Breakdowns". In: *Phys. Rev. Lett.* 85 (21), pp. 4626–4628. DOI: 10.1103/PhysRevLett.85.4626.

Danila, Bogdan, Yong Yu, John A. Marsh, and Kevin E. Bassler (2006). "Optimal transport on complex networks". In: *Phys. Rev. E* 74 (4), p. 046106. DOI: 10.1103/PhysRevE.74.046106.

Davidsen, J. and M. Paczuski (2005). "Analysis of the spatial distribution between successive earthquakes". In: *Phys. Rev. Lett.* 94.4, p. 48501.

Defays, Daniel (1977). "An efficient algorithm for a complete link method". In: *The Computer Journal* 20.4, pp. 364–366.

Deser, Clara (2000). "On the teleconnectivity of the 'Arctic Oscillation'". In: *Geophysical Research Letters* 27.6, pp. 779–782.

Ding, Qinghua and Bin Wang (2005). "Circumglobal Teleconnection in the Northern Hemisphere Summer*". In: *Journal of Climate* 18.17, pp. 3483–3505.

Dommenget, Dietmar and Mojib Latif (2002). "A cautionary note on the interpretation of EOFs". In: *Journal of Climate* 15.2, pp. 216–225.

Donges, J.F., Y. Zou, N. Marwan, and J. Kurths (2009a). "Complex networks in climate dynamics". In: *Eur. Phys. J. ST* 174.1, pp. 157–179.

– (2009b). "The backbone of the climate network". In: *Europhys. Lett.* 87, p. 48007.

Donges, Jonathan F., Hanna Schultz, Norbert Marwan, Yong Zou, and Juergen Kurths (2011). "Investigating the topology of interacting networks - Theory and application to coupled climate subnetworks". In: *The European Physical Journal B* 84.4, pp. 635–651.

Donner, R V, Y Zou, J F Donges, N Marwan, and J Kurths (2010). "Recurrence networks – A novel paradigm for nonlinear time series analysis". In: *New J. Phys.* 12.3, p. 33025.

Duch, Jordi and Alex Arenas (2005). "Community detection in complex networks using extremal optimization". In: *Physical review E* 72.2, p. 027104.

Durkee, Joshua D., Thomas L. Mote, and J. Marshall Shepherd (2009). "The contribution of mesoscale convective complexes to rainfall across subtropical South America". In: *Journal of Climate* 22.17, pp. 4590–4605. ISSN: 0894-8755. DOI: 10.1175/2009JCLI2858.1. URL: http://journals.ametsoc.org/doi/abs/10.1175/2009JCLI2858.1.

Ebert-Uphoff, I and Y Deng (2012a). "Causal Discovery for Climate Research Using Graphical Models". In: *Journal of Climate* 25.17, pp. 5648–5665.

Ebert-Uphoff, Imme and Yi Deng (2012b). "A new type of climate network based on probabilistic graphical models: Results of boreal winter versus summer". In: *Geophysical Research Letters* 39.19.

Eltahir, Elfatih A. B. and Bras Rafael L. (1993). "Precipitation recycling in the Amazon basin". In: *Quarterly Journal of the Royal Meteorological Society* 120.518, pp. 861–880. ISSN: 00359009. DOI: 10.1002/qj.49712051806. URL: http://doi.wiley.com/10.1002/qj.49712051806.

Erdős, P and A Rényi (1959). "On random graphs I." In: *Publ. Math. Debrecen* 6, pp. 290–297.

Everitt, BS, S Landau, and M Leese (2001). "Cluster analysis. 2001". In: *Arnold, London*.

Feng, Qing Yi and Henk Dijkstra (2014). "Are North Atlantic multidecadal SST anomalies westward propagating?" In: *Geophysical Research Letters*.

Fruchterman, Thomas MJ and Edward M Reingold (1991). "Graph drawing by force-directed placement". In: *Software: Practice and experience* 21.11, pp. 1129–1164.

Gallos, L.K., H.A. Makse, and M. Sigman (2012). "A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks". In: *Proceedings of the National Academy of Sciences* 109.8, pp. 2825–2830.

Gastner, M.T. and M.E.J. Newman (2006). "The spatial structure of networks". In: *Eur. Phys. J. B* 49.2, pp. 247–252.

Gerstengarbe, F-W and PC Werner (2009). "Klimaextreme und ihr Gefährdungspotential für Deutschland". In: *Geographische Rundschau* 9, pp. 12–19.

Ghil, M., MR Allen, MD Dettinger, K. Ide, D. Kondrashov, ME Mann, AW Robertson, A. Saunders, Y. Tian, F. Varadi, et al. (2002). "Advanced spectral methods for climatic time series". In: *Rev. Geophys.* 40.1, p. 1003.

Gill, A_E (1980). "Some simple solutions for heat-induced tropical circulation". In: *Quarterly Journal of the Royal Meteorological Society* 106.449, pp. 447–462.

Girvan, Michelle and Mark EJ Newman (2002). "Community structure in social and biological networks". In: *Proceedings of the National Academy of Sciences* 99.12, pp. 7821–7826.

Gozolchiani, A, K Yamasaki, O Gazit, and S Havlin (2008). "Pattern of climate network blinking links follows El Niño events". In: *EPL (Europhysics Letters)* 83.2, p. 28005.

Gozolchiani, Avi, Shlomo Havlin, and Kazuko Yamasaki (2011). "Emergence of El Niño as an autonomous component in the climate network". In: *Physical Review Letters* 107.14, p. 148501. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.107.148501. URL: http://link.aps.org/doi/10.1103/PhysRevLett.107.148501.

Graves, Sara, Rahul Ramachandran, and Todd Berendes (2012). "Using GLIDER for Knowledge Discovery in Climate Science to Visualize, Analyze and Mine Satellite Imagery". In: *High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:* IEEE, pp. 488–494.

Hannachi, A., I. T. Jolliffe, and D. B. Stephenson (2007). "Empirical orthogonal functions and related techniques in atmospheric science: A review". In: *Int. J. Clim.* 27.9, pp. 1119–1152. ISSN: 08998418.

Hauert, Christoph and Michael Doebeli (2004). "Spatial structure often inhibits the evolution of cooperation in the snowdrift game". In: *Nature* 428.6983, pp. 643–646.

Hawes, Nathan et al. (2013). "Code maps: A scalable visualisation technique for large codebases". In:

Heitzig, Jobst, Jonathan F Donges, Yong Zou, Norbert Marwan, and Jürgen Kurths (2012). "Node-weighted measures for complex networks with spatially embedded, sampled, or differently sized nodes". In: *The European Physical Journal B-Condensed Matter and Complex Systems* 85.1, pp. 1–22.

Henderson, JA and PA Robinson (2011). "Geometric Effects on Complex Network Structure in the Cortex". In: *Phys. Rev. Lett.* 107.1, p. 18102.

Hlinka, Jaroslav, David Hartman, Nikola Jajcay, Martin Vejmelka, Reik Donner, Norbert Marwan, Jürgen Kurths, and Milan Paluš (2014). "Regional and inter-regional effects in evolving climate networks". In: *Nonlinear Processes in Geophysics* 21.2, pp. 451–462.

Holten, Danny and Jarke J Van Wijk (2009). "Force-Directed Edge Bundling for Graph Visualization". In: *Computer Graphics Forum*. Vol. 28. 3. Wiley Online Library, pp. 983–990.

Hoskins, Brian J. and Tercio Ambrizzi (1993). "Rossby Wave Propagation on a Realistic Longitudinally Varying Flow". In: *Journal of the Atmospheric Sciences* 50.12, pp. 1661–1671. URL: http://journals.ametsoc.org/doi/abs/10.1175/1520-0469(1993)050<1661:RWPOAR>2.0.CO;2.

Huang, Norden E, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. Vol. 454. 1971. The Royal Society, pp. 903–995.

Huffman, George, David Bolvin, Eric Nelkin, David Wolff, Robert Adler, Guojun Gu, Yang Hong, Kenneth Bowman, and Erich Stocker (2007). "The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales". In: *J. Hydrometeor* 8.1, pp. 38–55. DOI: 10.1175/JHM560.1.

Itzkovitz, Shalev, Reuven Levitt, Nadav Kashtan, Ron Milo, Michael Itzkovitz, and Uri Alon (2005). "Coarse-graining and self-dissimilarity of complex networks". In: *Physical Review E* 71.1, p. 016127.

Jones, Kate E, Nikkita G Patel, Marc A Levy, Adam Storeygard, Deborah Balk, John L Gittleman, and Peter Daszak (2008). "Global trends in emerging infectious diseases." In: *Nature* 451.7181, pp. 990–3. ISSN: 1476-4687. DOI: 10.1038/nature06536.

Jorgetti, Tatiana, Pedro Leite da Silva Dias, and Edmilson Dias de Freitas (2013). "The relationship between South Atlantic SST and SACZ intensity and positioning". In: *Climate Dynamics* 42.11-12, pp. 3077–3086. ISSN: 0930-7575. DOI: 10.1007/s00382-013-1998-z. URL: http://link.springer.com/10.1007/s00382-013-1998-z.

Kamada, Tomihisa and Satoru Kawai (1989). "An algorithm for drawing general undirected graphs". In: *Information processing letters* 31.1, pp. 7–15.

Kawamura, Ryuichi (1994). "A rotated EOF analysis of global sea surface temperature variability with interannual and interdecadal scales". In: *Journal of physical oceanography* 24.3, pp. 707–715.

Klein, D and G Menz (2003). "Niederschlag im Jahresverlauf". In: *2003): Klima, Pfanzen und Tierwelt. In: Institut für Länderkunde: Nationalatlas der Bundesrepublik Deutschland. Leipzig*.

Klovdahl, A.S., J.J. Potterat, D.E. Woodhouse, J.B. Muth, S.Q. Muth, and W.W. Darrow (1994). "Social networks and infectious disease: The Colorado Springs study". In: *Soc. Sci. Med.* 38.1, pp. 79–88.

Lambert, Antoine, Romain Bourqui, and David Auber (2010). "3D edge bundling for geographical data visualization". In: *Information Visualisation (IV), 2010 14th International Conference*. IEEE, pp. 329–335.

Liebmann, Brant, George N. Kiladis, Carolina S. Vera, Celeste Saulo, and Leila M. V. Carvalho (2004). "Subseasonal variations of rainfall in South America in the

vicinity of the low-level jet east of the Andes and comparison to those in the South Atlantic convergence zone". In: *Journal of Climate* 17.19, pp. 3829–3842.

Ludescher, Josef, Avi Gozolchiani, Mikhail I Bogachev, Armin Bunde, Shlomo Havlin, and Hans Joachim Schellnhuber (2013). "Improved El Niño forecasting by cooperativity detection". In: *Proceedings of the National Academy of Sciences* 110.29, pp. 11742–11745.

MacQueen, James et al. (1967). "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 14. California, USA, pp. 281–297.

Malik, N., B. Bookhagen, N. Marwan, and J. Kurths (2011). "Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks". In: *Clim. Dyn.* 10.1007/s00382-011-1156-4, pp. 1–17.

Malik, Nishant, Bodo Bookhagen, Norbert Marwan, and Jürgen Kurths (2012a). "Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks". In: *Climate dynamics* 39.3-4, pp. 971–987.

Malik, Nishant, Bodo Bookhagen, Norbert Marwan, and Jürgen Kurths (2012b). "Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks". In: *Climate Dynamics* 39.3, pp. 971–987. ISSN: 0930-7575. DOI: 10.1007/s00382-011-1156-4. URL: http://www.springerlink.com/index/10.1007/s00382-011-1156-4.

– (2012c). "Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks". In: *Climate Dynamics* 39.3, pp. 971–987.

Marengo, J A, B Liebmann, A M Grimm, V Misra, P L Silva Dias, I F A Cavalcanti, L M V Carvalho, E H Berbery, T Ambrizzi, C S Vera, Others, P L Silva Dias, A C Saulo, J Nogues-paegle, E Zipser, A Seth, and L M Alves (2012). "Recent developments on the South American monsoon system". In: *International Journal of Climatology* 32.1, pp. 1–21. DOI: 10.1002/joc.2254.

Marengo, José A., Wagner R. Soares, Celeste Saulo, and Matilde Nicolini (2004). "Climatology of the low-level jet east of the Andes as derived from the NCEP-NCAR reanalyses: Characteristics and temporal variability". In: *Journal of Climate* 17.12, pp. 2261–2280.

Marengo, José A., Maria Valverde, and G Obregon (2013). "Observed and projected changes in rainfall extremes in the Metropolitan Area of São Paulo". In: *Climate Research* 57, pp. 61–72. ISSN: 0936-577X. DOI: 10.3354/cr01160. URL: http://www.int-res.com/abstracts/cr/v57/n1/p61-72/.

Marwan, N., J. Kurths, and P. Saparin (2007a). "Generalised Recurrence Plot Analysis for Spatial Data". In: *Phys. Lett. A* 360.4–5, pp. 545–551. DOI: 10.1016/j.physleta.2006.08.058.

Marwan, N, M.Carmen Romano, M Thiel, and J Kurths (2007b). "Recurrence plots for the analysis of complex systems". In: *Physics Reports* 438.5-6, pp. 237–329. ISSN: 03701573. DOI: 10.1016/j.physrep.2006.11.001.

Marwan, N., J. F. Donges, Y. Zou, R. V. Donner, and J. Kurths (2009). "Complex network approach for recurrence analysis of time series". In: *Phys. Lett. A* 373.46, pp. 4246–4254. DOI: 10.1016/j.physleta.2009.09.042.

Menck, Peter J, Jobst Heitzig, Jürgen Kurths, and Hans Joachim Schellnhuber (2014). "How dead ends undermine power grid stability". In: *Nature communications* 5.

Monahan, Adam H, John C Fyfe, Maarten HP Ambaum, David B Stephenson, and Gerald R North (2009). "Empirical orthogonal functions: The medium is the message". In: *Journal of Climate* 22.24, pp. 6501–6514.

Mudelsee, Manfred, Michael Börngen, Gerd Tetzlaff, and Uwe Grünewald (2004). "Extreme floods in central Europe over the past 500 years: Role of cyclone pathway 'Zugstrasse Vb'". In: *Journal of Geophysical Research: Atmospheres (1984–2012)* 109.D23.

N. M. Ferguson, G. P. Garnett (2000). "More Realistic Models of Sexually Transmitted Disease Transmission Dynamics: Sexual Partnership Networks, Pair Models, and Moment Closure". In: *Sex. Transm. Dis.* 27 (10), pp. 600–609.

Newman, Mark (2010). *Networks: An Introduction*. 1st ed. Oxford University Press, USA. ISBN: 0199206651.

Newman, Mark EJ (2005). "A measure of betweenness centrality based on random walks". In: *Social networks* 27.1, pp. 39–54.

Newman, Mark EJ and Michelle Girvan (2004). "Finding and evaluating community structure in networks". In: *Physical review E* 69.2, p. 026113.

Newman, Mark EJ and Juyong Park (2003). "Why social networks are different from other types of networks". In: *Physical Review E* 68.3, p. 036122.

Newman, M.E.J. (2003). "The Structure and Function of Complex Networks". In: *SIAM review* 45.2, pp. 167–256.

Nicolini, Matilde, A. Celeste Saulo, Juan C. Torres, and Paola Salio (2002). "Enhanced precipitation over southeastern South America related to strong low-level jet events during austral warm season". In: *Meteorologica, Special Issue for the South American Monsoon System* 27.1-2, pp. 59–69.

Nogués-Paegle, Julia and Kingtse C. Mo (1997). "Alternating wet and dry conditions over South America during summer". In: *Monthly Weather Review* 125.2, pp. 279–291.

Paluš, M., D. Hartman, J. Hlinka, and M. Vejmelka (2011). "Discerning connectivity from dynamics in climate networks". In: *Nonlin. Processes Geophys.* 18, pp. 751–763.

Pastor-Satorras, Romualdo, Miguel Rubí, and Albert Diaz-Guilera (2003). *Statistical mechanics of complex networks*. Vol. 625. Springer.

Pearson, Karl (1901). "LIII. On lines and planes of closest fit to systems of points in space". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11, pp. 559–572.

Pupyrev, Sergey, Lev Nachmanson, and Michael Kaufmann (2011). "Improving layered graph layouts with edge bundling". In: *Graph Drawing*. Springer, pp. 329–340.

Quiroga, Rodrigo Quian, Thomas Kreuz, and Peter Grassberger (2002). "Event synchronization: A simple and fast method to measure synchronicity and time delay patterns". In: *Physical Review E* 66.4, p. 41904.

Reynolds, Richard W, Nick A Rayner, Thomas M Smith, Diane C Stokes, and Wanqiu Wang (2002). "An improved in situ and satellite SST analysis for climate". In: *Journal of climate* 15.13, pp. 1609–1625.

Rheinwalt, Aljoscha, Norbert Marwan, Jürgen Kurths, Peter Werner, and Friedrich-Wilhelm Gerstengarbe (2012). "Boundary effects in network measures of spatially embedded networks". In: *EPL (Europhysics Letters)* 100.2, p. 28002.

Rienecker, Michele M, Max J Suarez, Ronald Gelaro, Ricardo Todling, Julio Bacmeister, Emily Liu, Michael G Bosilovich, Siegfried D Schubert, Lawrence Takacs, Gi-Kong Kim, Stephen Bloom, Junye Chen, Douglas Collins, Austin Conaty, Arlindo Da Silva, Wei Gu, Joanna Joiner, Randal D Koster, Robert Lucchesi, Andrea Molod, Tommy Owens, Steven Pawson, Philip Pegion, Christopher R Redder, Rolf Reichle, Franklin R Robertson, Albert G Ruddick, Meta Sienkiewicz, and Jack Woollen (2011). "MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications". In: *Journal of Climate* 24, pp. 3624–3648. ISSN: 08948755. DOI: 10.1175/JCLI-D-11-00015.1.

Rodwell, Mark J. and Brian J. Hoskins (2001). "Subtropical anticyclones and summer monsoons". In: *Journal of Climate* 14.15, pp. 3192–3211. URL: http://journals.ametsoc.org/doi/full/10.1175/1520-0442(2001)014<3192:SAASM>2.0.CO;2.

Romano, M Carmen, Marco Thiel, Jürgen Kurths, Konstantin Mergenthaler, and Ralf Engbert (2009). "Hypothesis test for synchronization: twin surrogates revisited." In: *Chaos (Woodbury, N.Y.)* 19.1, p. 015108. DOI: 10.1063/1.3072784.

Runge, Jakob, Jobst Heitzig, Norbert Marwan, and Jürgen Kurths (2012). "Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy". In: *Phys. Rev. E* 86.6, p. 61121.

Salio, Paola, Matilde Nicolini, and Edward J. Zipser (2007a). "Mesoscale Convective Systems over Southeastern South America and Their Relationship with the South American Low-Level Jet". In: *Monthly Weather Review* 135.4, pp. 1290–1309. ISSN: 0027-0644. DOI: 10.1175/MWR3305.1. URL: http://journals.ametsoc.org/doi/abs/10.1175/MWR3305.1.

– (2007b). "Mesoscale convective systems over southeastern South America and their relationship with the South American low-level jet". In: *Monthly Weather Review* 135.4, pp. 1290–1309. ISSN: 0027-0644. DOI: 10.1175/MWR3305.1. URL: http://journals.ametsoc.org/doi/abs/10.1175/MWR3305.1.

Santos, Emanuele, Jorge Poco, Yaxing Wei, Shishi Liu, Bob Cook, Dean N Williams, and Claudio T Silva (2013). "UV-CDAT: Analyzing Climate Datasets from a User's Perspective". In: *Computing in Science & Engineering* 15.1, pp. 94–103.

Sarma, Anita, Larry Maccherone, Patrick Wagstrom, and James Herbsleb (2009). "Tesseract: Interactive visual exploration of socio-technical relationships in software development". In: *Software Engineering, 2009. ICSE 2009. IEEE 31st International Conference on*. IEEE, pp. 23–33.

Schmah, Tanya, Norbert Marwan, Jesper Skovhus Thomsen, and Peter Saparin (2011). "Long range node-strut analysis of trabecular bone microarchitecture." en. In: *Medical Physics* 38.9, pp. 5003–11.

Schreiber, Thomas and Andreas Schmitz (1996). "Improved surrogate data for nonlinearity tests". In: *Physical Review Letters* 77.4, p. 635.

– (2000). "Surrogate time series". In: *Physica D: Nonlinear Phenomena* 142.3, pp. 346–382.

Seluchi, Marcelo E. and René D. Garreaud (2006). "Influence of the subtropical Andes on baroclinic disturbances: A cold front case study". In: *Monthly Weather Review* 134.11, pp. 3317–3335. URL: http://journals.ametsoc.org/doi/pdf/10.1175/MWR3247.1.

Siqueira, Jose R. and L. A. Toledo Machado (2004). "Influence of the frontal systems on the day-to-day convection variability over South America". In: *Journal of Climate* 17.9, pp. 1754–1766. URL: http://journals.ametsoc.org/doi/abs/10.1175/1520-0442(2004)017<1754:IOTFSO>2.0.CO;2.

Solomonoff, Ray and Anatol Rapoport (1951). "Connectivity of random nets". In: *The bulletin of mathematical biophysics* 13.2, pp. 107–117.

Sporns, O., D.R. Chialvo, M. Kaiser, and C.C. Hilgetag (2004). "Organization, development and function of complex brain networks". In: *Trends Cognit. Sci.* 8.9, pp. 418–425.

Squartini, Tiziano and Diego Garlaschelli (2011). "Analytical maximum-likelihood method to detect patterns in real networks". In: *New Journal of Physics* 13.8, p. 083001.

Stam, C.J., J.C. Reijneveld, et al. (2007). "Graph theoretical analysis of complex networks in the brain". In: *Nonlin. Biomed. Phys.* 1.1, p. 3.

Steinhaeuser, K., N.V. Chawla, and A.R. Ganguly (2010). "An exploration of climate data using complex networks". In: *ACM SIGKDD Explorations Newsletter* 12.1, pp. 25–32.

Steinhaeuser, K, A R Ganguly, and N V Chawla (2012a). "Multivariate and multiscale dependence in the global climate system revealed through complex networks". In: *Climate Dynamics* 39.3-4, pp. 889–895.

Steinhaeuser, Karsten, Auroop R. Ganguly, and Nitesh V. Chawla (2012b). "Multivariate and multiscale dependence in the global climate system revealed through complex networks". In: *Climate Dynamics* 39.3-4, pp. 889–895. ISSN: 0930-7575. DOI: 10.1007/s00382-011-1135-9. URL: http://www.springerlink.com/index/10.1007/s00382-011-1135-9.

Thiel, M, M. C Romano, J Kurths, M Rolfs, and R Kliegl (2006). "Twin surrogates to test for complex synchronisation". In: *Europhysics Letters (EPL)* 75.4, pp. 535–541. ISSN: 0295-5075. DOI: 10.1209/epl/i2006-10147-0.

Thiel, M, M C Romano, J Kurths, M Rolfs, and R Kliegl (2008). "Generating surrogates from recurrences." In: *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 366.1865, pp. 545–57. ISSN: 1364-503X. DOI: 10.1098/rsta.2007.2109.

Tominski, Christian, James Abello, and Heidrun Schumann (2009). "CGV – An interactive graph visualization system". In: *Computers & Graphics* 33.6, pp. 660–678.

Torrence, Christopher and Gilbert P. Compo (1998). "A Practical Guide to Wavelet Analysis". In: *Bull. Amer. Met. Soc.* 79.1, pp. 61–78. ISSN: 15200477.

Trenberth, Kevin E, David P Stepaniak, and Julie M Caron (2000). "The global monsoon as seen through the divergent atmospheric circulation". In: *Journal of Climate* 13.22, pp. 3969–3993.

Tsonis, A A and K L Swanson (2008). "Topology and predictability of El Niño and La Niña networks". In: *Physical Review Letters* 100.22, p. 228502.

Tsonis, A.A. and P.J. Roebber (2004a). "The architecture of the climate network". In: *Physica A* 333.0, pp. 497–504.

Tsonis, AA and PJ Roebber (2004b). "The architecture of the climate network". In: *Physica A: Statistical Mechanics and its Applications* 333, pp. 497–504.

Tsonis, A.A., K.L. Swanson, and P.J. Roebber (2006a). "What do networks have to do with climate?" In: *Bull. Am. Meteorol. Soc.* 87.5, pp. 585–596.

Tsonis, Anastasios A. and Paul J. Roebber (2004c). "The architecture of the climate network". In: *Physica A: Statistical Mechanics and its Applications* 333, pp. 497–504. ISSN: 03784371. DOI: 10.1016/j.physa.2003.10.045. URL: http://www.sciencedirect.com/science/article/pii/S0378437103009646.

Tsonis, Anastasios A, Kyle L Swanson, and Paul J Roebber (2006b). "What do networks have to do with climate?" In: *Bulletin of the American Meteorological Society* 87.5, pp. 585–595.

Tuckwell, H. C., Laurent Toubiana, and Jean-Francois Vibert (1998). "Spatial epidemic network models with viral dynamics". In: *Phys. Rev. E* 57 (2), pp. 2163–2169. DOI: 10.1103/PhysRevE.57.2163.

Van Der Mheen, Mirjam, Henk A. Dijkstra, Avi Gozolchiani, Matthijs Den Toom, Qingyi Feng, Jürgen Kurths, and Emilio Hernandez-Garcia (2013). "Interaction network based early warning indicators for the Atlantic MOC collapse". In: *Geophysical Research Letters* 40.11, pp. 2714–2719. ISSN: 00948276. DOI: 10.1002/grl.50515.

Vera, C., W. Higgins, J. Amador, T. Ambrizzi, R. D. Garreaud, D. Gochis, D. Gutzler, D Lettenmaier, J. A. Marengo, C. R. Mechoso, J. Nogues-Paegle, P.L. Silva Dias, and C. Zhang (2006). "Toward a unified view of the American monsoon systems". In: *Journal of Climate* 19.20, pp. 4977–5000. URL: http://journals.ametsoc.org/doi/pdf/10.1175/JCLI3896.1.

Webster, Peter J, Vo Oo Magana, TN Palmer, J Shukla, RA Tomas, M u Yanai, and T Yasunari (1998). "Monsoons: Processes, predictability, and the prospects for prediction". In: *Journal of Geophysical Research: Oceans (1978–2012)* 103.C7, pp. 14451–14510.

Webster, Peter J, Andrew M Moore, Johannes P Loschnigg, and Robert R Leben (1999). "Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–98". In: *Nature* 401.6751, pp. 356–360.

Werner, PC and FW Gerstengarbe (2010). "Katalog der Großwetterlagen Europas (1881–2009) nach Paul Hess und Helmut Brezowsky". In: *Potsdam Institut für Klimafolgenforschung, Postdam, Germany*.

Wiedermann, Marc, Jonathan F Donges, Jobst Heitzig, and Jürgen Kurths (2013). "Node-weighted interacting network measures improve the representation of real-world complex systems". In: *EPL (Europhysics Letters)* 102.2, p. 28007.

Wiese, Roland, Markus Eiglsperger, and Michael Kaufmann (2004). *yfiles – visualization and automatic layout of graphs*. Springer.

Xia, Jianguo, Maia J Benner, and Robert EW Hancock (2014). "NetworkAnalyst-integrative approaches for protein–protein interaction network analysis and visual exploration". In: *Nucleic acids research* 42.W1, W167–W174.

Yamasaki, K., A. Gozolchiani, and S. Havlin (2008a). "Climate networks around the globe are significantly affected by El Nino". In: *Phys. Rev. Lett.* 100.22, p. 228501.

Yamasaki, Kazuko, Avi Gozolchiani, and Shlomo Havlin (2008b). "Climate networks around the globe are significantly affected by El Nino". In: *Physical Review Letters* 100.22, p. 228501.

Yatagai, Akiyo, Kenji Kamiguchi, Osamu Arakawa, Atsushi Hamada, Natsuko Yasutomi, and Akio Kitoh (2012). "APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges". In: *Bulletin of the American Meteorological Society* 93.9, pp. 1401–1415.

Zamora-López, Gorka, Changsong Zhou, and Jürgen Kurths (2010). "Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks." In: *Frontiers in neuroinformatics* 4, p. 1. ISSN: 1662-5196. DOI: 10.3389/neuro.11.001.2010.

Zemp, DC, M Wiedermann, J Kurths, A Rammig, and JF Donges (2014). "Node-weighted measures for complex networks with directed and weighted edges for studying continental moisture recycling". In: *EPL (Europhysics Letters)* 107.5, p. 58005.

Zhou, C., L. Zemanová, G. Zamora, C.C. Hilgetag, and J. Kurths (2006). "Hierarchical organization unveiled by functional connectivity in complex brain networks". In: *Phys. Rev. Lett.* 97.23, p. 238103.

Zhou, Jiayu and K.-M. Lau (1998). "Does a monsoon climate exist over South America?" In: *Journal of Climate* 11.5, pp. 1020–1040.

Zipser, Edward J., Daniel J. Cecil, Chuntao Liu, Stephen W. Nesbitt, and David P. Yorty (2006). "Where Are the Most Intense Thunderstorms on Earth?" In: *Bulletin of the American Meteorological Society* 87.8, pp. 1057–1071. ISSN: 0003-0007. DOI: 10.1175/BAMS-87-8-1057. URL: http://journals.ametsoc.org/doi/abs/10.1175/BAMS-87-8-1057.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 3. Februar 2016 Aljoscha Rheinwalt