

# Complex Network Analysis of Extreme Rainfall in South America

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium

(Dr. rer. nat.)

im Fach Physik

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

von

**Dipl.-Phys. Niklas Felix Boers M.Sc.**

Präsident der Humboldt-Universität zu Berlin:

Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:

Prof. Dr. Elmar Kulke

Gutachter:

1. Jürgen Kurths
2. José A. Marengo
3. Thomas Stemler

**Tag der mündlichen Prüfung:** 30. April 2015



*To Elfi.*



## Abstract

Based on the theory of networks, a general framework is developed to study collective synchronization phenomena of extreme events in complex systems. The method relies on observational time series encoding the variability of the single parts of the system, and is intended to reveal emerging patterns of extreme event synchronization on the macroscopic level. For this purpose, the time series obtained from an interactive system under consideration are identified with network nodes, and the possibly delayed and non-linear interdependence of extreme events in different time series is represented by network links connecting the nodes. In this way, the complex internal synchronization structure of the system becomes accessible in terms of the topology of the network, which can be analyzed by introducing suitable network measures. The methodology can thus be seen as a tool for exploring empirical or simulation-derived data, and can form the basis for the development of scientific hypotheses concerning the physical mechanisms underlying the emergent synchronization patterns. But in addition to the pure analysis of a given system, this tool can also be used for statistical prediction of extreme events, given that the system exhibits sufficiently concise synchronization patterns.

The methodology is applied to satellite-derived rainfall time series of high spatiotemporal resolution in order to investigate the collective dynamics of extreme rainfall events in South America. The purpose of this application is threefold: First, it is shown how the methodology can be used for climatic analysis by revealing climatological mechanism from the spatial patterns exhibited by different network measures. This is partly intended to serve as a proof of concept, but also adds new insights into the functioning of the climate system in situations where traditional techniques to study spatial patterns of co-variability of climatic observables are not applicable. This is the case for spatial characteristics of extreme event synchronicity, which cannot be derived nor analyzed on the basis of linear covariance measures. Second, networks encoding the synchronization structure of extreme rainfall events are constructed in a way that resolves their temporal order. These directed networks are used to assess the predictability of extreme rainfall at the eastern slopes of the Andes, which are frequently exposed to rainfall-induced natural hazards in form of floods and landslides. By introducing the concept of network divergence, sink and source regions of extreme events can be identified, allowing to track their directed synchronization pathways through the network. On this basis, a climatological mechanism is revealed that causes large rainfall clusters to propagate from southeastern South America towards the Central Andes. A simple statistical forecast rule is finally derived from these insights, predicting substantial fractions of extreme rainfall events in the Central Andes. Third, the methodology and the insights developed in the first two steps are used to evaluate the dynamical representation of extreme events in different datasets, and in particular their dynamical implementation in three state of the art climate models.



## Zusammenfassung

Basierend auf der Theorie von Netzwerken wird zunächst ein allgemeines Rahmenwerk entwickelt, um kollektive Synchronisationsphänomene von Extremereignissen in komplexen Systemen zu studieren. Die Methode vergleicht die Variabilität der einzelnen Teile des Systems auf Grundlage von Beobachtungszeitreihen mit dem Ziel, emergente Synchronisationsmuster von Extremereignissen auf makroskopischer Ebene aufzudecken. Zu diesem Zweck werden die einzelnen Zeitreihen eines interaktiven Systems mit den Knoten eines Netzwerks identifiziert und die Abhängigkeiten zwischen diesen durch die Kanten des Netzwerks dargestellt, wobei mögliche Nicht-Linearitäten sowie variierende Verzögerungen zwischen Extremereignissen in verschiedenen Zeitreihen speziell berücksichtigt werden. Die komplexe interne Synchronisationsstruktur des Systems wird so in Form der Netzwerktopologie mathematisch zugänglich gemacht und kann durch die Einführung geeigneter Netzwerkmaße analysiert werden. Die entwickelte Methode stellt somit ein Werkzeug zur Untersuchung von empirischen oder aus Simulationen gewonnenen Datensätzen dar und kann den Ausgangspunkt zur Entwicklung wissenschaftlicher Hypothesen über die physikalischen Mechanismen, die den emergenten Synchronisationsmustern zu Grunde liegen, bilden. Über die reine Analyse eines Systems hinaus kann diese Methodik jedoch auch zur statistischen Vorhersage von Extremereignissen verwendet werden, sofern ausreichend prägnante Synchronisationsmuster vorliegen.

Die Methode wird im Folgenden auf räumlich und zeitlich hochaufgelöste Regendaten aus Satellitenmessungen angewendet, um die kollektive Dynamik extremer Regenereignisse in Südamerika zu untersuchen. Diese Anwendung verfolgt drei Ziele: Erstens wird gezeigt, wie die hier entwickelte Methode zur klimatologischen Analyse verwendet werden kann. Dazu werden klimatologische Mechanismen auf Grundlage der räumlichen Muster verschiedener Netzwerkmaße aufgedeckt. Dies soll einerseits als Nachweis dazu dienen, dass die Methode ihren Zweck erfüllt, trägt aber andererseits auch zum Verständnis des Klimasystems in Situationen bei, in denen traditionelle Techniken zur Studie räumlicher Kovariabilitätsmuster klimatischer Observablen nicht anwendbar sind. Für räumliche Charakteristika der Synchronisation extremer Ereignisse ist dies der Fall, da diese auf Grundlage linearer Kovarianzmaße weder ableitbar noch analysierbar sind. Zweitens werden Netzwerke konstruiert, die die Synchronisationsstruktur extremer Regenereignisse unter Berücksichtigung ihrer zeitlichen Reihenfolge kodieren. Diese gerichteten Netzwerke werden verwendet, um die Vorhersagbarkeit extremer Regenereignisse an den Osthängen der Anden einzuschätzen, welche häufig zu Fluten und Landrutschen in diesen Regionen führen. Durch die Einführung des Konzeptes der Netzwerkdivergenz können Quellen und Senken von Extremereignissen identifiziert werden. Dies erlaubt es, die gerichteten Netzwerkpfade, entlang derer Extremereignisse synchronisieren, nachzuverfolgen. Auf dieser Grundlage wird ein klimatologischer Mechanismus entdeckt, der für die Propagation von großen Regensystemen vom Südosten Südamerikas bis in die Zentralanden verantwortlich ist. Aus diesen Erkenntnissen wird eine statistische Regel gewonnen, die beträchtliche Anteile der extremen Regenereignisse in den Zentralanden vorhersagt. Drittens werden die bis dahin entwickelten Methoden und gewonnenen Einsichten dazu verwendet, die Darstellung extremer Regenereignisse in verschiedenen Datensätzen zu vergleichen. Insbesondere

wird in diesem Kontext die Implementierung solcher Ereignisse in drei gängigen Klimamodellen evaluiert.

# List of publications

This dissertation is partly based on the following publications. The identifiers given below (e.g. P1) are cited in the text to highlight passages that are connected to these studies.

- P1 **N. Boers**, B. Bookhagen, N. Marwan, J. Kurths, J. Marengo, *Complex networks identify spatial patterns of extreme rainfall events of the South American Monsoon System*, *Geophysical Research Letters* **40**, 4386–4392 (2013).
- 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* **41**, 73977405 (2014).
- P3 A. Rheinwalt, **N. Boers**, N. Marwan, J. Kurths, F. Gerstengarbe, P. Werner, *Non-Linear Time Series Analysis of Precipitation Events Using Regional Climate Networks for the Region of Germany*, *Climate Dynamics* (accepted).
- P4 **N. Boers**, B. Bookhagen, N. Marwan, J. Kurths, *Spatiotemporal Characteristics and Synchronization of Extreme Rainfall in South America with Focus on the Andes Mountain Range*, *Climate Dynamics* (accepted).
- P5 **N. Boers**, B. Bookhagen, H.M.J. Barbosa, N. Marwan, J. Kurths, J. Marengo, *Prediction of Extreme Floods in the Eastern Central Andes based on a Complex Networks Approach*, *Nature Communications* 5:5199 (2014).
- P6 **N. Boers**, B. Bookhagen, H.M.J. Barbosa, N. Marwan, J. Kurths, J. Marengo, *Propagation of strong rainfall events from southeastern South America to the Central Andes* (in revision).
- P7 **N. Boers**, B. Bookhagen, J. Marengo, N. Marwan, J. v. Storch, J. Kurths, *Extreme rainfall of the South American monsoon system: A dataset comparison using complex networks*, *Journal of Climate* 28:3, 1031-1056 (2015).
- P8 **N. Boers**, R. Donner, B. Bookhagen, J. Kurths, *Complex network analysis helps to identify impacts of the El Niño Southern Oscillation on moisture divergence in South America* *Climate Dynamics* (accepted).

Berlin, May 29, 2015



# Acknowledgements

I am deeply grateful to Jürgen Kurths for supervising and promoting me in all aspects of scientific life. Furthermore, I want to thank José Marengo for his great support, in particular during my stay at the Centre for Earth System Research in Brazil. I am very thankful to Bodo Bookhagen for all his efforts, for his patience, for his hospitality in Santa Barbara, and in particular for helping me out whenever my very limited knowledge of the geosciences did just not suffice to make sense of the plots. Seriously, what would I have done! In addition, I am obliged to Norbert Marwan for his great advices, and for having an answer to basically any question on time series analysis.

I am indebted to the Potsdam Institute for Climate Impact Research, the German Research Foundation, and the International Research and Training Group “Dynamical Phenomena in Complex Networks” for giving me the opportunity to make a living from something that is most of the time a lot of fun.

Many thanks go to Bedartha Goswami, Aljoscha Rheinwalt, and Dominik Traxl. For countless discussions, for the great times we had at conferences, for being extremely critical at times, but most of all for their friendship. To be continued ... I am truly looking forward to that!

Finally, I want to thank my family. Don't get me started what for. They're just great!



# Contents

List of publications	ix
Acknowledgements	x
List of Figures	xvii
List of Tables	xxi
List of frequently used symbols and abbreviations	1
<b>1. Introduction</b>	<b>3</b>
1.1. Motivation . . . . .	3
1.2. Climatological Setting: The South American Monsoon System . . . .	6
1.3. Arrangement of this thesis . . . . .	9
<b>I. Theoretical Framework</b>	<b>11</b>
<b>2. Measures of Similarity</b>	<b>13</b>
2.1. Introduction . . . . .	13
2.2. Event Synchronization . . . . .	14
2.3. Comparison between Pearson's correlation coefficient and Event Syn- chronization . . . . .	17
<b>3. Network Theory</b>	<b>19</b>
3.1. Introduction . . . . .	19
3.2. Mathematical representation of networks . . . . .	19
3.3. Construction of climate networks from empirical data . . . . .	20
3.4. Complex network measures . . . . .	22
3.5. The influence of the spatial embedding . . . . .	26
<b>II. Applications</b>	<b>29</b>
<b>4. Spatial Patterns of Extreme Rainfall Co-Variability</b>	<b>31</b>
4.1. Summary . . . . .	31
4.2. Introduction . . . . .	31
4.3. Data . . . . .	32

*Contents*

4.4. Methods . . . . .	34
4.5. Results . . . . .	36
4.6. Discussion . . . . .	40
4.7. Conclusion . . . . .	41
<b>5. Extreme Rainfall Associated with the South American Rainfall Dipole</b>	<b>43</b>
5.1. Summary . . . . .	43
5.2. Introduction . . . . .	43
5.3. Data . . . . .	45
5.4. Methods . . . . .	46
5.5. Results . . . . .	47
5.6. Discussion . . . . .	51
5.7. Conclusion . . . . .	53
<b>6. Spatiotemporal Connectivity of Extreme Rainfall in the Andes</b>	<b>55</b>
6.1. Summary . . . . .	55
6.2. Introduction . . . . .	56
6.3. Data . . . . .	58
6.4. Methods . . . . .	58
6.5. Results . . . . .	61
6.6. Discussion . . . . .	71
6.7. Conclusion . . . . .	73
<b>7. Prediction of Extreme Floods in the Eastern Central Andes</b>	<b>75</b>
7.1. Summary . . . . .	75
7.2. Introduction . . . . .	75
7.3. Data . . . . .	76
7.4. Methods . . . . .	77
7.5. Results . . . . .	78
7.6. Discussion . . . . .	85
7.7. Conclusion . . . . .	85
<b>8. Dataset and Model Intercomparison</b>	<b>87</b>
8.1. Summary . . . . .	87
8.2. Introduction . . . . .	87
8.3. Data . . . . .	89
8.4. Methods . . . . .	90
8.5. Results and Discussion . . . . .	91
8.6. Conclusion . . . . .	116
<b>9. Impacts of the El Niño Southern Oscillation on Extreme Moisture Divergence</b>	<b>119</b>
9.1. Summary . . . . .	119
9.2. Introduction . . . . .	119

9.3. Data . . . . .	120
9.4. Methods . . . . .	122
9.5. Results . . . . .	125
9.6. Discussion . . . . .	135
9.7. Conclusion . . . . .	141
<b>10. Conclusion</b>	<b>143</b>
10.1. Contributions of this thesis . . . . .	143
10.2. Outlook . . . . .	146
<b>Appendix</b>	<b>148</b>
<b>A. Additional figures for chapter 4</b>	<b>151</b>
<b>B. Additional figures for chapter 5</b>	<b>163</b>
<b>C. Additional figures for chapter 6</b>	<b>167</b>
<b>D. Additional figures for chapter 7</b>	<b>171</b>
<b>Bibliography</b>	<b>187</b>



# List of Figures

1.1.	Topography of South America and key features of the South American monsoon system. . . . .	6
1.2.	Second EOF of daily rainfall during the monsoon system. . . . .	7
1.3.	The Influence of ENSO on monthly rainfall in South America. . . . .	8
2.1.	Influence of the event rates on the outcome of Event Synchronization. . . . .	17
2.2.	Comparison of Pearson’s correlation coefficient and Event Synchronization. . . . .	18
3.1.	Exemplary visualization of a climate network derived from rainfall data over South America. . . . .	22
3.2.	Artificial example network demonstrating the network measures betweenness centrality, clustering coefficient, mean geographical distance, and long-ranged directedness . . . . .	25
4.1.	The South American monsoon system and rainfall climatology for the monsoon season. . . . .	33
4.2.	The influence of the spatial embedding on the spatial distribution of various network measures. . . . .	37
4.3.	Complex Network measures reveal key features of the South American monsoon system. . . . .	38
4.4.	Difference between long-ranged directedness for rainfall events above the 95th and 90th percentiles. . . . .	39
5.1.	Topography and time series of the number of extreme events in the two study regions. . . . .	44
5.2.	Composites and anomalies of rainfall, geopotential height, and wind for the two phases of the South American rainfall dipole. . . . .	48
5.3.	Network measures degree (DR) and directionality (DR) for the two phases of the South American rainfall dipole. . . . .	49
5.4.	Difference between Degree fields for the two phases of the South American rainfall dipole. . . . .	50
6.1.	Geographical and hydrological setting. . . . .	57
6.2.	Mean, median, and 90th percentiles of rainfall distributions. . . . .	62
6.3.	Number of events of the four types LSE, LLE, SEE, and SLE defined in the text. . . . .	63

List of Figures

6.4.	Total fraction of total DJF rainfall accounted for by the four types of events. . . . .	65
6.5.	Fraction of total DJF rainfall accounted for by each burst of the four types of events. . . . .	66
6.6.	90th percentiles of rainfall cluster sizes. . . . .	68
6.7.	Regional Connectivity of the Altiplano-Puna Plateau. . . . .	69
6.8.	Regional Connectivity of the main catchments at the Andean foothills. . . . .	70
7.1.	Topographical and Climatological Setting. . . . .	79
7.2.	Network divergence, Regional Connectivity of southeastern South America (SESA), as well as extreme event propagation from SESA to the eastern Central Andes. . . . .	81
7.3.	Atmospheric conditions for propagation and non-propagation times. . . . .	82
8.1.	Topography of South America and key features of the South American Monsoon System. . . . .	90
8.2.	Mean daily rainfall. . . . .	93
8.3.	90th percentiles of daily rainfall. . . . .	95
8.4.	95th percentiles of daily rainfall. . . . .	96
8.5.	Fraction of total DJF rainfall contributed by events above the 90th percentile. . . . .	97
8.6.	Difference between 90th and the 50th percentiles. . . . .	99
8.7.	Difference between 95th and the 90th percentiles. . . . .	100
8.8.	Rainfall distributions. . . . .	102
8.9.	Degree for events above the 90th percentile. . . . .	104
8.10.	Betweenness centrality for events above the 90th percentile. . . . .	106
8.11.	Clustering for events above the 90th percentile. . . . .	108
8.12.	Regional Connectivity of the central Amazon Basin. . . . .	111
8.13.	Regional Connectivity of southeastern South America. . . . .	112
8.14.	Regional Connectivity of the South Atlantic Convergence Zone. . . . .	113
8.15.	Differences of network measures for various datasets and event percentiles. . . . .	115
9.1.	Example of a daily time series of moisture divergence. . . . .	122
9.2.	Composites of mean daily moisture divergence for different ENSO phases. . . . .	126
9.3.	Composites of 90th percentiles of moisture divergence for different ENSO phases. . . . .	127
9.4.	Composites of 10th percentiles of moisture divergence for different ENSO phases. . . . .	128
9.5.	Spearman's rank correlation coefficient between the ENSO index MEI and mean, 90th percentiles, and 10th percentiles of moisture divergence. . . . .	130
9.6.	$L^1$ -distance matrix $\mathbf{L}$ between the ranks of the local clustering coefficients (RLC) for events above the 90th percentile. . . . .	131
9.7.	$L^1$ -distance matrix $\mathbf{L}$ between the ranks of the local clustering coefficients (RLC) for events above the 90th percentile. . . . .	132

9.8.	ENSO index MEI and temporal evolution of the clustering coefficients for events above the 90th percentile. . . . .	132
9.9.	ENSO index MEI and temporal evolution of the clustering coefficients for events below the 10th percentile. . . . .	133
9.10.	Composites of the local clustering coefficient based on strong evapotranspiration events for different ENSO phases. . . . .	133
9.11.	Composites of the average size of connected components of simultaneous extremes based on strong evapotranspiration events for different ENSO phases. . . . .	134
9.12.	Composites of mean daily moisture divergence for two different types of El Niño events. . . . .	138
9.13.	Composites of 90th percentiles of moisture divergence for two different types of El Niño events. . . . .	139
9.14.	Composites of the local clustering coefficient for two different types of El Niño events. . . . .	140
A.1.	Rainfall Climatology for the fall season from March to May. . . . .	152
A.2.	Rainfall Climatology for the winter season from June to August. . . . .	153
A.3.	Rainfall Climatology for the spring season from September to November. . . . .	154
A.4.	Network measures for events above the 90th percentile for the fall season from March to May. . . . .	155
A.5.	Network measures for events above the 90th percentile for the winter season from June to August. . . . .	156
A.6.	Network measures for events above the 90th percentile for the spring season from September to November. . . . .	157
A.7.	Network measures for events above the 95th percentile for the fall season from March to May. . . . .	158
A.8.	Network measures for events above the 95th percentile for the winter season from June to August. . . . .	159
A.9.	Network measures for events above the 95th percentile for the winter season from June to August. . . . .	160
A.10.	Network measures for events above the 90th percentile for the spring season from September to November. . . . .	161
B.1.	Same as Figure 5.2 in chapter 5, but the isochrones (right column) computed for $\tau_{max} = 2$ days. . . . .	164
B.2.	Same as Figure 5.2 in chapter 5, but the isochrones (right column) computed for $\tau_{max} = 1$ day. . . . .	165
C.1.	Average number of 3-hourly events per burst of consecutive events of the four different types. . . . .	168
C.2.	Average percentage of total DJF rainfall contributed to the catchments C1 to C7 by each single burst of consecutive events of the four different types. . . . .	169

*List of Figures*

D.1. Linear trends for the magnitudes of rainfall extremes in the ECA. . .	171
D.2. Linear trends for the frequency of rainfall extremes in the ECA. . . .	172
D.3. Spatially resolved linear trends for the magnitudes of rainfall extremes in South America. . . . .	173
D.4. Network measure strength into and out of ECA. . . . .	174
D.5. Average spatial extent of area receiving extreme events during predic- tion times and subsequent two days. . . . .	175
D.6. Relationship between elevation and relative fraction of extreme events during prediction times in the ECA. . . . .	176
D.7. Same as Figure D.6, but resolved by ENSO phase. . . . .	177
D.8. Relative fraction of extreme events during prediction times and subse- quent two days. . . . .	178
D.9. Relative fraction of total DJF rainfall during prediction times and subsequent two days. . . . .	179
D.10. Same as Figure D.8 but resolved by ENSO phase. . . . .	180
D.11. Same as Figure D.9 but resolved by ENSO phase. . . . .	180
D.12. Heidke-Skill-Score as a function of the threshold for the definition of SESA times and the number of extreme events in the ECA. . . . .	181
D.13. Same as Figure D.12 but resolved by ENSO phase. . . . .	182
D.14. Network divergence (left) and strength out of SESA (right) for the 95th to 99th percentiles as rainfall event thresholds. . . . .	183
D.15. Comparison of the sum of strength into and out of SESA for five different datasets. . . . .	184
D.16. Composites of rainfall and geopotential height and wind fields at 850mbar showing the propagation from SESA to ECA from 12 hours before to 30 hours after rainfall peaks at SESA. . . . .	185

# List of Tables

7.1. Different conditions used to determine the climatic mechanism and to formulate the forecast rule. . . . .	82
7.2. Contingency table used for computing the Heidke-Skill-Score. . . . .	84
7.3. Specific values used to compute the Heidke-Skill-Score. . . . .	84



# List of frequently used abbreviations

## Climatological and geographical abbreviations

ECA	eastern Central Andes
ITCZ	Intertropical Convergence Zone
MCS	mesoscale convective systems
SACZ	South Atlantic Convergence Zone
SALLJ	South American Low-Level Jet
SAMS	South American Monsoon System
SEBRA	southeastern Brazil
SESA	southeastern South America

## Mathematical symbols

<b>S</b>	a general similarity matrix
<b>A</b>	a general network adjacency matrix
$ES^{sym}$	symmetric version of the similarity measure Event Synchronization
$ES^{dir}$	directed version of the similarity measure Event Synchronization
DG	network measure Degree
DR	network measure Directionality
BC	network measure Betweenness Centrality
MD	network measure Mean Geographical Distance
CC	network measure Clustering Coefficient
LD	network measure Long-ranged Directedness
RC	network measure Regional Connectivity
$S^{in}, S^{out}$	directed network measures in-strength and out-strength
$\Delta S$	network divergence



# Chapter 1.

## Introduction

### 1.1. Motivation

The aim of this thesis is to deepen the understanding of the collective dynamics of extreme rainfall events at different locations by analyzing observational time series in a suitable way. For this purpose, a general methodological framework to analyze the joint characteristics of synchronization of extreme events in different time series will be developed on the basis of complex network theory. We will modify a non-linear synchronization measure such that it meets our requirements and introduce new ways to construct networks encoding the synchronization structure of extreme events. Several new network measures will be introduced, which are designed to quantify specific aspects of the network topology that are relevant for understanding synchronization phenomena of extreme events.

The framework will be applied to spatially sampled rainfall data in South America, and we will show how it can be used to derive spatial patterns encoding the synchronization structure of extreme events at different locations on the surface of the earth. These spatial patterns will be used to draw inferences about the climatic mechanisms governing the occurrence and synchronization of extreme events. As we will show below, existing techniques are not suitable for this purpose due to problems concerning the distributions of the time series and varying delays between events. We will focus on three domains of application: climatic analysis of extreme rainfall, prediction of extreme rainfall, and evaluation of climate models with respect to their dynamical implementation of extreme rainfall.

The study of spatial patterns of co-variability of time series measured at different locations constitutes an integral part of geoscientific research. In most situations, such patterns are derived from observational data for exploratory purposes, providing the basis for the subsequent identification of the geophysical mechanisms underlying the interdependencies responsible for the patterns. In principle, this kind of analysis can be divided into two subsequent steps: First, a suitable measure that quantifies the co-variability of time series at different locations has to be determined. This choice strongly depends on the data distribution and on the kind of dependency that is assumed between the individual time series. Second, spatial patterns are to be derived from these co-variabilities: Given a set of time series measured at  $N$  different locations, one can compute the measure of co-variability for all possible pairs of time series, resulting in a *similarity matrix*  $\mathbf{S}$  of dimension  $N \times N$ . However, in

many applications, such a matrix is too complex in the sense that it provides more information than can be directly conceived, calling for a dimension reduction to the order of  $N$ . The resulting vector will contain a value for each of the  $N$  time series, and exhibit the spatial patterns of interest. Of, course, the interpretation of these patterns will depend greatly on the choice of the similarity measure as well as on the specific method used for dimension reduction.

Probably the simplest example for this kind of analysis are so-called correlation maps, which show the correlation<sup>1</sup> values of one given reference time series to time series measured at different locations (see Figure 1.3 for an example showing the influence of ENSO on rainfall in South America). This approach corresponds to the rather trivial dimension reduction by simply choosing one row or column from  $\mathbf{S}$ . While it is certainly useful in many cases, in many other situations, one might not be interested in the co-variability of just one reference time series with a set of time series at hand, but rather in the spatial characteristics of the joint co-variability of this set of time series. A common tool for the latter purpose are empirical orthogonal functions (EOFs), which have been applied for decades by meteorologist and climatologists to study the spatial characteristics of co-variability of climatic observables. Such EOFs are derived from principal component analysis (PCA) of the covariance matrix computed for a given set of time series, which takes the role of the similarity matrix  $\mathbf{S}$ . The dimensionality reduction is achieved by spectral decomposition of  $\mathbf{S}^T \mathbf{S}$ , and the key assumption of this approach is that the relevant information of  $\mathbf{S}$  is stored in its eigenvectors (i.e., the EOFs, see Figure 1.2 for an example showing the second EOF of daily rainfall of the South American monsoon). However, all approaches based on PCA of the covariance matrix carry some technical caveats that have to be considered. First, if the data are not normally distributed, the resulting EOFs will by construction only be orthogonal, i.e. uncorrelated, but in general not independent (Dommenget and Latif, 2002; Monahan et al., 2009). Thus, the spatial patterns exhibited by EOFs of different order do in general not correspond to independent processes. This has to be taken into account when interpreting the spatial patterns in a climatic context, since one dynamical mechanism may be responsible for the variability patterns of several EOFs. Second, only linear dependencies between the time series can be discovered on the basis of the covariance matrix. This should be considered a too strong restriction in most contexts, for most equations governing the dynamics underlying geophysical data are highly non-linear. Third, as the approach is based on the covariance matrix, it only captures information about the first two statistical moments of the data. In particular, any information about interdependencies of extreme events in the different time series is lost.

Of course, there exists a plethora of similarity measures and techniques to reduce the dimensionality of the matrix  $\mathbf{S}$ . However, for the purpose of this thesis, namely the analysis of the spatial synchronization structure of extreme rainfall events, to our knowledge no suitable methodology exists in the literature. We intend to fill this gap

---

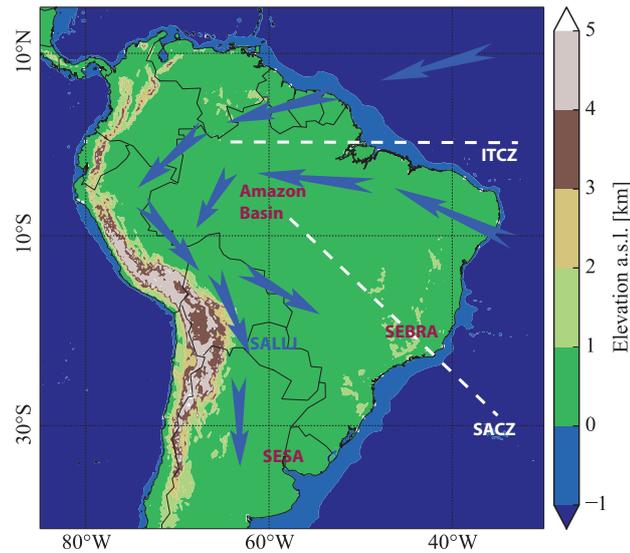
<sup>1</sup>For now, think of Pearson's correlation coefficient, or Spearman's Rho, for instance.

by representing the synchronization structure of extreme events by complex networks, and analyzing it using the language of complex network theory.

During the past decade, complex network theory has proven to provide powerful tools for analyzing the spatial characteristics of co-variability of climate time series (e.g. Tsonis and Roebber, 2004; Tsonis et al., 2007; Yamasaki et al., 2008; Donges et al., 2009b; Malik et al., 2012; Steinhäuser et al., 2012; Berezin et al., 2012; Ludescher et al., 2013), resulting in the popular *climate network* approach. We will build upon these approaches, refine and further develop them in order to obtain a methodology that is suitable for analyzing the synchronization structure of extreme events in large sets of time series. For this purpose, new ways to construct networks from the synchronization of extreme events will be introduced with special focus on the statistical significance of the networks. Furthermore, several new measures on undirected as well as directed and weighted networks will be introduced, which are specifically designed to quantify the topological aspects of the networks that are relevant for climatological applications.

In general terms, the methodology that will be developed in the following chapters is based on the idea that relevant and important features or mechanisms of a given complex system influence the way how extreme events in the time series of this system synchronize. A network is constructed by representing strong and statistically significant synchronizations by network links, and the topology of this network is assumed to encode the influences of the driving features or mechanisms. By means of suitable measures, it should then be (and is in fact, as we will show) possible to extract this information from the network's topology, and in this way infer the underlying features and mechanisms from the time series. The methodology can thus be understood as a data exploration tool that can be used to develop scientific hypotheses about the mechanisms driving a given complex system under consideration. Furthermore, in situations where information on the temporal order of events is available, we will show how directed networks can be employed to statistically track the propagation of these events through the network, which can in certain situations be used for prediction of extreme events.

Specifically for the case of rainfall extremes, we assume that the mechanisms underlying a climatic system like the South American monsoon influence the way how extreme rainfall events synchronize at different locations. Spatially embedded networks will be derived from the synchronization characteristics of extreme events at different locations in space. Along the lines of the general approach, suitable network measures will be introduced and interpreted in a climatic context in order to reveal relevant climatic features underlying the monsoon system from the internal structure of the networks. In the following chapters, we shall show how this can be used for climatic analysis of extreme rainfall, but also for statistical prediction of these events as well as for evaluating their implementation in climate models.



**Figure 1.1.:** Topography of South America and key features of the South American monsoon system, such as the Intertropical Convergence Zone (ITCZ), the South Atlantic Convergence Zone (SACZ), the South American Low-Level Jet (SALLJ), as well as the three regions SESA, SEBRA, and the Amazon Basin which are referred to in the text.

## 1.2. Climatological Setting: The South American Monsoon System

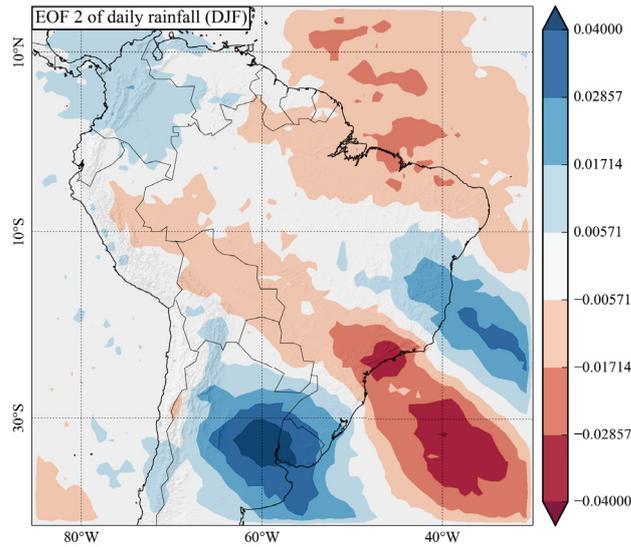
Rainfall in most parts of South America critically depends on moisture inflow from the tropical Atlantic Ocean at low atmospheric levels. This inflow is provided all year round by the trade winds, which transport large amounts of moisture along the Intertropical Convergence Zone (ITCZ) towards the Amazon Basin (Figure 1.1). During austral summer (December to February), differential heating between ocean and land, as well as precipitation-induced latent heat release over the Amazon Basin, substantially enhance the low-level flow from the tropical Atlantic Ocean towards the continent (Zhou and Lau, 1998; Rodwell and Hoskins, 2001). Furthermore, the southward displacement of the ITCZ during this time of the year leads to a complex interplay between the strengthened low-level moisture inflow, the Andean topography, and frontal systems originating from the southern tip of the continent. The climatic system emerging from this interplay is commonly referred to as South American Monsoon System (SAMS, (Vera et al., 2006; Marengo et al., 2012)).

Upon recycling large fractions of their moisture content through precipitation and evapotranspiration over the Amazon Basin (Eltahir and Bras, 1993; Marengo, 2006), the low-level winds are ultimately blocked by the Andes cordillera to the West. The shape of the mountain range causes the flow to be redirected southward along the slopes of the northern Central Andes, where they lead to high rainfall amounts due

## 1.2. Climatological Setting: The South American Monsoon System

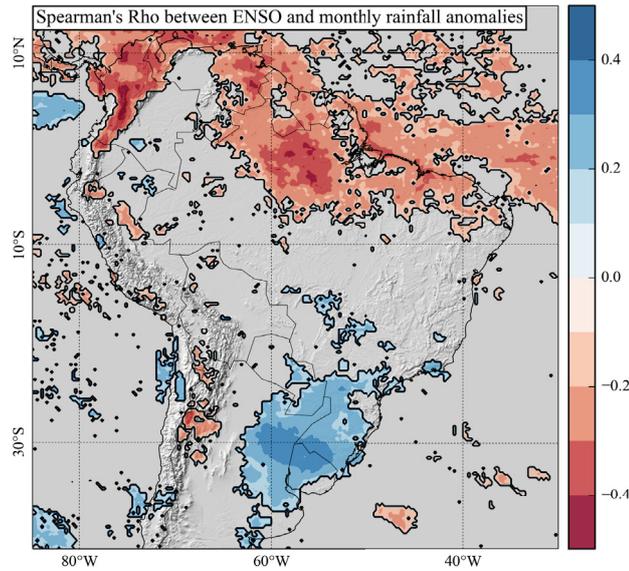
to orographic lifting (Bookhagen and Strecker, 2008).

There exists considerable variability in the direction and strength of the subsequent low-level flow to the subtropics, with typical exit regions ranging from central Argentina to eastern Brazil. Two dominant circulation regimes have been identified and extensively discussed in this context: A pronounced southward component of the flow south of approximately  $15^{\circ}\text{S}$  establishes the South American Low-Level Jet (SALLJ, (Marengo et al., 2004)) and, in case of particularly strong southward anomalies, the Chaco Jet (Salio et al., 2002; Saulo et al., 2004). These wind systems east of the slopes of the Andes are associated with enhanced rainfall in southeastern South America (SESA) (Liebmann et al., 2004), and specifically with the formation of mesoscale convective systems (MCS, (Maddox, 1980; Nicolini et al., 2002; Salio et al., 2007; Durkee and Mote, 2009)) in this region. These MCS contribute substantial fractions of total seasonal precipitation in large parts of subtropical South America, and in particular in SESA (Durkee et al., 2009). In contrast, if the flow exhibits a pronounced eastward component, it transports moist air to the South Atlantic Convergence Zone (SACZ, (Liebmann et al., 2004; Carvalho et al., 2004; Jorgetti et al., 2013)), a convective band of precipitation extending from the central Amazon to southeastern Brazil (SEBRA).



**Figure 1.2.:** Second empirical orthogonal function (EOF) computed from the covariance matrix of daily rainfall data for the core monsoon season from December to February (DJF) for the time period from 1998 to 2012. Rainfall data were obtained from the gauge-calibrated satellite product TRMM 3B42 V7 (Huffman et al., 2007).

The alternation between these two regimes, with enhanced rainfall in SESA when rainfall in SEBRA is suppressed and vice versa, is commonly called the South American rainfall dipole, and constitutes the dominant pattern of intra-seasonal variability of the SAMS (Nogués-Paegle and Mo, 1997; Carvalho et al., 2002; Liebmann et al., 2004;



**Figure 1.3.:** Spearman's Rho between the El Niño Southern Oscillation Index MEI (Wolter and Timlin, 1993) and monthly anomalies of rainfall over South America for the time period from 1998 to 2011. Rainfall data were obtained from the gauge-calibrated satellite product TRMM 3B42 V7 (Huffman et al., 2007).

Vera et al., 2006). The inverse co-variability of rainfall in these two regions can be easily observed using EOFs of daily rainfall during the monsoon season (see Figure 1.2). The oscillation between these regimes is related to frontal systems approaching from southern Argentina, caused by Rossby waves in the polar jet streams. The associated low-pressure systems interact with the tropical low-level circulation and force the flow of warm and moist tropical air masses along their isobars to the subtropics (Liebmann et al., 1999; Siqueira and Machado, 2004). The onset of the monsoon is related to these frontal systems becoming stationary over SEBRA, and thereby establishing the SACZ (Nieto-Ferreira et al., 2011).

The El Niño Southern Oscillation (ENSO) strongly influences the climate of South America (Vera et al., 2006; Marengo et al., 2012). In particular, rainfall in large parts of South America shows strong dependencies on ENSO variability (Barros et al., 2008; Grimm and Tedeschi, 2009; Bookhagen and Strecker, 2010; Tedeschi et al., 2013). For example, monthly rainfall anomalies in eastern tropical South America are negatively correlated with the multivariate ENSO index MEI (Wolter and Timlin, 1993), while monthly rainfall anomalies in SESA are positively correlated (Figure 1.3). Therefore, during warm ENSO conditions (i.e., El Niño phases), rainfall in eastern tropical South America can be expected to be suppressed, while rainfall in SESA is typically enhanced. The latter fact is associated with a strengthening of the SALLJ during warm ENSO episodes (Silva et al., 2009).

### 1.3. Arrangement of this thesis

According to the general rationale outlined above, this thesis is organized as follows. In chapter 2, possible measures of similarity to quantify the degree of co-variability are discussed, and it is argued why Event Synchronization provides the most suitable starting point to analyze the synchronization of extreme events. Several modifications will be described in this chapter, and the question of how to determine statistical significance will be addressed in this chapter.

In chapter 3, we introduce the relevant elements of complex network theory. It will be described how networks are constructed from the values provided by the similarity measure, and existing as well as new network measures will be introduced. Furthermore, caveats with spatially embedded networks will be discussed, and a method to overcome these problems will be introduced.

A first application of the methodology to daily rainfall data is presented in chapter 4. It will be shown how suitable measures on undirected and unweighted networks reveal the key features of the SAMS. This chapter mainly serves as a proof of concept for the methodology, as the obtained findings are consistent with existing results on this well-known climate system.

Due to a missing theoretical framework, the aforementioned rainfall dipole between SESA and SEBRA could so far only be analyzed with respect to mean conditions, but not specifically with respect to its role concerning the dynamics of extreme events. In chapter 5, by constructing networks separately for the two phases, we will show how the synchronization pathways of extreme events differ among the two regimes. Furthermore, evidence is provided in this chapter that the rainfall dipole is only the most pronounced mode of an oscillation that extends over the entire South American continent.

In chapter 6, we shall derive directed networks of extreme event synchronization in order to reveal the geographical origins of spatially extensive, long-lasting rainfall clusters over the main catchments along the eastern slopes of the Andes. By revealing such teleconnections of extreme rainfall events, insights into their functioning and the responsible climatic processes are obtained, but also the potential predictability of these events is assessed.

Using directed and weighted networks, we shall proceed to establish a general framework for the prediction of extreme events in complex systems in chapter 7. Applying this methodology to three-hourly rainfall data reveals that, under specific atmospheric conditions, extreme rainfall events propagate from SESA toward the Central Andes, where they are likely to lead to severe floods and landslides. We will show how a simple forecast rule for these events can be formulated on the basis of the general theoretical framework.

In chapter 8, several methodological concepts developed in the previous chapters will be collected to perform an extensive evaluation of observational data and climate models with respect to the representation of the static and dynamic properties of extreme rainfall events. We will show that climate models, while performing

## *Chapter 1. Introduction*

reasonably well in reproducing average rainfall values, have substantial difficulties with the dynamics of extreme events.

In chapter 9, we shall extend the methods developed in the previous chapters to a dynamical analysis using a sliding window approach. We will combine the methods derived to analyze the spatial synchronization structure of extreme events with elements from pattern recognition theory and recurrence networks in order to analyze impacts of ENSO on the clustering properties of strong rainfall and evapotranspiration events over South America.

Part I.

# Theoretical Framework



# Chapter 2.

## Measures of Similarity

### 2.1. Introduction

The key idea of the climate network approach that will be presented in the next chapter is to represent interdependencies between time series of climatic observables at different locations by network links. Many different measures have been employed to quantify these interdependencies, and in general terms we will refer to such measures as *similarity measures*. In the following, we will show that none of the standard measures fulfills all the requirements we desire to analyze the synchronization of extreme events, and how the measure that satisfies most of the requirements has to be modified to become perfectly suited for our purposes.

The probably most widely-used similarity measure is Pearson's correlation coefficient (PCC). For two time series  $x$  and  $y$  of length  $T$  with existing means  $\bar{x}, \bar{y}$  and standard deviations  $\sigma_x, \sigma_y$ , it is defined as the bilinear form

$$PCC(x, y) := \frac{Cov(x, y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^T (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^T (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^T (y_i - \bar{y})^2}} \quad (2.1)$$

and thus  $PCC(x, y) \in [-1, +1] \forall x, y$ . This measure is well-suited to quantify *linear* dependencies between  $x$  and  $y$ : Geometrically, PCC is the normalized scalar product between the two vectors  $x$  and  $y$ , and  $PCC(x, y) = 1$  or  $PCC(x, y) = -1$  thus implies that the two vectors can be written as linear combinations of each other, while  $PCC(x, y) = 0$  implies that they are perpendicular. However, this does not exclude possible non-linear dependencies between  $x$  and  $y$ . Note that, even if  $x$  and  $y$  are binary event time series, where an entry  $x_i = 1$  indicates an event at the corresponding time, and an entry  $x_i = 0$  indicates no event, PCC can be applied to count the normalized number of simultaneous events in  $x$  and  $y$ . The fact that PCC is restricted to linear interrelations between  $x$  and  $y$  has motivated the introduction of many more general similarity measures. For example, the non-parametric similarity measures given by Spearman's and Kendall's rank correlation coefficients quantify general monotonic dependencies between  $x$  and  $y$ , including non-linearities. Another example of a non-linear similarity measure is Mutual Information, which quantifies the "joint information" contained in  $x$  and  $y$ . Since many processes underlying the variability and interdependency of climatic observables are known to be highly non-linear, a suitable measure of similarity between rainfall event time series should be capable

of accounting for non-linearities. However, another subtlety has to be addressed in this context: There will typically be a temporal delay between events at different locations, and this delay cannot be expected to be constant in time. For example, think of a situation where rainfall clusters frequently propagate from the location where time series  $x$  is measured to the location where time series  $y$  is measured. The propagation of rainfall may be driven by low-level winds (e.g., at 850mb), which have varying speeds. The delay between rainfall events in  $x$  and associated events in  $y$  thus depends on these varying wind speeds and will vary in time itself. Delays between  $x$  and  $y$  can be accounted for by shifting the time series  $x$  and  $y$  against each other by prescribed time windows (the lead or lag), and then computing one of the above similarity measures. However, in such a “lead-lag analysis”, only one single lead (or lag) is provided for the pair  $(x, y)$ , and assumed to be valid for the entire time range.

Therefore, we have to look for a non-linear similarity measure that is suitable for binary data, provides a unique association between events, and allows for a *dynamical* delay, i.e. varying time intervals between events of one time series and events of the other time series. To our knowledge, Event Synchronization (ES), first introduced in (Quian Quiroga et al., 2002), is the only measure which fulfills all these desiderata. In the following section, we will first introduce the original version of this measure and then introduce several modifications necessary to meet specific requirements concerning the temporal order of events, as well as their distribution in the time series.

## 2.2. Event Synchronization

The original version of ES is defined as follows (Quian Quiroga et al., 2002): Consider a set of  $N$  time series  $\{x_1, \dots, x_N\}$ , each of length  $T$ , and let  $(x_i, x_j)$  denote a pair of these time series. We define corresponding *event series*  $e_i$  and  $e_j$  as sequences containing the time indices for which the values of  $x_i$  and  $x_j$  fulfill certain prescribed conditions<sup>2</sup>. These sequences shall be ordered from lower to higher values. Consider two events  $e_i^\mu$  and  $e_j^\nu$ , with  $0 \leq \mu \leq l_i$  and  $0 \leq \nu \leq l_j$ , where  $l_i$  ( $l_j$ ) denotes the total number of events in time series  $x_i$  ( $x_j$ ). In order to decide if the two events  $e_i^\mu$  and  $e_j^\nu$  can be uniquely assigned to each other, compute for  $d_{ij}^{\mu,\nu} := e_i^\mu - e_j^\nu$  the *dynamical delay*

$$\tau = \min \left( \frac{\{d_{ii}^{\mu,\mu-1}, d_{ii}^{\mu,\mu+1}, d_{jj}^{\nu,\nu-1}, d_{jj}^{\nu,\nu+1}\}}{2} \right). \quad (2.2)$$

---

<sup>2</sup>There are many possible choices of a specific event definition. For example, events can be points in time at which the value of the time series is above a global threshold, points in time for which this value is above a given percentile of the distribution of the values of the time series, or points in time for which the values of the time series drastically change. Since such a choice depends on the specific research question and application, we will keep the definition of events in this chapter as general as possible.

In addition, a filter can be introduced by declaring a maximum delay  $\tau_{max}$  between  $e_i^\mu$  and  $e_j^\nu$ , which allows to analyze processes on different time scales. Put

$$C_{ij}^{\mu\nu} = \begin{cases} 1 & \text{if } 0 < d_{ij}^{\mu,\nu} \leq \tau \quad \text{and} \quad d_{ij}^{\mu,\nu} \leq \tau_{max} , \\ \frac{1}{2} & \text{if } d_{ij}^{\mu,\nu} = 0 , \\ 0 & \text{otherwise ,} \end{cases} \quad (2.3)$$

and compute the normalized sum of uniquely associated events. The latter summation can either be done in a symmetric fashion,

$$ES_{ij}^{sym} := \frac{\sum_{\mu=1}^{l_i} \sum_{\nu=1}^{l_j} (C_{ij}^{\mu\nu} + C_{ji}^{\mu\nu})}{\sqrt{l_i l_j}} , \quad (2.4)$$

or in an anti-symmetric fashion

$$ES_{ij}^{anti-sym} := \frac{\sum_{\mu=1}^{l_i} \sum_{\nu=1}^{l_j} (C_{ij}^{\mu\nu} - C_{ji}^{\mu\nu})}{\sqrt{l_i l_j}} . \quad (2.5)$$

While  $ES_{ij}^{sym} \in [0, 1]$  gives the normalized total number of *synchronous* (i.e., uniquely associable) events regardless of their temporal order,  $ES_{ij}^{anti-sym} \in [-1, 1]$  yields the normalized difference between synchronous events which occurred at  $j$  before they occurred at  $i$  and events which occurred at  $i$  before they occurred at  $j$ .

For some applications, however, this specific definition of  $ES_{ij}^{anti-sym}$  is not appropriate. Given two time series  $x_i$  and  $x_j$ , measured at locations  $i$  and  $j$ , one may be interested in the total number of synchronous events which first occurred at  $j$  and thereafter at  $i$ , and, separately, in the total number of synchronous events which first occurred at  $i$  and thereafter at  $j$ . For this reason, a modified version of *directed Event Synchronization* will be introduced in the following, where the corresponding sums will be stored separately instead of computing their difference. Furthermore, in particular when applying ES to data with high temporal resolution, there may occur situations where, typically, several events occur during consecutive time steps. Such sequences of events in a row will be referred to as *bursts*. In such situations, only the first will be considered as an event, weighted by the number of subsequent events, which are themselves discarded from the summation. Thus, for each event  $e_i^\mu$ , there is a weight  $w_i^\mu$ . We then put

$$W_{ij}^{\mu\nu} = \begin{cases} \min(w_i^\mu, w_j^\nu) & \text{if } 0 < d_{ij}^{\mu,\nu} \leq \tau \quad \text{and} \quad d_{ij}^{\mu,\nu} \leq \tau_{max} , \\ 0 & \text{otherwise ,} \end{cases} \quad (2.6)$$

where it should be emphasized that events at the very same time do not contribute, since they do not allow to determine the temporal order. The introduction of weights  $w_i^\mu$  above assures that in situations where a burst in a given event series  $e_j$  is followed by a burst in another event series  $e_i$ , but such that there is temporal overlap between the two bursts, all events are still counted in a time-ordered manner. We define

directed Event Synchronization (see also P4 and P5) by

$$ES_{ij}^{dir} := \frac{\sum_{\mu=1}^{l_i} \sum_{\nu=1}^{l_j} W_{ij}^{\mu\nu}}{\sqrt{l_i l_j}}, \quad (2.7)$$

which is neither necessarily symmetric nor anti-symmetric: in general  $ES_{ij}^{dir} \neq ES_{ji}^{dir}$  and  $ES_{ij}^{dir} \neq -ES_{ji}^{dir}$ .

We emphasize again that, by virtue of equation (2.2), none of these versions of ES assumes temporal homogeneity between the time series because the possible delay between events is dynamical (within the range of  $\tau_{max}$ ), contrary to the static delay in classical lead-lag correlation analysis.

### 2.2.1. The influence of the event rate

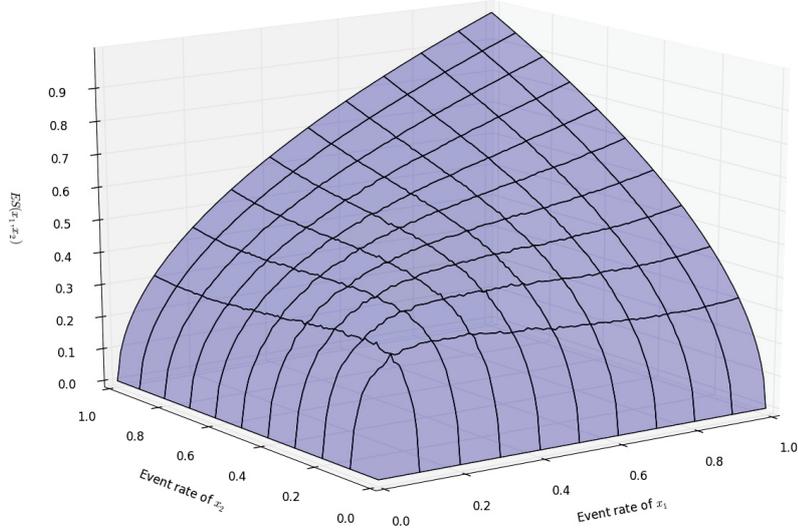
We define the *event rate* of a time series  $x_i$  as the quotient of the number of events  $l_i$  and the length  $T$  of  $x_i$ :  $r_i = l_i/T$ . Despite the normalization factor  $\sqrt{l_i l_j}^{-1}$  in equations (2.4) and (2.5), the value of ES depends on event rates  $r_i$  and  $r_j$  if  $\tau_{max}$  is finite (which will be the case for all practical applications), since the probability of “random” synchronizations increases with increasing event rates (Figure 2.1). The values of the matrix  $\mathbf{ES} \in [0, 1]^{N \times N}$  computed for different pairs of event series are thus not directly comparable if the event rate varies between the event series.

The most obvious way to nevertheless obtain comparable values of ES is to define events in such a way that the event rate is equal for all event series under consideration. However, some common event definitions do by construction not allow equal event rates in all event series (e.g., if events are defined as time steps for which the corresponding values are above a global threshold), thus demanding a more sophisticated solution to this problem. A suitable approach in such situations is to compare the values of ES on the level of their statistical significance, which itself depends on the event rates  $r_i$  and  $r_j$  of the considered event series  $e_i$  and  $e_j$ . Appropriate statistical null models for ES depend on the specific event definition and will be described in detail in the corresponding following chapters. For now, we shall assume that the probability distribution consistent with the null hypothesis is already obtained. Denoting the corresponding probability density function by  $H_{r_i, r_j}$ , statistical significance of a given empirical value  $\overline{ES}_{ij}$  can be estimated by the probability of obtaining a value larger or equal to  $\overline{ES}_{ij}$  given the null hypothesis:

$$P(ES_{ij} \geq \overline{ES}_{ij}) = \int_{\overline{ES}_{ij}}^1 H_{r_i, r_j}(s) ds \quad (2.8)$$

In contrast to the values  $\overline{ES}_{ij}$  themselves, their *positions* in the respective null model distribution (i.e.  $P(ES_{ij} \geq \overline{ES}_{ij})$ ) are comparable among pairs of event series with different event rates  $(r_i, r_j)$ . Note that in this approach, one may omit the normalization by  $\sqrt{l_i l_j}$  in equations (2.4) and (2.5) altogether.

### 2.3. Comparison between Pearson's correlation coefficient and Event Synchronization



**Figure 2.1.:** The influence of the event rates ( $r_1$  and  $r_2$ ) on the values of ES for finite  $\tau_{max}$ . For each specific combination  $(r_i, r_j)$  (corresponding to a single point in the  $x_1 - x_2$ -plane) we construct 1000 surrogate time series pairs  $(\bar{x}_i, \bar{x}_j)$  by uniformly randomly placing events such that the resulting event rate is  $r_i$  respectively  $r_j$ , and compute  $ES^{sym}$  for all these pairs. The values shown in the figure are the average values of  $ES^{sym}$  over the 1000 values obtained for each set of surrogate pairs.

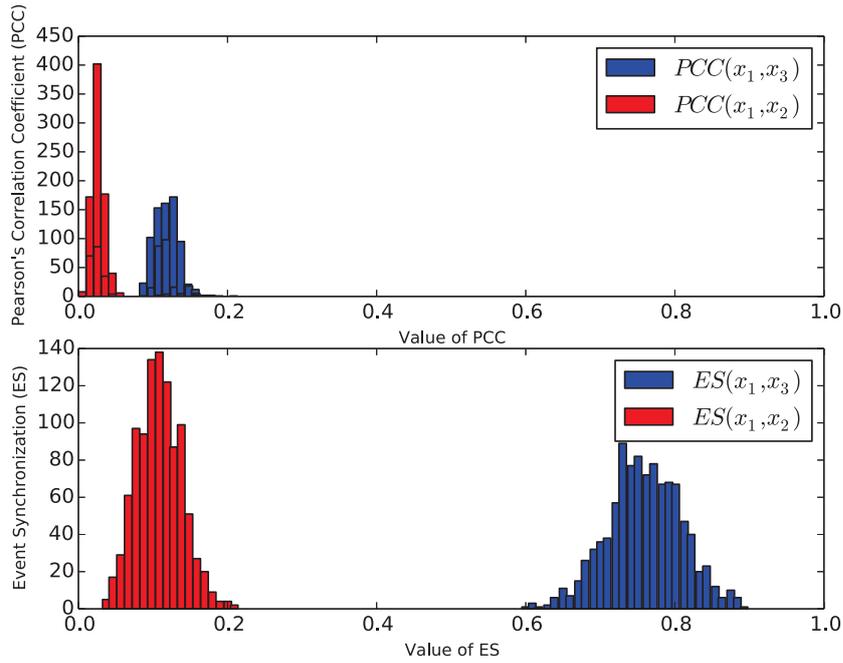
### 2.3. Comparison between Pearson's correlation coefficient and Event Synchronization

Lead-lag analysis on the basis of Pearson's correlation coefficient (PCC) is the most straightforward and most-often applied approach to derive directed network links encoding the temporal order of interrelations between time series at different locations. However, while certainly possible, lead-lag analysis using PCC is not the best-suited approach for our purpose for the following reasons:

1. In the following chapters, we will investigate binary time series of extreme rainfall events, for which only few of the entries will be 1, while the remaining entries will be 0. We cannot expect that the delay between events in two different time series will be constant in time. However, this is a necessary condition for a lead-lag analysis, since it only provides one single lead (or lag), namely the length of the time window by which one time series is shifted against the other.

2. The lead-lag analysis may not provide unique correlations, since there are often several maxima of PCC over the range of leads and lags.

To exemplify the difference between the PCC lead-lag analysis and our approach, we construct the following time series: Let  $x_1$  and  $x_2$  denote two event time series of length  $t = 10000$ , with 100 independently and uniformly randomly chosen entries



**Figure 2.2.:** Comparison of Pearson's correlation coefficient (PCC) and Event Synchronization (ES) for binary data with varying delay between events. The event time series  $x_1$  and  $x_2$  are independent, while  $x_1$  and  $x_3$  are dependent in the sense that events in  $x_3$  follow events in  $x_1$  within 16 time steps.

equal to 1, and the remaining entries equal to 0 (the event rate corresponds to an event threshold at the 99th percentile of all entries, which we will for example employ in chapter 7).

Let furthermore  $x_3$  denote the time series obtained from  $x_1$  by shifting each entry 1 to the right by an integer uniformly randomly chosen from the set  $\{1, \dots, 16\}$ . Thus,  $x_3$  is a time series for which all events follow events at  $x_1$  within a time window of 16 time steps, corresponding to  $\tau_{max} = 16$ . This maximum delay corresponds to the choice in chapter 7. We can now compute  $ES(x_1, x_2)$ ,  $ES(x_1, x_3)$ , as well as  $\max_{i \in \{1, \dots, 16\}} PCC(x_1(0, \dots, t-i), x_2(i, \dots, t))$ , and  $\max_{i \in \{1, \dots, 16\}} PCC(x_1(0, \dots, t-i), x_3(i, \dots, t))$ . Since  $x_1$  and  $x_2$  are independent, but  $x_1$  and  $x_3$  are strongly dependent, we expect low correlation values between  $x_1$  and  $x_2$ , but high correlation values between  $x_1$  and  $x_3$ . Repeating this procedure 1000 times, we obtain histograms of the corresponding values of PCC and ES (Figure 2.2). It can be observed that the values for  $ES(x_1, x_3)$  (around 0.75) are substantially higher than  $ES(x_1, x_2)$ . In contrast, this is not the case for PCC, for which values of  $PCC(x_1, x_3)$  are distributed around 0.1. This example explains why, in contrast to the ES approach we chose, employing a lead-lag analysis using PCC might produce misleading results, namely small correlation values for strongly dependent time series.

# Chapter 3.

## Network Theory

### 3.1. Introduction

In this chapter, we will present the elements of network theory which are relevant for the applications presented in the following chapters. For a more complete and detailed introduction to network theory, we refer to (Newman, 2010). The methodological framework outlined in this chapter also provides the theoretical foundations of the associated publications P1 to P8, and some of the following paragraphs closely follow the corresponding sections in the respective publications. We will introduce the standard notions and terminologies of network theory, including some standard network measures, but also introduce several new network measures, which are defined to quantify specific properties of the network topology that are relevant in the context of extreme event synchronization.

### 3.2. Mathematical representation of networks

In mathematical graph theory, a *graph*  $\mathcal{G}$  is defined as a set of *vertices*  $\mathcal{V}$  together with a set of *edges*  $\mathcal{E}$  connecting the vertices (Newman, 2010):

$$\mathcal{G} := (\mathcal{V}, \mathcal{E})$$

Especially in physics, graphs are usually referred to as *networks*, vertices are often referred to as *nodes*, and the edges are called *links*. Throughout this thesis, the respective terms will be used as synonyms. A graph for which there are no edges from a vertex to itself (so-called *self-loops*) and at most one edge between any two vertices is called *simple graph*. A common representation of a simple graph is in terms of the adjacency matrix:

$$A_{ij} = \begin{cases} 1 & \text{if the vertices } i \text{ and } j \text{ are connected by an edge,} \\ 0 & \text{otherwise.} \end{cases} \quad (3.1)$$

An adjacency matrix defined in this way is symmetric, since the condition in (3.1) is symmetric under exchange of  $i$  and  $j$ . In this sense, the corresponding network is *undirected*. In many applications, however, it is beneficial to drop this property and

allow for directionality of the links, resulting in a *directed* network, represented by a non-symmetric adjacency matrix:

$$A_{ij} = \begin{cases} 1 & \text{if there exists an edge from vertex } j \text{ to vertex } i, \\ 0 & \text{otherwise.} \end{cases} \quad (3.2)$$

Furthermore, it is possible to assign weights  $w_{ij}$  to the links, which leads to a *weighted* (and possibly still directed) network. This can be represented by an adjacency matrix with entries equal to the weights:  $A_{ij} = w_{ij}$ .

### 3.3. Construction of climate networks from empirical data

Consider a set of  $N$  empirical time series  $\{x_i\}_{1 \leq i \leq N}$  of some climatological variable (e.g. temperature, pressure, or rainfall). Let further  $\mathbf{S} = \{S_{ij}\}_{1 \leq i, j \leq N}$  denote a general, possibly non-symmetric similarity matrix for this set of time series, with entries  $S_{ij}$  equal to the values of the corresponding similarity measure  $S$  computed for the pairs  $(x_i, x_j)$ . Many different similarity measures have been employed for the construction of networks from climate data (in the following called *climate networks*). Many approaches are based on the covariance matrix of the set of time series, which includes using Pearson's correlation coefficient (Tsonis et al., 2006; Tsonis and Swanson, 2008; Gozolchiani et al., 2008; Yamasaki et al., 2008; Donges et al., 2009b; Gozolchiani et al., 2011)). However, the non-linearities immanent in the climate system have motivated to also use similarity measures capturing non-linear dependencies, such as Mutual Information (Donges et al., 2009a; Hlinka et al., 2013; Hlinka et al., 2014). The right choice of a similarity measure ultimately depends on the specific climatological variable under consideration, its distribution, and on the dynamical processes underlying the dependencies of time series at different locations. As described in the last chapter, the best-suited measure for analyzing the synchronicity of extreme rainfall is Event Synchronization, which is employed in all of the applications presented in the following chapters. But since the network theoretic concepts described in the following do not depend on the specific similarity measure, we will present them for a general similarity matrix  $\mathbf{S}$ .

In general terms, a weighted and directed network can be derived from the matrix  $\mathbf{S}$  by identifying the time series with the network's nodes and taking the similarity matrix as the network's adjacency matrix:  $A_{ij} = S_{ij}$ . If all entries  $S_{ij}$  are non-zero, this yields a *fully connected* network. A not fully connected network can be obtained by specifying certain conditions on the entries  $S_{ij}$  under which a link will be placed. Formally denoting the set of possible values fulfilling these prescribed conditions on  $S_{ij}$  by  $\mathcal{C}_{ij}$ , this is captured by the adjacency matrix

$$A_{ij} = \begin{cases} S_{ij} & \text{if } S_{ij} \in \mathcal{C}_{ij} \text{ and } i \neq j, \\ 0 & \text{otherwise,} \end{cases} \quad (3.3)$$

### 3.3. Construction of climate networks from empirical data

where self-loops have been excluded. A corresponding unweighted network may be derived by setting  $A_{ij} = 1$  instead of  $A_{ij} = S_{ij}$  in equation (3.3). A common specification of the conditions  $\mathcal{C}$  is to set network links only if the corresponding entries of  $\mathbf{S}$  are larger than a prescribed value. For example in (Tsonis et al., 2006; Tsonis and Swanson, 2008), the entries of  $\mathbf{S}$  are Pearson's correlation coefficients between the respective time series, and links are placed between nodes if the corresponding absolute values of the entries  $S_{ij}$  are above 0.5. A different approach was taken for example in (Donges et al., 2009a; Donges et al., 2009b), where a link is placed between two nodes if the corresponding entry of  $\mathbf{S}$  (in these studies Pearson's correlation coefficient as well as Mutual Information) is above some  $p$ -th percentile threshold, resulting in a network with prescribed link density  $\rho = (100 - p)\%$ .

The prescribed conditions  $\mathcal{C}$  should in general assure that only statistically significant values  $S_{ij}$  are represented by network links (Paluš et al., 2011). In the two approaches described in the last paragraph, the threshold or link density can be adjusted in such a way. However, in the applications presented in chapters 5 and 6, the event rates in each time series are not fixed, and thus a different way to estimate statistical significance has to be found, which accounts for the specific distributions of events in the time series. In such situations, the significance threshold corresponding to a given confidence (e.g. a p-value of 0.05) depends on the event rates of the pair of time series  $(i, j)$ . Denoting the corresponding significance threshold by  $T_{ij}$ , equation (3.3) becomes

$$A_{ij} = \begin{cases} S_{ij} & \text{if } S_{ij} \in \mathcal{C}_{ij} \text{ and } i \neq j, \\ 0 & \text{otherwise,} \end{cases} \quad (3.4)$$

which for the unweighted case translates to

$$A_{ij} = \Theta(S_{ij} - T_{ij}) - \delta_{ij}, \quad (3.5)$$

where  $\Theta$  denotes Heaviside's function and Kronecker's  $\delta$  excludes self-loops. The specific statistical null models used to derive  $T_{ij}$  differ between the different applications and will be described in the corresponding chapters.

The climatological time series are derived from measurements at different locations, and are ideally available on homogeneous geographical grids. In cases where a given variable is only measured on the earth's surface (such as rainfall), the resulting climate network is then embedded on this two-dimensional surface (see Figure 3.1 for an example over South America). Each network node representing a time series then also corresponds to a grid cell, which can be uniquely referred to by a pair of angles  $(\lambda, \phi)$ , with latitude  $\lambda \in (-90^\circ, 90^\circ)$  and longitude  $\phi \in [0^\circ, 360^\circ)$ .



**Figure 3.1.:** Exemplary visualization of a climate network derived from rainfall data over South America.

## 3.4. Complex network measures

### 3.4.1. Undirected and unweighted networks

On undirected and unweighted climate networks, we are going to apply the well-established measures *degree*, *betweenness centrality*, and the *local clustering coefficient* (Newman, 2010). Furthermore, we will introduce the new network measures *mean geographical distance*, *long-ranged directedness*, *regional connectivity*, and *directionality*. All these measures are *local* in the sense that they assign a value (or two values for directionality) to each network node  $i \in \{1, \dots, N\}$  corresponding to a certain geographical position. Here, we will give the mathematical definitions of these network measures, while their respective climatological interpretations will be provided in context with their application in the respective following chapters.

The degree (DG) of a node  $i$  is defined as the sum of network nodes  $j$  to which  $i$  is connected to by network links:

$$DG_i := \sum_{j=1}^N A_{ij} \quad (3.6)$$

The measure betweenness centrality (BC) is based on the concept of shortest paths in the network. A shortest path in a network between two nodes  $k$  and  $l$  is the shortest

possible sequence of links that have to be passed to get from  $k$  to  $l$ . Betweenness centrality is then defined as

$$BC_i := \frac{\sum_{l < k \neq i} \sigma_{kl}(i)}{\sum_{l < k \neq i} \sigma_{kl}} \quad (3.7)$$

where  $\sigma_{kl}$  denotes the total number of shortest network paths from  $k$  to  $l$  and  $\sigma_{kl}(i)$  the number of shortest network paths from  $k$  to  $l$  which go through node  $i$ .

The clustering coefficient (CC) at a given node  $i$  is defined as the fraction of the number of triangles and the number of pairs formed by network links connected to  $i$ :

$$CC_i := \frac{\sum_{j < k} A_{ij} A_{jk} A_{ik}}{\sum_{j < k} A_{ij} A_{ik}} \quad (3.8)$$

We further define the mean geographical distance (MD, introduced in the associated publication P1) at a network node  $i$  as the arithmetic mean of the geographical distances of all network links attached to  $i$ :

$$MD_i := \frac{1}{DG_i} \sum_{j=1}^N A_{ij} dist(i, j) \quad (3.9)$$

where  $dist(i, j)$  denotes the great-circle distance (measured in  $[km]$ ) between the grid points corresponding to nodes  $i$  and  $j$ .

We combine the three network measures BC, CC, and MD in the definition of long-ranged directedness (LD, introduced in the associated publication P1). For this purpose, we calculate the normalized ranks of BC, CC, and MD, denoted by NRBC, NRCC, and NRMD, respectively, and put

$$LD_i = \frac{1}{2} NRBC_i + \frac{1}{2} NRMD_i - NRCC_i. \quad (3.10)$$

In many cases, one may be interested in the connectivity of different geographical regions in terms of their rainfall synchronicity. For this purpose, we define the measure regional connectivity (RC, introduced in the associated publication P7) For each node  $i$ , RC to a geographical region  $R$  is defined as the number of nodes (i.e., grid cells) in  $R$  which are connected to  $i$  by a network link:

$$RC_i(R) = \sum_{j \in R} A_{ij} \quad (3.11)$$

In addition, we are interested in the directions along which extreme events occur synchronously. For this purpose, we employ the measure *directionality* (DR, introduced in the associated publication P2). At each node  $i$ , this measure yields two values: the dominant angle  $DR_i^\phi$  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  $\phi_{ij}$  denote the angle between the meridian going through node  $i$  and the straight line between node  $i$  and  $j$ . For undirected networks, all angles  $\phi_{ij}$  are taken as *modulo*  $\pi$ , thus  $\phi_{ij} \in [0, \pi)$ . In the following, we will therefore refer to  $DR_i^\phi$  as an *orientation* rather than an angle. We first compute the frequency distribution of all orientations  $\varphi$  of links at  $i$ :

$$P_i(\varphi) = \sum_{j:\phi_{ij} \in (\varphi-\epsilon, \varphi+\epsilon)} A_{ij} , \quad (3.12)$$

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

$$DR_i^s = \max_{\varphi \in [0, \pi)} P_i(\varphi) \quad (3.13)$$

together with the corresponding orientation

$$DR_i^\phi = \arg \max_{\varphi \in [0, \pi)} P_i(\varphi) . \quad (3.14)$$

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 chapter 5 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 time (subject to the temporal resolution of the dataset). In this sense, they can be interpreted as *isochrones*.

### 3.4.2. Directed and weighted networks

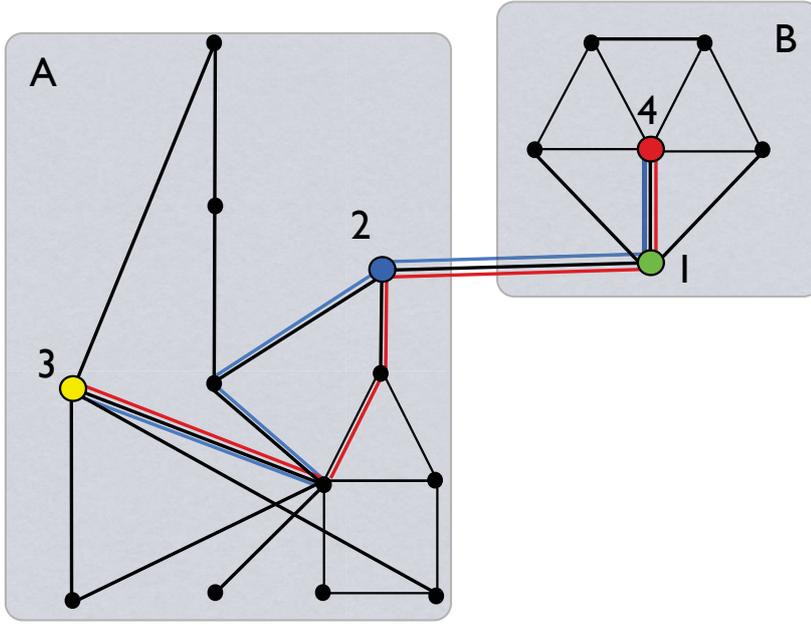
The adjacency matrix  $\mathbf{A}$  of a directed network is in general not symmetric. In this thesis, we choose the convention that a non-zero entry  $A_{ij}$  represents a link pointing from node  $j$  to node  $i$ . The generalizations of degree on directed and weighted networks are the *in-strength*  $\mathcal{S}_i^{in}$  and *out-strength*  $\mathcal{S}_i^{out}$  (Newman, 2010). At a network node  $i$ ,  $\mathcal{S}_i^{in}$  is defined as the sum of weights assigned to links pointing to node  $i$ ,

$$\mathcal{S}_i^{in} := \sum_{j=1}^N A_{ij} , \quad (3.15)$$

while  $\mathcal{S}_i^{out}$  is defined as the sum of weights assigned to links pointing from node  $i$ :

$$\mathcal{S}_i^{out} := \sum_{j=1}^N A_{ji} \quad (3.16)$$

On the basis of these measures, we define the measure *network divergence* ( $\Delta\mathcal{S}$ , introduced in the associated publication P5) as the difference of in-strength and



**Figure 3.2.:** An example network consisting of two subnetworks A and B is depicted. Nodes are called connected if there exists a link between them. Between nodes 3 and 4, there are two shortest paths, marked by the red and blue lines. Nodes 1 and 2 have high BC, as they lie on all shortest paths between nodes in A and nodes in B. Node 3 has high MD, as the mean of the spatial lengths of its links is high. Node 4 will have high CC, as many of its network neighbors are connected themselves. For node 1, CC is higher than for node 2, resulting in lower LD for node 1 than for node 2, although they have the same value of BC.

out-strength at each grid cell:

$$\Delta \mathcal{S}_i := \mathcal{S}_i^{in} - \mathcal{S}_i^{out}. \quad (3.17)$$

This measure provides us with the notion of sinks and sources of network strength, and will be of great use for estimating the predictability of extreme events.

The measure regional connectivity introduced above can also be generalized to directed and weighted networks. We define the strength into a geographical region  $R$  from a node  $i$  as

$$\mathcal{S}_i^{in}(R) = \frac{1}{|R|} \sum_{j \in R} A_{ji}, \quad (3.18)$$

and the strength out of a geographical region  $R$  to a node  $i$  as

$$\mathcal{S}_i^{out}(R) = \frac{1}{|R|} \sum_{j \in R} A_{ij}, \quad (3.19)$$

where  $|R|$  denotes the number of nodes contained in  $R$ . Note that this way, we defined the strength out of a region  $R$  to a node  $i$  as the average in-strength of  $i$ , restricted to the sources in  $R$  and vice versa. Thus, e.g.,  $\mathcal{S}_i^{out}(R) = 1$  would imply perfect connectivity from each node in  $R$  to  $i$ : there would be a link from each node in  $R$  to  $i$  and each of these links would have weight equal to 1.

### 3.5. The influence of the spatial embedding

The network measures defined in the previous section (except DR) yield scalar values for each node, and their distribution can be visualized by color-coded maps over the geographical domain of interest. However, the spatial embedding of the networks, and in particular the fact that they are confined to a given geographical domain, restricts the possibilities to place network links, and thereby influences the spatial distribution of network measure values (Gastner and Newman, 2006a; Gastner and Newman, 2006b; Barthelemy, 2011; Rheinwalt et al., 2012). This *bias* has to be considered and taken care of before interpreting the spatial distribution of network measure values in a climatological context. The effect of the spatial embedding on the network's topology is formally captured by the conditional probability of a link between two grid points given their geographical distance. The corresponding probability distribution  $P((i, j) \in \mathcal{E} | dist(i, j))$  is estimated from the actual network. Here,  $\mathcal{E}$  denotes the set of links in the network. For linear network measures, the spatial distribution that is expected from the spatial embedding alone can be analytically determined. For example, the expected degree is given by

$$\langle DG_i \rangle = \sum_{j=1}^N P((i, j) \in \mathcal{E} | dist(i, j)), \quad (3.20)$$

and for the expected mean geographical distance we have

$$\langle MD_i \rangle = \frac{1}{DG_i} \sum_{j=1}^N dist(i, j) P((i, j) \in \mathcal{E} | dist(i, j)). \quad (3.21)$$

The expectation values for the other linear measures defined above, namely regional connectivity, in- and out-strength, can be computed accordingly. For BC and CC, it is due to their non-linearity not as easy to obtain the expectation value analytically. Therefore, following (Rheinwalt et al., 2012), we numerically construct sufficient<sup>3</sup> surrogate networks preserving the probability distribution  $P((i, j) \in \mathcal{E} | dist(i, j))$ , compute the non-linear network measure for each surrogate network, and take the mean over all surrogate values at each node.

---

<sup>3</sup>For the applications in this thesis, 1000 surrogates have proven to be sufficient, for the results do not further change by increasing this number.

### 3.5. The influence of the spatial embedding

In order to correct for these effects of the spatial embedding, we divide each grid cell's value of a given measure  $M_i$  by the value expected from the spatial embedding  $\langle M_i \rangle$ :

$$M_i^{\text{corrected}} = \frac{M_i}{\langle M_i \rangle} \quad (3.22)$$

For a spatial plot of uncorrected, expected, and corrected versions of the measures BC, CC, and MD, we refer to Figure 4.2 in chapter 4.



Part II.

## Applications



## Chapter 4.

# Spatial Patterns of Extreme Rainfall Co-Variability

### 4.1. Summary

We apply symmetric Event Synchronization ( $ES^{sym}$ ) to daily rainfall events above the 90th and above the 95th percentiles of the South American monsoon season. Undirected Networks are constructed from the matrix  $\mathbf{ES}^{sym}$  and the classical measures betweenness centrality (BC), mean geographical distance (MD), clustering (CC), as well as the here introduced long-ranged directness (LD) are computed in order to analyze the networks' topology. Upon providing climatological interpretations for these measures, we show that their spatial distributions reveal the key features of the South American Monsoon System (SAMS), such as the main moisture pathways, areas with frequent development of Mesoscale Convective Systems (MCS), and the major convergence zones (ITCZ and SACZ). In addition, our results reveal substantial differences between the spatial structures of rainfall synchronicity above the 90th and above the 95th percentiles. Most notably, events above the 95th percentile contribute stronger to MCS in the La Plata Basin. The purpose of this chapter is twofold: Since the main features of the SAMS are relatively well-understood, the results presented here can serve as a proof of concept, showing that the proposed methodology yields results which are consistent with previous studies. In addition, however, the presented results reveal the specific role which these climatic features play for the spatial characteristics of extreme rainfall synchronicity, and – in this sense – go beyond the hitherto known. This chapter is based on the associated publication P<sub>1</sub>, and some of the following sections will closely follow the presentation in that study. Supplementary figures for this chapter can be found in appendix A.

### 4.2. Introduction

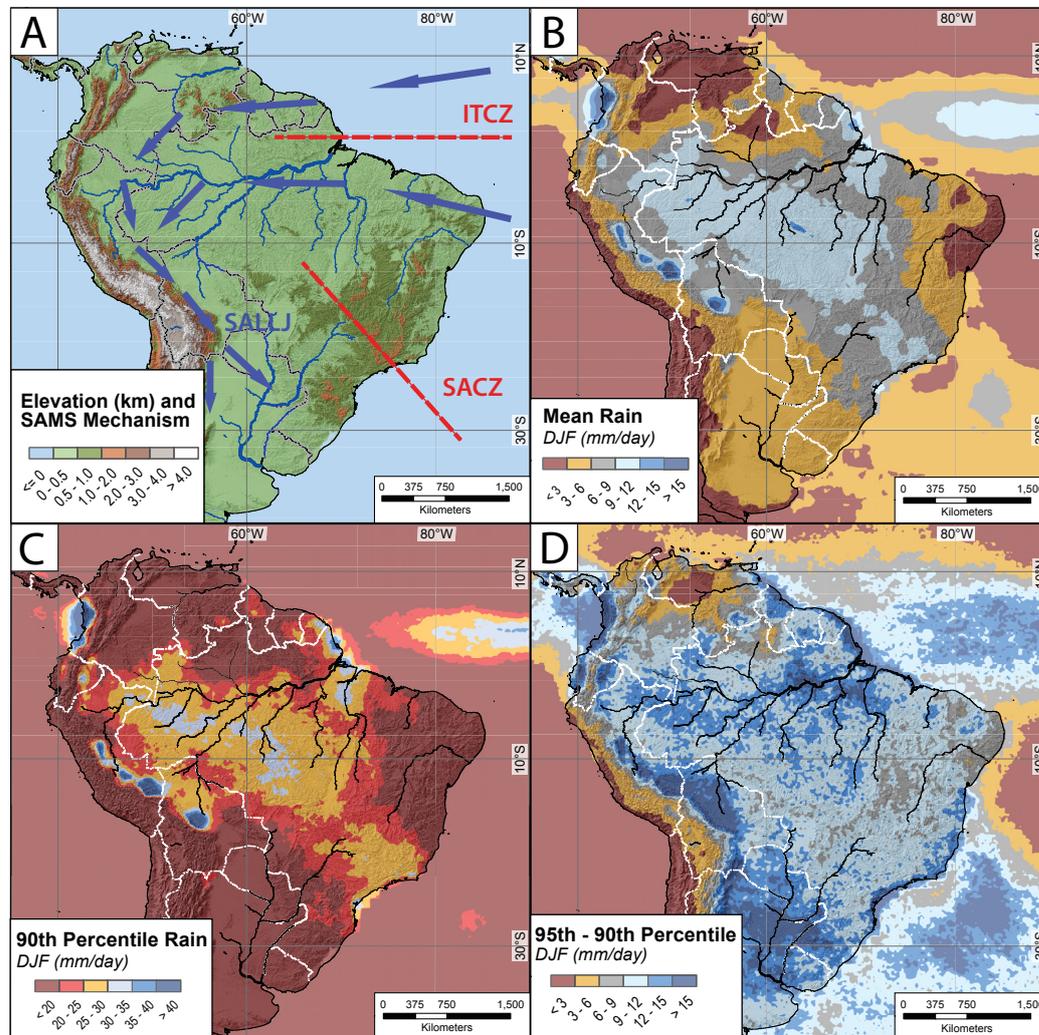
In this chapter, we use  $ES^{sym}$  and analyze extreme rainfall in the South American Monsoon System (SAMS) by means of undirected networks. This climate system is particularly well-studied and thus suitable for demonstrating how the proposed methodology reproduces known climatic features, but also how it can reveal previously undocumented results.

As explained in chapter 3, we assume that the underlying climatic mechanisms are encoded in the topological characteristics of networks calculated from the most synchronous extreme rainfall events. We quantify these topological properties by employing the network measures betweenness centrality (BC), mean geographical distance (MD), clustering coefficient (CC), as well as the here introduced long-ranged directness (LD). For each of these measures, we provide climatological interpretations, put these interpretations into relation with previous studies on the climatic mechanisms involved in the SAMS, and intend to show that our results are consistent with these climatological features.

It should be noted that most previous climate network approaches focused on large-scale analyses of low-to-moderate spatiotemporal resolution data (e.g. Tsonis et al., 2007; Gozolchiani et al., 2008; Yamasaki et al., 2008; Donges et al., 2009b). But high spatiotemporal resolution data derived from station networks or satellite data (such as the here used TRMM 3B42 daily satellite product) are crucial for identifying low-frequency and high-magnitude rainfall events. By analyzing these data with undirected networks, we unravel spatiotemporal connections between atmospheric processes that directly link to natural hazards occurring at the Earth's surface. We emphasize that the small-scale linkages, orographic rainfall barriers, and important climate patterns and meteorological mechanisms are only adequately resolved with data on high spatial and temporal scales.

### 4.3. Data

We employ the TRMM 3B42 V7 daily satellite product (Huffman et al., 2007) from January 1st 1998 to December 31st 2012 (15 years) with spatial coverage of 40°S to 15°N and 85°W to 30°W (Fig. 4.1) at a resolution of  $0.25^\circ \times 0.25^\circ$ , resulting in  $N = 48400$  time series. Rainfall sums are measured in [mm/day].



**Figure 4.1.:** A. Elevation and key features of the SAMS. B. Mean daily rainfall for the core monsoon season from December to February (DJF). C. The 90th percentile of rainfall for DJF. D. Difference of the 95th and the 90th percentile for DJF.

## 4.4. Methods

Here, we provide a brief revision of the methodology, with particular focus on testing statistical significance and the climatological interpretations of the employed network measures. We refer to chapters 2 and 3 for additional information and detailed mathematical definitions.

### 4.4.1. Event selection

At each grid point and for each three-months season (DJF, MAM, JJA and SON), we consider events above the 90th and 95th percentile of the rainfall time series (15 years of data  $\times$  90 days per season  $\times$  10% (5%) = 135 (68) events at each grid point). Grid points with less than 135 (68) wet days are discarded from the analysis (Fig. 4.1C). These are regions in the eastern Pacific and near the Atacama Desert in northern Chile with very low rainfall amounts and only very infrequent rainfall events. Thus, in order to overcome the problems related to the influence of the event rate on comparability of values of  $ES$  discussed in section 2.2.1, here we construct the event series such that we obtain the same number of events at each location.

### 4.4.2. Network construction

We compute the matrix  $\mathbf{ES}^{sym}$  (cf. equation (2.4)) using a maximal time delay of  $\tau_{max} = 3$  days. For each grid-point pair  $(i, j)$  we thus summed the values for both directions (from  $i$  to  $j$  and from  $j$  to  $i$ ), thereby losing all information about directionality. Upon calculating  $\mathbf{ES}^{sym}$  for all pairs of grid points, we consider events at a grid cell  $j$  to be synchronous to events at a grid cell  $i$  if the corresponding value  $ES_{ij}^{sym}$  is among the highest 2% of all entries of  $\mathbf{ES}^{sym}$ . We denote this by a link between  $i$  and  $j$  and will say that  $i$  and  $j$  are connected. In this way we obtain two undirected networks (one for each event threshold) for each three-months season. For the monsoon season (DJF), the lowest value of  $ES$  which is represented by a link is  $\overline{ES}^{sym} = 0.34$  (0.23) for events above the 90th (95th) percentile.

To estimate statistical significance of the network links, we construct 10000 independent event series with 135 (68) independently and uniformly randomly distributed events and calculate  $ES^{sym}$  for  $\tau_{max} = 3$  for all possible pairs. From the histogram  $H$  of all these values, we obtain an estimator of the probability of values of  $ES^{sym}$  to be larger than or equal to the network threshold  $\overline{ES}^{sym}$  (and hence the probability of a network link) given such a set of independent time series (cf. equation (2.8)):

$$P(ES^{sym} \geq \overline{ES}^{sym}) = \int_{\overline{ES}^{sym}}^1 H(s) ds \quad (4.1)$$

Note that  $H$  is in the case at hand independent from the event rates, as these are identical in each time series. For  $\overline{ES}^{sym} = 0.34$  (0.23), we obtain a  $p$ -value of  $P(ES^{sym} \geq \overline{ES}^{sym}) = 0.03$  (0.01), and thus all network links correspond to significant values of  $ES$  at the 5%-confidence level.

### 4.4.3. Network measures

We analyze the topology of these networks and thereby derive spatial patterns with the following four network measures: betweenness centrality (BC), mean geographical distance (MD), the clustering coefficient (CC), and long-ranged directedness (LD).

First, we employ betweenness centrality (BC, equation (3.7)). As explained in section 3.4.1, it is based on the concept of shortest geodesic paths in the network, which are the shortest sequences of links leading from one grid point to another. For given grid points  $j$  and  $k$ , we first look for all shortest paths between  $j$  and  $k$  and then calculate the ratio of the number of such shortest network paths that pass through a third grid point  $i$ , and the total number of shortest network paths between  $j$  and  $k$ . BC at grid point  $i$  is then defined as the sum of these ratios over all  $j$  and  $k$ . If a grid point lies on many such shortest paths between any pairs of grid points, its BC will be high and we interpret it to be important for the propagation of extreme rainfall events, in particular over large spatial distances.

Second, we calculate the mean geographical distance (MD, equation (3.9)) of the connections at a grid point. We use this measure to estimate the spatial scales at which a grid point is connected to other regions. Areas with high MD are thus likely to be part of teleconnection patterns between regions of high synchronization.

Third, we calculate the clustering coefficient (CC, equation (3.8)), which at a given grid point is defined as the relative frequency of pairs of network neighbors of this grid point that are network neighbors themselves. A region with high CC is interpreted to exhibit large spatial coherence of extreme rainfall events.

Fourth, we introduced a combination of these three measures called long-ranged directedness (LD, equation (3.10)). The three measures BC, MD, and CC illustrate complementary aspects of the network and its topology, and our analysis of the SAMS is based on their combination. This allows us to obtain more detailed and robust insights into the underlying mechanisms of the SAMS than with just one network measure alone. We classify a region's extreme rainfall characteristics from regionally clustered to long-ranged and directed by introducing LD. For example, we expect extreme rainfall in a region with low LD (resulting from low BC and low MD in combination with high CC) to occur spatially coherently (and extensively within that region) with little long-range impacts. In particular, frequent spatially extensive contiguous events will lead to this combination of network measures, which can thus be an indicator for regions with more frequent development of Mesoscale Convective Systems (MCS). On the other hand, extreme rainfall in a region with high LD values resulting from high BC and high MD in combination with low CC is expected to involve highly dynamical, directed moisture transport processes along narrow transport routes and over large spatial scales.

The fact that we only consider a bounded spatial domain can influence the spatial distribution of these four network measures, as it may impose restrictions on the placement of possible network links. This is commonly referred to as the effect of spatial embedding of the network (see section 3.5). We correct for this effect following Rheinwalt et al. (2012) by dividing the value of each measure at each grid point ( $M_i$ )

by the value expected from the spatial embedding alone (cf. equation (3.22)). The expectation value  $\langle M_i \rangle$  is either obtained analytically, or – for measures where no analytical formula exists – obtained by averaging  $M_i$  over a collection of 1000 random surrogate networks preserving the relative frequency of links given their geographical distance (Figure 4.2).

## 4.5. Results

### 4.5.1. Extreme rainfall (> 90th percentile)

Here, we present our results for the 90th percentile of the peak monsoon season (DJF). Figures for the remaining seasons and for events above the 95th percentile can be found in Appendix A. For all network measures we show the values *after* correcting for the effects of the spatial embedding (Figure 4.2).

We observe a region of high BC and MD but rather low CC and therefore high LD over the tropical Atlantic Ocean between the equator and 5°N (Fig. 4.3). This corresponds to the climatological position of the ITCZ.

The same combination of BC, CC, and MD, resulting in even higher LD, can be observed over a large continuous area in the central Amazon Basin around 5°S, 60°W (Fig. 4.3).

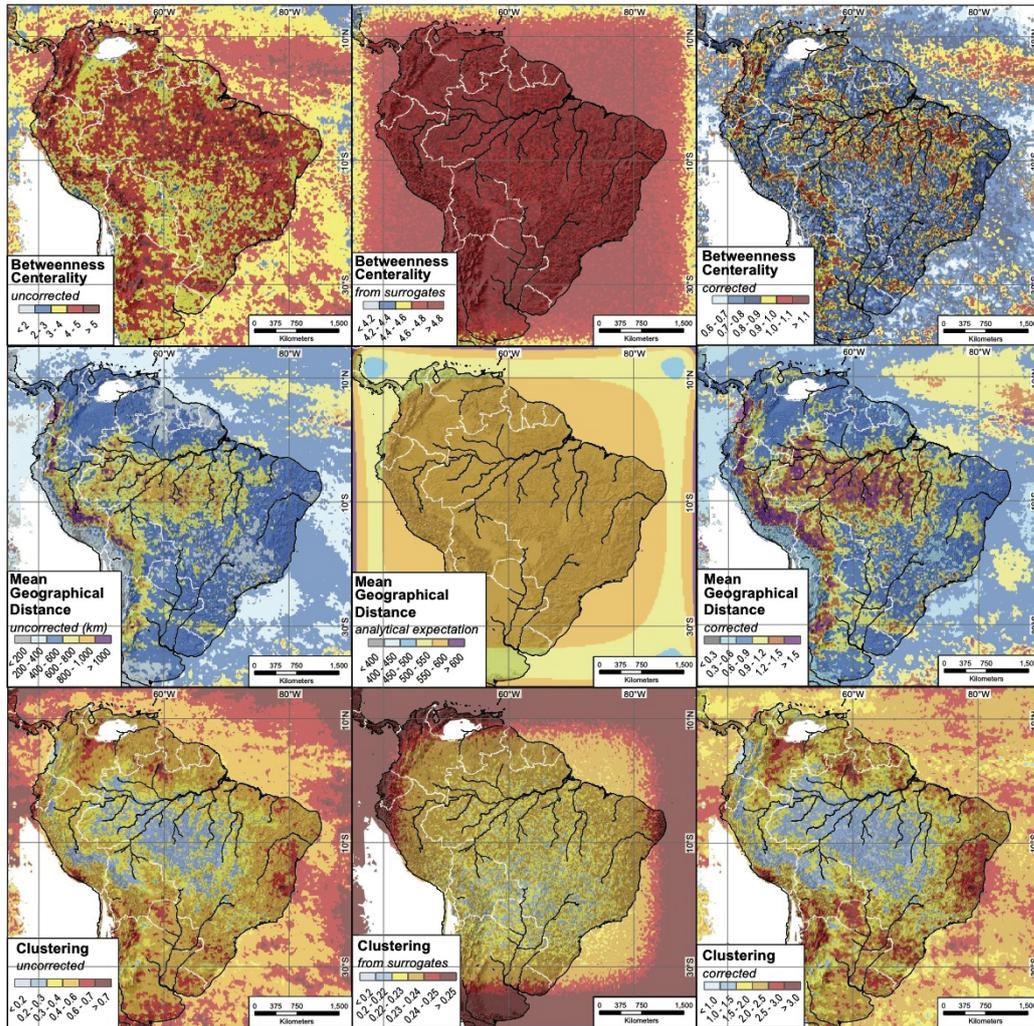
Moderately high LD-values extend from the central Amazon southeastward towards southeast Brazil (Fig. 4.3D). This wide branch, meeting the coast at about 20°S, corresponds to the climatological position of the continental part of the SACZ (Carvalho et al., 2004).

Starting at about 10°S, 70°W, a concise narrow band of high LD extends south-eastwards along the eastern slope of the Andes. In northern Argentina, this band splits into two branches, one leading southwards until approximately 30°S (Fig. 4.3), and one leading eastwards, reaching the Atlantic coast in southernmost Brazil. The position of the northern part of this signature, before it splits into two branches, coincides with that of strong orographic rainfall at the eastern Andean slopes associated with low-level winds transporting moisture from the Amazon Basin towards the subtropics (Bookhagen and Strecker, 2008), (Fig. 4.1).

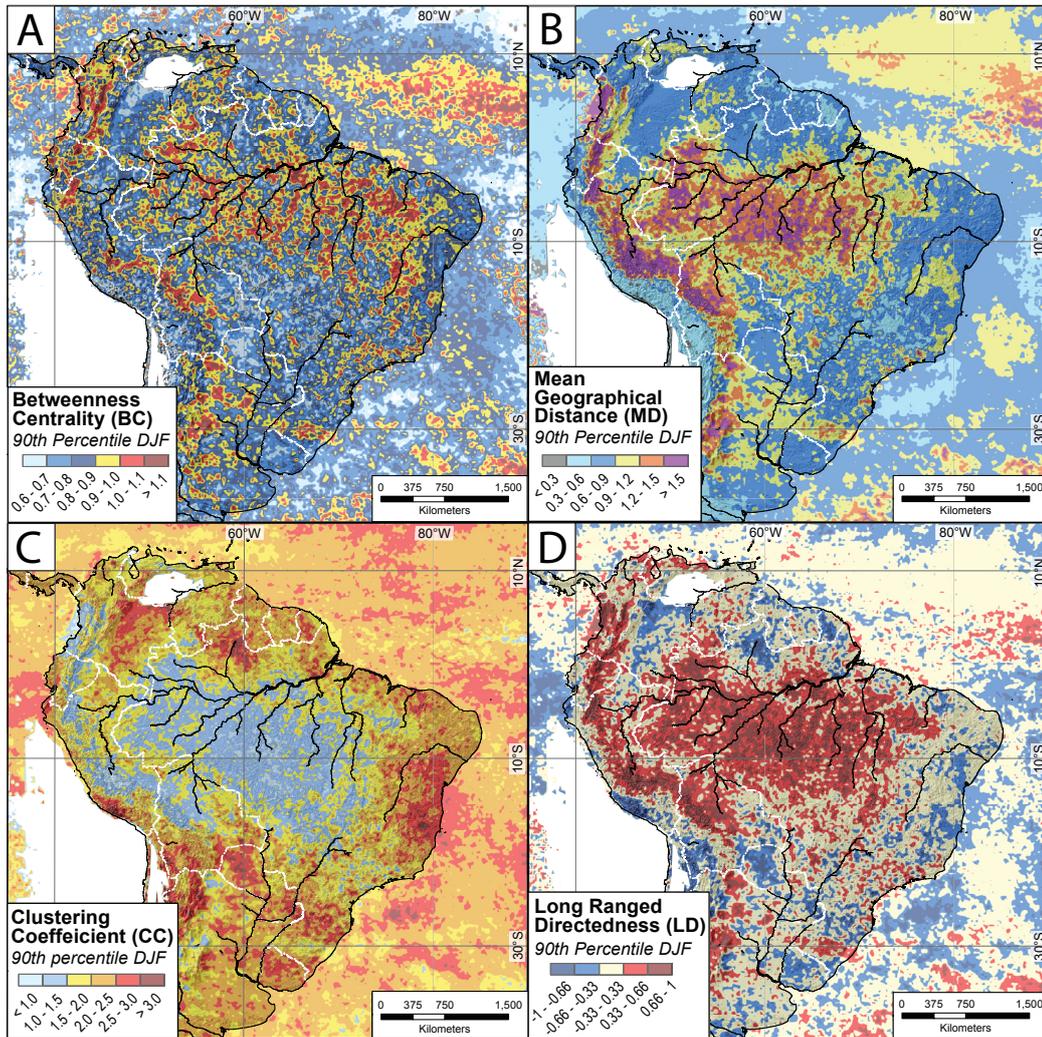
Large parts of the La Plata Basin in Paraguay, southern Brazil, NE Argentina and Uruguay are characterized by high CC, rather low BC and MD and thus low LD (Fig. 4.3). These areas are known for their frequent development of MCS (Salio et al., 2007). This region is intermitted by the eastern branch of high LD mentioned in the previous paragraph.

### 4.5.2. Most extreme rainfall (> 95th percentile)

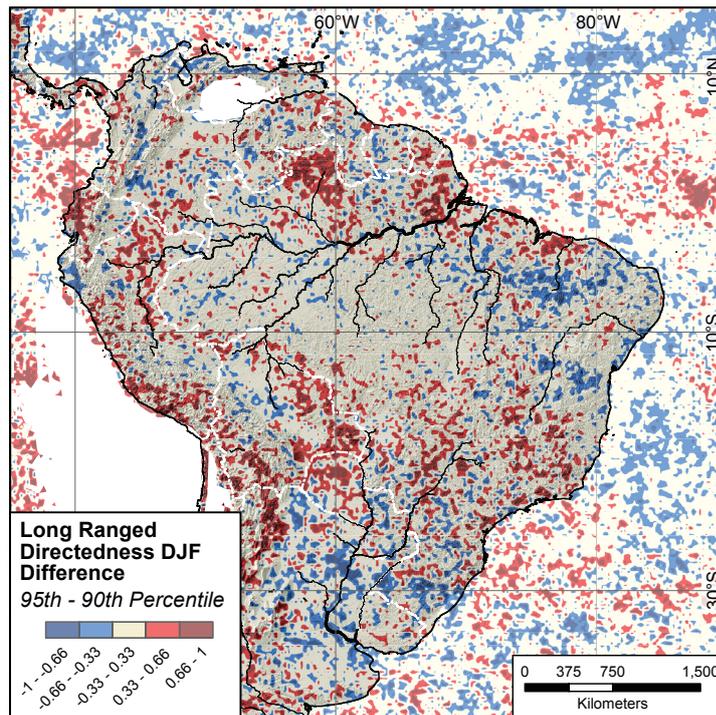
When only events above the 95th percentile are taken into account, LD is decreased as compared to the 90th percentile in the La Plata Basin. In contrast, LD exhibits increased values in the SACZ, in western Paraguay and at the outlet of the Amazon River (Fig. 4.4).



**Figure 4.2.:** The influence of the spatial embedding on the spatial distribution of the network measures betweenness centrality (BC, top row), mean geographical distance (MD, middle row), and clustering coefficient (CC, bottom row). The left column shows the uncorrected measure distributions, the middle column shows the spatial distribution of the three measures that would be expected from the spatial embedding alone, and the right column shows the corrected values. Corrections are carried out by dividing the uncorrected values by the values expected from the spatial embedding (cf. equation (3.22) and the corresponding explanations).



**Figure 4.3.:** A. Betweenness centrality (BC) on a logarithmic scale for DJF. This measure indicates the importance of a grid point for the long-range spatial propagation of extreme rainfall events. We observe high values in several areas that are crucial for water-vapor transport, especially in the central Amazon Basin and along the orographic barrier of the eastern Andes. B. Mean geographical distance (MD) for the same time period. For example, high values are found in the central Amazon Basin and along the eastern Andean slopes, which can be explained by their important role for moisture transport to the subtropics and the large spatial scales involved. C. The clustering coefficient (CC) for the same time period. It measures the degree of spatial coherence in the distribution of extreme rainfall. Note the high values in southeastern South America, where Mesoscale Convective Systems (MCS) frequently develop. D. Long-ranged directedness (LD) for the same time period. Low values indicate areas where extreme rainfall occurs regionally coherently as e.g. in MCS, while high values indicate areas where extreme rainfall propagates on narrow transport routes over large spatial distances, as e.g. along the eastern Andes slopes.



**Figure 4.4.:** Difference of LD of the 95th and the 90th percentile for DJF. High values indicate areas which are more important for the large-scale propagation of events above the 95th percentile, while low values indicate areas where regional clustering of events (as e.g. in MCS) enhances for events above the 95th percentile.

## 4.6. Discussion

### 4.6.1. Rainfall > 90th percentile

The region adjacent to the ITCZ is identified as the most important source for the large-scale distribution of extreme rainfall over the entire continent, as it is the only oceanic region in the considered domain with high BC.

The importance of the central Amazon Basin for the spatial distribution of extreme rainfall to most of the remaining continent can be explained by its geographical position combined with its dense vegetation cover: Rapid propagation of squall lines from the northeastern coast of Brazil into the Amazon Basin are responsible for most of the spatial distribution of extreme rainfall in this region (Garreaud and Wallace, 1997). The landfall of these thunderstorm systems leads to locally high clustering near the outlet of the Amazon River (Fig. 4.3C). These storm lines are followed by enhanced moisture inflow and convective activity because of latent heat release and the resulting strengthening of low-level easterlies (Cohen et al., 1995). Evapotranspiration recharges the moisture content of the low-level atmosphere (Eltahir and Bras, 1993; Lettau et al., 1979) and the deep convection leads to further extreme rainfall. At the same time, the enhanced low-level winds transport the moisture further west and south, which leads to synchronized rainfall there and thus explains the high BC values in the central Amazon Basin (Fig. 4.3A).

The SACZ is shaped by the continental wind patterns mainly resulting from the interplay of trade winds and the orographic barrier of the Andes (Carvalho et al., 2004). This suggests that extreme rainfall in this area is dictated by long-ranged, directed transport processes, which is consistent with high LD (Fig. 4.3D).

The band of high LD along the eastern slopes of the Bolivian Andes (Fig. 4.3D) highlights the importance of this region for the large-scale moisture transport (we emphasize the particularly high MD in this region, Fig. 4.3B). Its geographical position suggests that this network signature is due to the interplay of orographic effects and the conveyor belt providing moisture for subtropical South America (Bookhagen and Strecker, 2008; Arraut et al., 2012). It is located slightly farther west than the climatological position of SALLJ (Marengo et al., 2004), presumably because orographic effects overprint the moisture advection signal related to these wind systems. In other words, since our approach is only based on rainfall events, the resulting network measures reveal locations that are crucial for the synchronicity of extreme rainfall associated with large-scale moisture transport, but not the transport processes themselves. These rainfall events are strongly enhanced by orographic effects, which shift the patterns towards the Andes, i.e. westwards from the climatological position of the main wind routes. We stress that only high spatiotemporal resolution data are able to uncover the spatial distinction of these features.

The position of the branch of high LD extending southwards from the border of Bolivia and Argentina suggests that this pattern corresponds to extreme rainfall associated with moisture transport by the Chaco Jet, a southward extension of the SALLJ controlled by the Chaco Low (Salio et al., 2002; Saulo et al., 2004). We

interpret the branch of high LD extending eastward from the border of Bolivia and Argentina to be the transport route towards the La Plata Basin. This area is characterized by high synchronicity of extreme events, and it is consistent in shape with the climatological moisture transport route (Arraut et al., 2012).

Furthermore, the network measures identify regions with frequent occurrences of MCS in southeastern South America, characterized by low LD. In particular, we emphasize the high CC in Paraguay, the typical exit region of the SALLJ, which provides moisture for MCS. We note that areas corresponding to the eastern branch of high (instead of low) LD have been previously described to exhibit frequent MCS development as well (Durkee et al., 2009; Durkee and Mote, 2009). This discrepancy stems from an additional climatological process, namely the directed propagation of extreme events along the rather narrow eastern branch of high BC, which seems to overprint the expected signature of high CC.

#### 4.6.2. Comparison of spatial patterns for rainfall above the 90th and above the 95th percentiles

The substantial differences in LD between extreme ( $> 90$ th percentile) and most extreme ( $> 95$ th percentile) rainfall suggest different spatial patterns for the propagation of events (Fig. 4.4). For the most extreme events, the role of the SACZ is enhanced (Fig. 4.4). We attribute this to higher synchronicity of events above the 95th percentile along the SACZ due to an increased number of events during phases of intensified SACZ, as was previously described by Carvalho et al. (2004). Decreased LD values for events above the 95th percentile in the La Plata Basin indicate an enhanced role of MCS for the most extreme rainfall events in this region because, as described above, low LD is interpreted to indicate regions with frequent MCS development. On the other hand, increased values of LD in western Paraguay suggest that the SALLJ is more important for the directed propagation of events above the 95th percentile in this region. Similarly, we infer that the squall lines in NE Brazil are more important for the large-scale propagation of the most extreme events.

## 4.7. Conclusion

We analyzed the synchronicity of extreme daily rainfall events of the South American Monsoon System (SAMS) by means of undirected networks. The networks were constructed from the symmetric version of Event Synchronization,  $ES^{sym}$ . This similarity measure was applied to daily seasonal extreme events, which were obtained from high spatiotemporal resolution rainfall data (daily data at  $0.25^\circ \times 0.25^\circ$ ) derived from remote-sensing measurements. We introduced a new combination of network measures, called long-ranged directedness, which quantifies the characteristics of spatial distribution and geographic connectivity of extreme rainfall. The spatial patterns exhibited by this new measure reveal the most important large-scale features of the SAMS when considering events above the 90th percentile. These include the propagation of mesoscale storm systems, areas of deep convection in the Amazon Basin,

the main rainfall-transport routes to the subtropics including the South American Low-Level Jet and its interplay with orographic processes at the eastern Andean slopes, as well as the Intertropical Convergence Zone (ITCZ) and the South Atlantic Convergence Zone (SACZ). In a final step, we have analyzed the differences in spatial patterns between network measures derived for rainfall events above the 90th and above the 95th percentile. The resulting patterns reveal that the South American Low-Level Jet as well as Mesoscale Convective Systems play an enhanced role for the most extreme events. The simplicity of the approach, together with the consistency of the results with previous studies of the SAMS, shows the unique potential of the developed methodology for climatological studies.

## Chapter 5.

# Extreme Rainfall Associated with the South American Rainfall Dipole

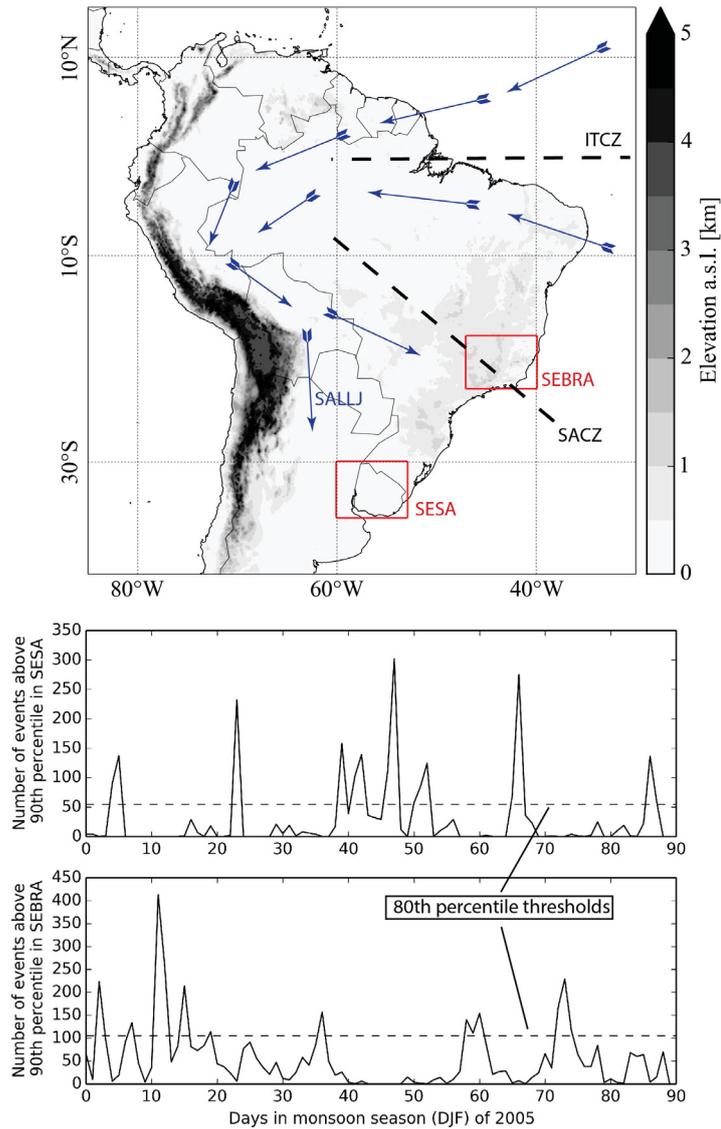
### 5.1. Summary

Intraseasonal rainfall variability of the South American monsoon system is characterized by a pronounced dipole between the region around Buenos Aires in southeastern South America and southeastern Brazil, including São Paulo and Rio de Janeiro. Here, we analyze the dynamical properties of extreme rainfall events associated with the rainfall dipole between these two regions by computing the matrix  $\mathbf{ES}^{sym}$  separately for the two phases of the dipole. By construction, the event rates will typically be different from one location to another, calling for an adapted way to determine statistical significance of the entries of  $\mathbf{ES}^{sym}$ . We make the following main observations in this chapter: i) Our approach reveals the dominant climatological propagation routes of extreme events corresponding to the two dipole phases. ii) While extreme rainfall propagation in the tropics is directly driven by the trade winds and their deflection by the Andean Cordillera, 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 suggests that the influence of Rossby waves emanating from the western Pacific Ocean, which cause frontal systems over South America and impact large-scale circulation patterns, extends beyond the equator. This chapter is related to the associated publications P2 and P3 and the following presentation will closely follow their content. Supplementary figures for this chapter can be found in appendix B.

### 5.2. 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 São Paulo and Rio de Janeiro between between 18°S and 23°S (see Figure 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,



**Figure 5.1.:** Top: Topography of South America and key features of the South American monsoon system, including 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 ( $30^{\circ}\text{S}$  to  $35^{\circ}\text{S}$  and  $60^{\circ}\text{W}$  to  $53^{\circ}\text{W}$ ) and SEBRA ( $18^{\circ}\text{S}$  to  $23^{\circ}\text{S}$  and  $47^{\circ}\text{W}$  to  $40^{\circ}\text{W}$ ) 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.

but also particularly vulnerable to damages caused by extreme rainfall and associated floodings and landslides (Berbery and Barros, 2002; Carvalho et al., 2002; Barros et al., 2006; Marengo et al., 2013b).

As outlined in section 1.2, rainfall in South America 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 mountains and channeled southwards. 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., 2007; 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., 2004). The dipolar behavior of rainfall between SESA and SEBRA has been described as 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, as explained in chapter 1: 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 in the context of the South American rainfall dipole by employing and refining the methodology developed in the previous chapters.

### 5.3. Data

As in the previous chapter, we employ the satellite-derived and gauge-calibrated rainfall data product from the Tropical Rainfall Measurement Mission (TRMM 3B42 V7, (Huffman et al., 2007)) with daily temporal and  $0.25^\circ \times 0.25^\circ$  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  $1/2^\circ \times 2/3^\circ$  spatial resolutions. The spatial domain is confined to the coordinates  $40^\circ\text{S}$  to  $15^\circ\text{N}$  and  $85^\circ\text{W}$  to  $30^\circ\text{W}$  (Figure 5.1). All datasets are analyzed for the monsoon seasons (December to February, DJF) between 1998 and 2012.

## 5.4. Methods

### 5.4.1. Extreme rainfall events

For each of the  $N = 48400$  time series  $x_i$ , we define extreme rainfall events as days on which rainfall amounts exceed the 90th percentile of the rainfall distribution restricted to *wet days*, i.e. days with rainfall sums larger than  $0.01\text{mm/day}$ . This percentile threshold is local in the sense that it depends on the respective grid cell's rainfall distribution. Different grid cells will thus by construction exhibit different numbers of events  $l$  and thus also different event rates  $r$ .

### 5.4.2. 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 80th percentile (Figure 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 procedure 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 networks derived from  $\mathbf{ES}^{sym}$  will in the following be constructed separately for these two phases.

### 5.4.3. Network construction

We will construct two separate networks for the two phases of the dipole in the following way: For the computation of the Event Synchronization matrix  $\mathbf{ES}^{sym}$  for the SESA (SEBRA) phase, we will only consider those binary entries in the event 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 empirical value  $\overline{ES}_{ij}^{sym}$ . For this purpose, we build 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: Upon constructing 1000 surrogate pairs of event time series for each combination of event rates  $(r_i, r_j)$ , we compute  $ES_i^{sym} j$  for all these surrogate pairs

$(i, j)$ , which yields a relative frequency distribution of values of  $ES^{sym}$  consistent with the assumptions of the null model. We infer the score of the 95% significance level from this distribution following the procedure outlined in section 2.2.1 by determining the threshold  $T_{r_i, r_j}$  such that

$$P(ES_{ij} \geq T_{r_i, r_j}) = \int_{T_{r_i, r_j}}^1 H_{r_i, r_j}(s) ds < 0.05 . \quad (5.1)$$

A network link will be placed between grid cells  $i$  and  $j$  if the corresponding empirical value  $\overline{ES_{ij}^{sym}}$  is above  $T_{r_i, r_j}$ . For two arbitrary grid cells  $i$  and  $j$ , this can be formally written in terms of the adjacency matrix  $\mathbf{A}$  as

$$A_{ij} = \Theta(\overline{ES_{ij}^{sym}} - T_{r_i, r_j}) - \delta_{ij}, \quad (5.2)$$

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

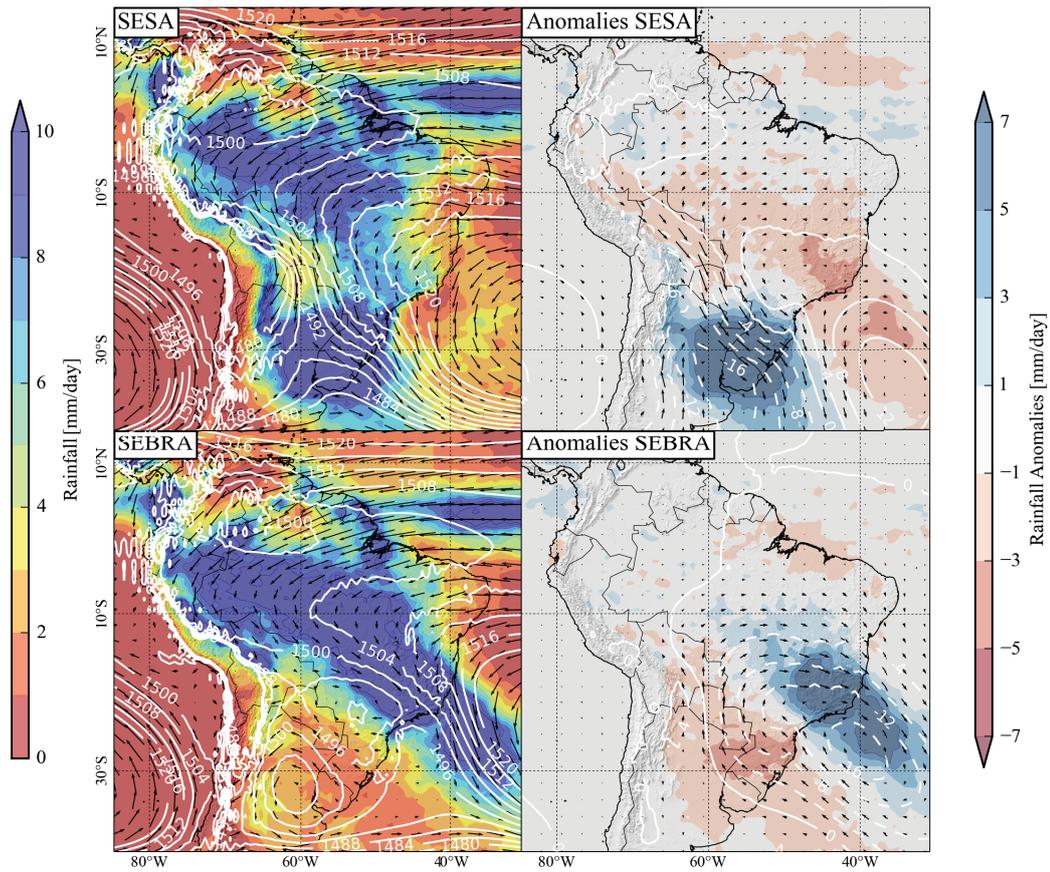
#### 5.4.4. Network measures

In this chapter, we consider two network measures in order to reveal the synchronization pathways of extreme events corresponding to the two dipole phases. First, we compute the *degree* (DG) as defined in equation (3.6). At each location  $i$ , this measure yields the number of other grid cells where extreme events occur synchronously with extreme events at  $i$ . In this sense, we expect DG to provide a local (i.e., based on next neighbors in the network) estimate for a location's importance for the propagation of extreme events. Second, we are interested in the directions along which extreme events occur synchronously. For this purpose, we employ the measure *directionality* (DR) defined in equations (3.13) and (3.14). Following the explanations of this measure in chapter 3, under the assumption that the temporal resolution of one day is sufficiently high, we expect that rainfall clusters typically propagate perpendicular to the streamlines by which this measure is visualized, which we refer to as isochrones.

## 5.5. Results

### 5.5.1. Atmospheric Conditions

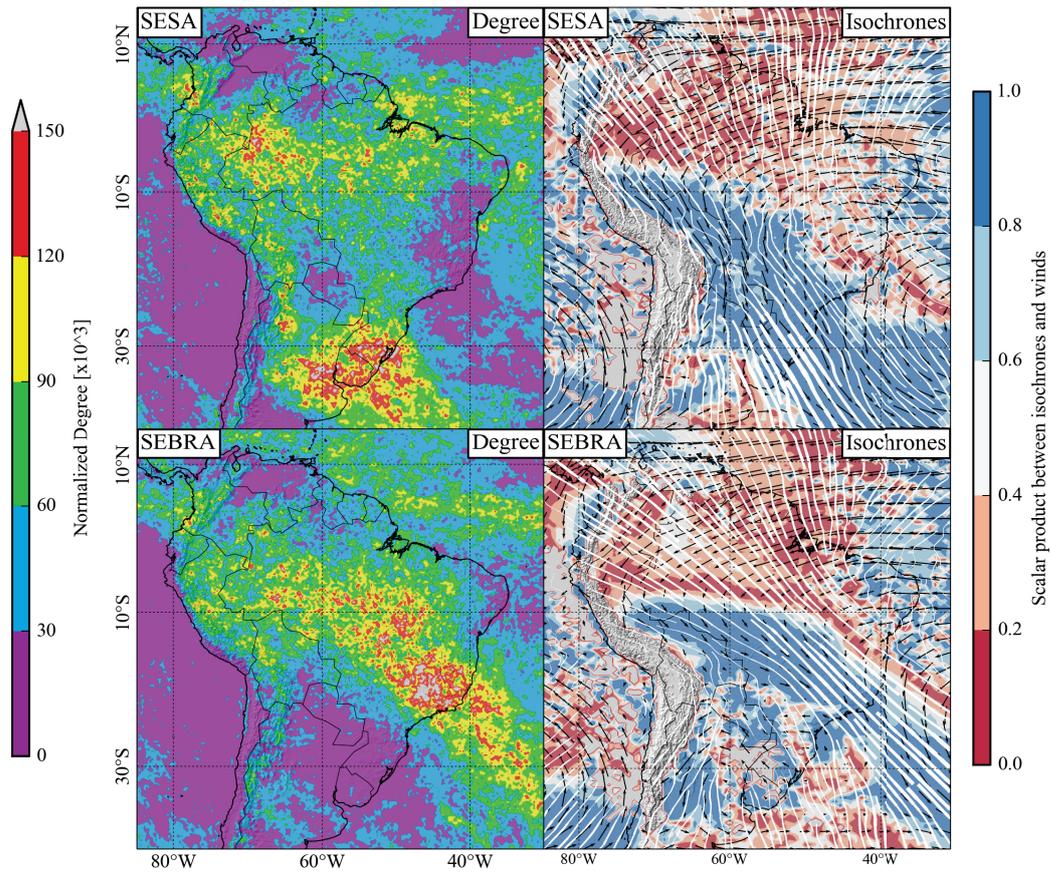
Composites of geopotential height and wind fields constructed separately for the two different dipole phases show distinctively different features (Figure 5.2). As expected, we find anomalously high rainfall amounts over SESA for the SESA phase, but negative anomalies over SEBRA (top row of Figure 5.2).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



**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.

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 Figure 5.2), rainfall composites show pronounced positive 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 (e.g. Carvalho et al., 2002; Liebmann et al., 2004; Carvalho et al., 2004).



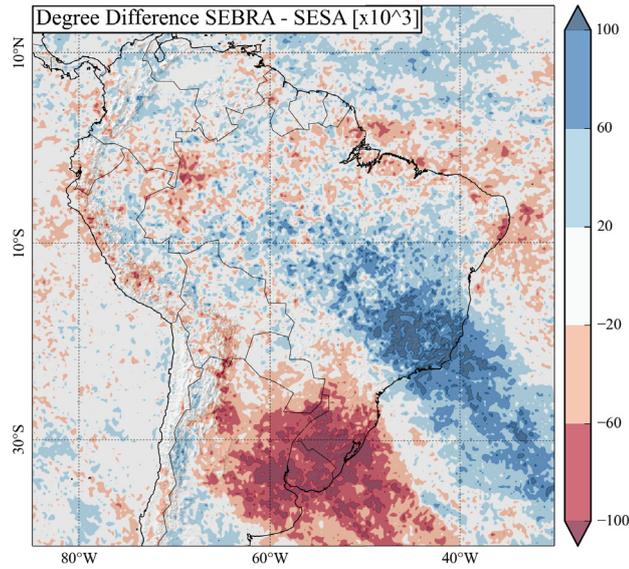
**Figure 5.3.:** Left: Degree (DG) normalized by the respective link density of the network for the SESA (top) and SEBRA (bottom) phase. Right: Isochrones, wind fields at 850mb as well as the absolute value of the scalar product between wind vectors and isochrones for the SESA (top) and the SEBRA (bottom)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

### 5.5.2. 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 (see Figures B.1 and B.2 in Appendix B).

### 5.5.3. Degree (DG)

For the SESA phase (top left in Figure 5.3), we observe high values of DG over the ITCZ, the entire Amazon Basin, along the eastern slopes of the Andes from northern



**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.

Peru to northern Argentina, as well as over SESA. In contrast, we find low DG values over SEBRA.

For the SEBRA phase (bottom left in Figure 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  $\tau$  (equation (2.2)), 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  $\tau_{max} = 2$  (shown in the appendix chapter B).

When subtracting the DG field for the SEBRA phase from the DG field obtained for the SESA phase (Figure 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.

#### 5.5.4. Directionality (DR)

For the reason explained in section 5.4.4, directionality is calculated for networks constructed from ES with  $\tau_{max} = 0$ , i.e. only events at the very same day are counted as synchronous. 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 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 than 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 Figure 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°S, 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.

## 5.6. 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 (e.g. Zhou and Lau, 1998; Vera et al., 2006; Marengo et al., 2012), 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 Figure 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  $20^\circ$ , 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 Figure 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 (Figure 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 than 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  $40^\circ\text{S}$  and  $15^\circ\text{N}$ . 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 (Lenters and Cook, 1999; Rodwell and Hoskins, 2001). The observed oscillation suggests that these Rossby waves control extreme-rainfall variability over

the entire South American continent.

## 5.7. 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 São 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 could 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 SEBRA. By comparing climatological wind directions with the orientations of streamlines of synchronous extreme rainfall, we revealed 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.

# Spatiotemporal Connectivity of Extreme Rainfall in the Andes

### 6.1. Summary

The South American Andes are frequently exposed to intense rainfall events, which lead to landslides and downstream flooding, resulting in significant infrastructural and socio-economic damage. These events have varying moisture sources and precipitation-forming processes. In this chapter, we assess their spatiotemporal characteristics and geographical origins over the South American continent. We define four different types of rainfall events: (1) Events which are determined by their high intensity alone (magnitude), (2) long-lasting events (temporal extent), (3) events with large spatial extent, and (4) high magnitude, long-lasting, and spatially extensive events. In a first step, we analyze the frequencies, durations, and spatial extents of these events and integrate their impact on the underlying hydrologic catchments. Our results indicate that events of type 1 make the overall highest contributions to total seasonal rainfall (up to 50%). However, each consecutive episode of the infrequent events of type 4, lasting on average 6 to 12 hours, still accounts for up to 20% of total seasonal rainfall in the subtropical Argentinean plains. In a second step, we determine the geographic source regions of intense rainfall on the high-elevation Altiplano-Puna Plateau as well as in the main river catchments along the foothills of the Andes mountain range using directed Event Synchronization ( $ES^{dir}$ ) and constructing directed networks. We resolve the synchronization structure of intense rainfall events at different locations and unravel long-ranged climatic linkages. Our results suggest that one or two particularly large squall lines per season, originating from northeastern Brazil, propagate towards the eastern Andes and reach the high-elevation Altiplano. In general, we observe that extreme rainfall in the catchments north of approximately 20°S typically originates from the Amazon Basin, while extreme rainfall at the Andean foothills south of 20°S and the Puna Plateau originates from southeastern South America. This chapter is based on the associated publication P<sub>4</sub>, and the following sections will closely follow that publication. Supplementary figures can be found in appendix C.

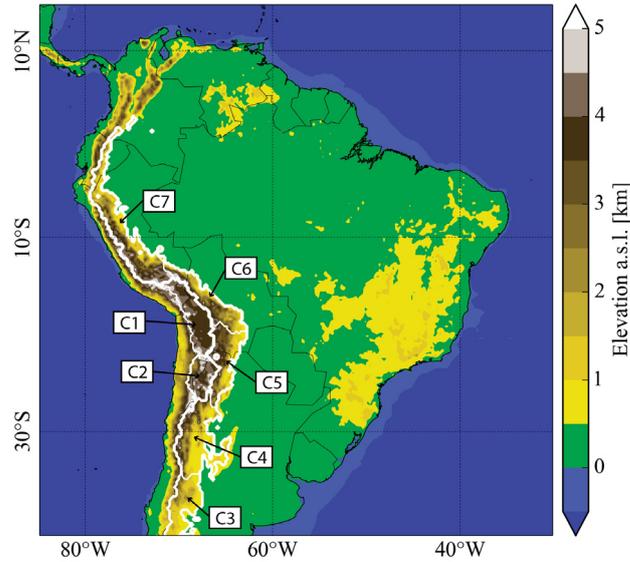
## 6.2. Introduction

During the monsoon season from December to February, the South American Andes are exposed to intense rainfall events that frequently lead to landslides and downstream flooding, with severe socio-economic impacts (Moreiras, 2005a; Harden, 2006; Barros et al., 2006; Programa de las Naciones Unidas para el Desarrollo (PNUD), 2011). These events frequently occur as large and long-lasting thunderstorms. Characteristic examples are the so-called Amazonian squall lines (Cohen et al., 1995) as well as mesoscale convective systems (Maddox, 1980; Griffiths et al., 2009; Romatschke and Houze, 2013; Boers et al., 2013), which attain largest spatial scales in southeastern South America (Durkee et al., 2009; Durkee and Mote, 2009). In addition, the interplay between the Andean orography and low-level easterly winds carrying large amounts of moisture from the Amazon Basin (e.g. Poveda et al., 2014) is responsible for intense rainfall peaks at the eastern slopes of the Andes (Bookhagen and Strecker, 2008; Houze Jr., 2012). The connection between large orographic thunderstorms and flooding has for example been studied by Smith et al. (2011) for the case of the North American Appalachians.

The local characteristics of extreme rainfall in terms of their frequency distribution has been thoroughly studied in the literature (e.g. Renard et al., 2013; Papalexiou and Koutsoyiannis, 2013; Serinaldi and Kilsby, 2014). In contrast, the temporal durations and spatial extensions of extreme rainfall events in South America have attracted much less attention (see e.g. Khan et al., 2007, for an exception), although these characteristics are certainly of great hydrological importance. Furthermore, the dynamical properties of extreme rainfall events, whilst taking into account their spatiotemporal characteristics, need to be known for a robust assessment of their geographical origins, and thereby of their potential predictability. This becomes particularly relevant when identifying the sources of extreme rainfall in given river catchments. Here, we will focus on four different types of rainfall events: (1) events which are solely determined by their high magnitude; (2) events which are characterized by their long duration *and* high magnitude; (3) spatially extensive events; and (4) high magnitude, long-lasting and spatially extensive events.

In a first step, we will analyze the frequencies, intensities, and spatial extensions of the four proposed event types and quantify their respective contribution to the overall water budget of the main drainage basins along the Andean mountain range, including several sub-basins of the Amazon and the La Plata rivers. We are particularly interested in the spatial scales of rainfall clusters. We will determine the size of these clusters by the number of grid cells belonging to a connected component of simultaneous events and analyze the frequency distribution of these cluster sizes.

In a second step, we will investigate the geographic source areas of the four proposed types of rainfall events for the mountainous Andean drainage basins, including the high-elevation Altiplano-Puna Plateau. The spatiotemporal characteristics of rainfall events on the Altiplano-Puna Plateau are crucial for understanding and sustaining water resources in this region. While this area forms an almost continuous topographic plateau with a mean elevation of  $4041 \pm 524$  m and constitutes the



**Figure 6.1.:** Topography of South America and the seven catchments C1 to C7 referred to in the text, delineated by white contour lines. Their spatial extensions are derived from SRTM data (Farr et al., 2007).

second-largest plateau on earth (after the Tibetan Plateau), the geomorphic and climatic characteristics of its northern and southern parts are distinct (Messerli et al., 1997; Vuille, 1999; Garreaud and Aceituno, 2001; Garreaud et al., 2003; Bookhagen and Strecker, 2012). We therefore divide this arid 1800km long and 350km to 400km wide region into the northern Altiplano (C1) and the southern Puna de Atacama (C2) Plateaus (Figure 6.1). In addition to the high-elevation Altiplano-Puna Plateau, the origins of extreme rainfall are of particular hydrological importance in the mountainous drainage basins along the Andes. We will divide them into five sub-catchments roughly corresponding to the southern Andes (C3), the southern Central Andes (C4), the Central Andes (C5), the northern Central Andes (C6), and the northern Andes (C7) (Figure 6.1). We will occasionally refer to the northern and northern Central Andes (C6 and C7) as tropical, and to the remaining parts as subtropical. In order to determine the geographical origins of extreme rainfall in these river catchments, we employ directed Event Synchronization ( $ES^{dir}$ , cf. equation (2.7)) and construct directed networks. In addition, we intend to take into account different temporal and spatial scales of rainfall events. For this purpose, we first carry out several steps of temporal and spatial pre-processing. We then resolve the source regions of rainfall in the mountainous Andean catchments (C1 to C7) with respect to the four types of rainfall events described above.

### 6.3. Data

As before, we employ the gauge-calibrated remote-sensing derived rainfall dataset TRMM 3B42 V7 of the Tropical Rainfall Measurement Mission (Huffman et al., 2007) for the time period from 1998 to 2012 for the spatial domain 15°N to 40°S and 85°W to 30°W with a spatial resolution of  $0.25^\circ \times 0.25^\circ$ . However, here we employ the 3-hourly version of this dataset. For each 3-hourly time step, values are provided as average rainfall rate in units of  $[mm/h]$ .

River catchments have been delineated from resampled topographic data (Shuttle Radar Topographic Mission, (Farr et al., 2007)) at a spatial resolution of 15 arcseconds ( $\approx 500m$ ).

### 6.4. Methods

#### 6.4.1. Definition of extreme events

We will focus on the monsoon season in South America from December to February (DJF). We analyze the following three different spatial and temporal characteristics of 3-hourly rainfall events: Their *intensity*, *duration*, and *spatial extent*.

Our goal here is to disentangle the impacts and origins of extreme rainfall in the main Andean drainage basins with respect to their intensity, duration, and spatial extent. Depending on these characteristics, we apply several steps of preprocessing:

1. Let the rainfall time series at location  $i$  be denoted by  $\tilde{t}_i$ , where  $1 \leq i \leq N = 48400$ . We can apply a running-mean filter in order to account for the temporal duration of rainfall events:

$$t_i^\sigma := \frac{\sum_{s=-w}^w \tilde{t}_i^{\sigma+s}}{2w+1}, \quad (6.1)$$

where  $\sigma$  is the time index and  $2w+1$  is the width of the running-mean filter. For ease of notation, we omit making the dependency of  $t_i^\sigma$  on  $w$  explicit. The application of this running-mean filter will in the following be referred to as *smoothing*.

2. For each filtered time series  $t_i$ , we define local event thresholds  $T_i^p$  by computing the  $p$ th percentile of all  $t_i^\sigma$  with  $t_i^\sigma > 0.2mm$ . The latter condition assures that only significant rainfall amounts ("wet times") are taken into account when computing the percentile thresholds. *Local events* are then defined as time steps  $\sigma$  for which  $t_i^\sigma$  is above this threshold:

$$\tilde{e}_i := \{\sigma | t_i^\sigma > T_i^p\} \quad (6.2)$$

Again for the sake of clarity, we omit to make explicit that  $\tilde{e}_i$  depends on  $w$  as well as on  $p$ .

3. We are also interested in the spatial extent of events, given by the size of connected components  $\{C_m^\sigma\}_{1 \leq m \leq N_\sigma}$  with simultaneous events  $\tilde{e}_i$  at a given time  $\sigma$ . Here, grid points are considered to be connected if they are horizontally, vertically,

or diagonally adjacent in space, and  $N_\sigma$  denotes the total number of connected components at time step  $\sigma$ . To each grid point  $i$ , we then assign the size of the component it belongs to:

$$SC_i^\sigma = \sum_{\{\mathcal{C}_m^\sigma\}} \delta_{\mathcal{C}_m^\sigma}(i) |\mathcal{C}_m^\sigma|, \quad (6.3)$$

where  $|\mathcal{C}_m^\sigma|$  denotes the cardinality of  $\mathcal{C}_m^\sigma$  (i.e., the number of elements in that set) and  $\delta_{\mathcal{C}_m^\sigma}(i) = 1$  if  $i \in \mathcal{C}_m^\sigma$  and  $\delta_{\mathcal{C}_m^\sigma}(i) = 0$  otherwise. We then define a component size threshold  $S_i^q$  at each grid cell  $i$  as the  $q$ th percentile of all  $SC_i^\sigma$  for which  $SC_i^\sigma > 0$  (i.e., the  $q$ th percentile of the set  $\{SC_i^\sigma | \sigma \text{ such that } SC_i^\sigma > 0\}$ ).

We then define *spatially extensive events* as local events which belong to a connected component that is larger than the corresponding  $S_i^q$ :

$$e_i := \{\tilde{e}_i | SC_i^{\tilde{e}_i} > S_i^q\}, \quad (6.4)$$

which depends on the width of the temporal filter  $w$ , the intensity threshold  $T_i^p$ , and the threshold for the cluster size  $S_i^q$ .

We define the following four types of events that we will analyze in the following sections:

1. Local and short extreme events (LSE), which are intense rainfall events (high magnitude), regardless of their duration and spatial extent:  $w = 0$ ,  $p = 90$ , and  $q = 0$
2. Local and long-lasting extreme events (LLE), which are long-lasting intense rainfall events, but without condition on their spatial extent:  $w = 5$ ,  $p = 90$ , and  $q = 0$
3. Spatially extensive events (SEE), which are not necessarily very intense or long-lasting:  $w = 0$ ,  $p = 50$ , and  $q = 90$
4. Spatially extensive and long-lasting extreme events (SLE):  $w = 5$ ,  $p = 90$ , and  $q = 90$

We note that these four definitions are deliberately not disjoint, i.e., given events may be part of two or more of these classes. The reason for this is that for a given event type, we are interested in the characteristics of all events of that type, in contrast to the characteristics of events which only fulfill the conditions of one event type, but none of the conditions of the remaining three event types.

#### 6.4.2. Network construction

We compute the directed synchronization matrix  $\mathbf{ES}^{dir}$  (equation (2.7) in chapter 2) for the four events types described above using a maximal delay  $\tau_{max} = 8$  time steps of three hours, corresponding to one day. Since the event rates at each location are

typically different, we determine statistical significance and thus the condition for placing a network link in a similar way as in chapter 5. However, the specific null hypotheses differs slightly from the one assumed there: Due to the high temporal resolution of three hours, many events will occur subsequently<sup>4</sup>. We therefore construct independent surrogates which preserve the event rate as before, but now they also preserve this block structure of subsequent events. From each original time series (48400 in total), we thus construct surrogate time series by uniformly randomly distributing original blocks of subsequent events. On the basis of this null model, we proceed to determine statistical significance in the same way as described in chapter 5 (cf. equation (5.1)), but in this case for directed network links. As explained in chapter 3, a non-zero entry  $A_{ij}$  of the non-symmetric adjacency matrix represents a network link pointing from node  $j$  to node  $i$ .

### 6.4.3. Application of Complex Networks

We will use the directed networks constructed in this way to determine the geographic origins of extreme rainfall in a given region of interest  $R$ . This methodological framework is well-suited for this task, since it is designed to reveal possibly non-linear climatic linkages on the basis of event time series.

We modify the regional connectivity  $RC$  of  $R$  to the directed case by counting for each grid point  $i$  the number of links pointing *into*  $R$ :

$$RC_i(R) = \sum_{j \in R} A_{ji} \tag{6.5}$$

This measure yields a value for each grid cell  $i$ , indicating the number of grid cells inside  $R$  where extreme events occurred synchronously after they occurred at  $i$ . In order to be able to visualize  $RC$  for different regions on the same spatial map, we will in the following use a binary version of this measure: A grid point  $i$  will be considered to have strong linkages to a region  $R$  if the number of network links pointing from  $i$  to  $R$  is more than ten times larger than what should be expected from a uniformly random placement of all available links.

By indicating where rainfall events typically occur *before* they synchronously occur in a region under consideration,  $RC$  estimates the geographic *origins* of rainfall events in that region. These directed climatic linkages are helpful for understanding the meteorological processes causing rainfall events, but also provide essential information for assessing the predicability of extreme events in any region  $R$ .

In section 6.5.2, we will investigate these climatic linkages separately for the four different event types proposed above.

---

<sup>4</sup>Also refer to the corresponding discussion before equation (2.6) in section 2.2.

## 6.5. Results

### 6.5.1. Frequency, intensity, and spatial extent of rainfall events

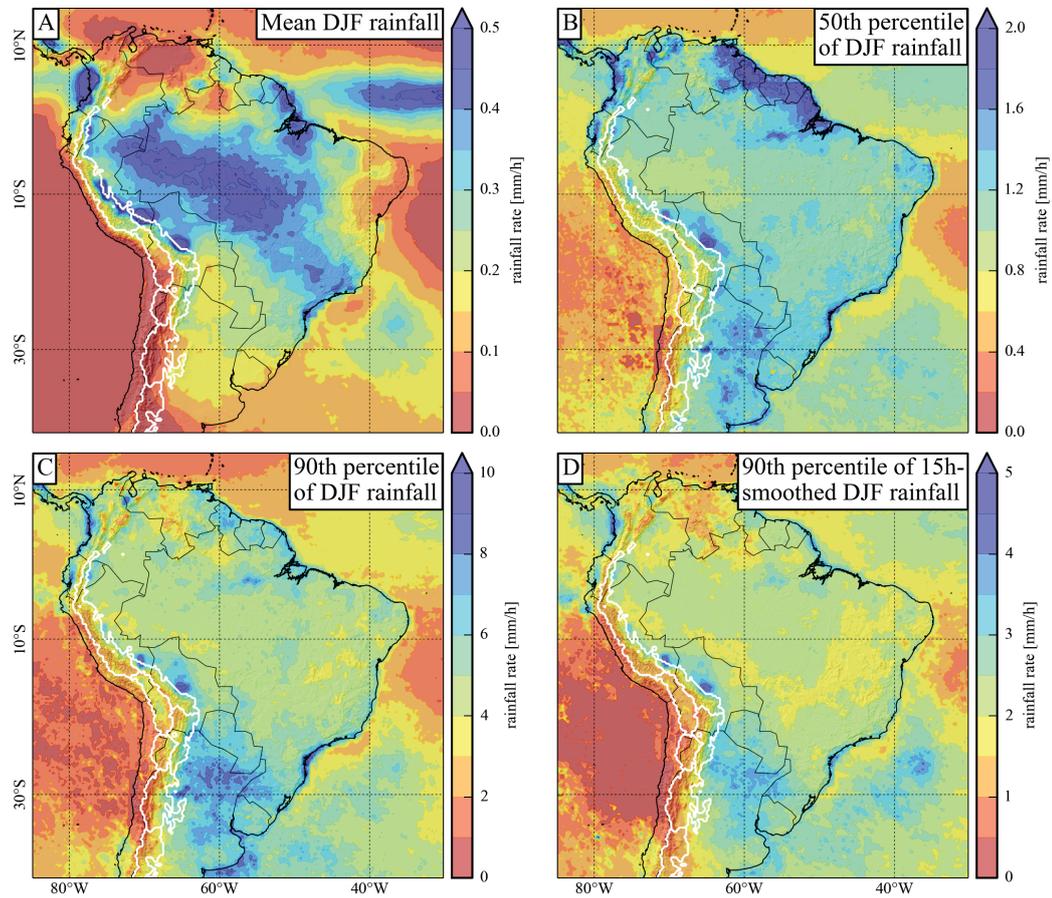
**Rainfall Intensity.** The monsoon season in South America is characterized by high mean rainfall values in the vicinity of the Intertropical Convergence Zone (ITCZ), in the Amazon Basin and along the South Atlantic Convergence Zone (SACZ), as well as along the eastern slopes of the Peruvian and Bolivian Andes (Figure 6.2A). The 50th percentile scores of wet times (hourly rainfall larger than 0.2mm/h) shows highest values at the northeastern South American coast, the eastern slopes of the Bolivian Andes as well as in northern Argentina (Figure 6.2B). 90th percentile scores of 3-hourly wet times exhibit highest values at the eastern slopes of the Bolivian Andes as well as in northeastern Argentina (Figure 6.2C). While the spatial distribution of 90th percentile scores for 3-hourly wet times of the 15-hour smoothed time series (Figure 6.2D) resembles that obtained for the non-smoothed case, the assumed values are considerably lower due to the temporal smoothing.

**Frequencies of the four event types.** In the previous section, we have defined four types of extreme events (which can be considered *end member scenarios*) that we will analyze in more detail: LSE, LLE, SEE, and SLE. Each event type occurs with different frequencies, although the relative spatial distribution of these frequencies remains similar (Figure 6.3). We emphasize that by *frequency*, we refer to the number of 3-hourly time steps which have an event, implying that, e.g., 4 consecutive time steps fulfilling the respective event condition are counted as 4 events.

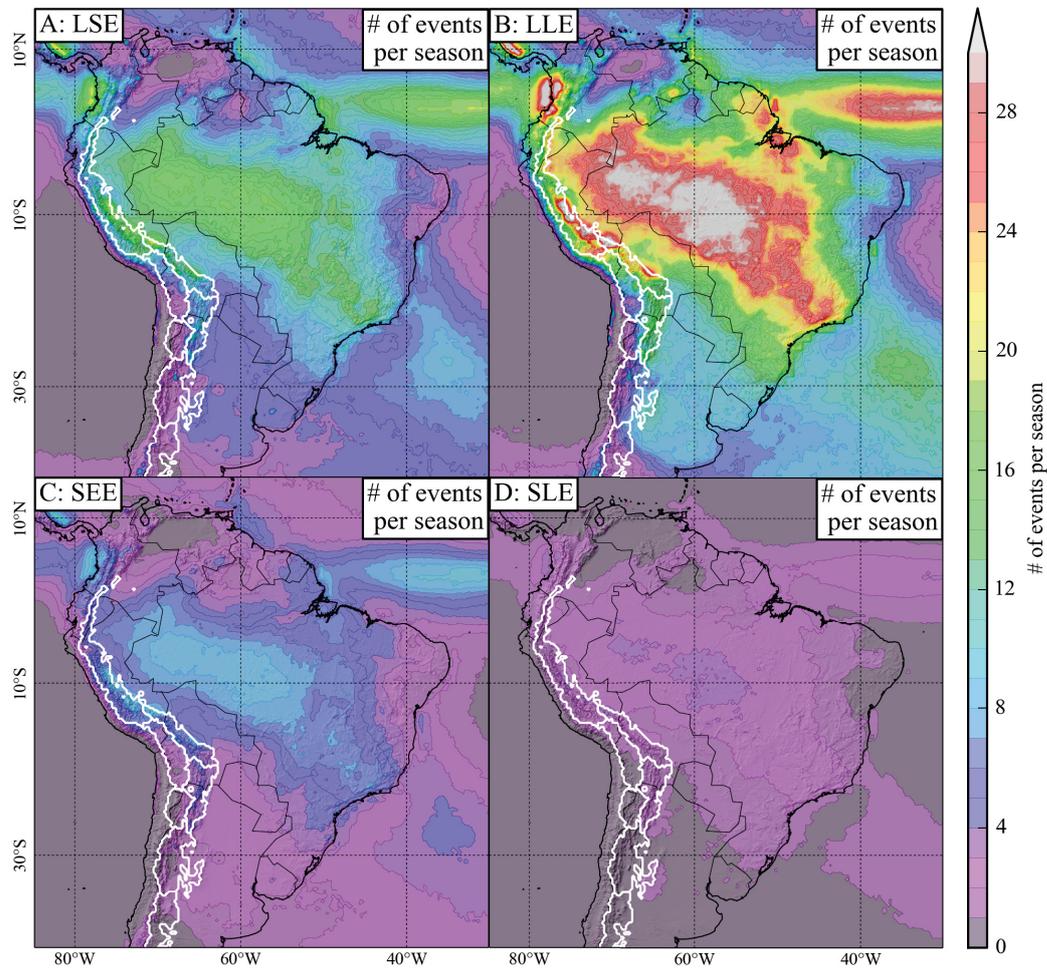
1. For LSE, we obtain typical values between 12 and 18 events per DJF season in the Amazon Basin and at the eastern Andean slopes north of 20°S, but only around 4 to 6 events in the subtropical plains of southern Bolivia, Paraguay, northern Argentina, and the adjacent Andean slopes (Figure 6.3A).
2. For LLE, these numbers are 25 to 30 and 7 to 12 respectively (Figure 6.3B).
3. For SEE, there are between 7 and 10 events in the Amazon Basin at the slopes of the northern Central Andes and between 1 and 4 events in the subtropical plains (Figure 6.3C).
4. For the most extreme SLE, we find 1 to 4 events in the Amazon Basin and the adjacent foothills of the Andes, but less than one such event per season in the subtropics of northern Argentina and Uruguay (Figure 6.3D).

Most notably, for all four event types, we observe a pronounced latitudinal gradient of event frequencies between the tropics and the subtropics: the northern, tropical regions have high event frequencies, while the subtropics exhibit considerably lower frequencies.

**Contributions to total seasonal rainfall.** The fraction of total DJF rainfall sums (not confined to wet times) accounted for by the four event types varies considerably:



**Figure 6.2.:** A. Mean hourly rainfall for DJF from 1998 to 2012. B. 50th percentile of rainfall distribution confined to wet times (i.e., 3-hourly time steps with average hourly rainfall larger than 0.2mm/h). C. 90th percentile of rainfall distribution confined to wet times. D. 90th percentile of 15-hour smoothed (using a 5-point moving average) rainfall distribution confined to wet times.



**Figure 6.3.:** Number of events per DJF season for local and short extreme events (A: LSE), local and long-lasting extreme events (B: LLE), spatially extensive events (C: SEE), and spatially extensive long-lasting extreme events (D: SLE).

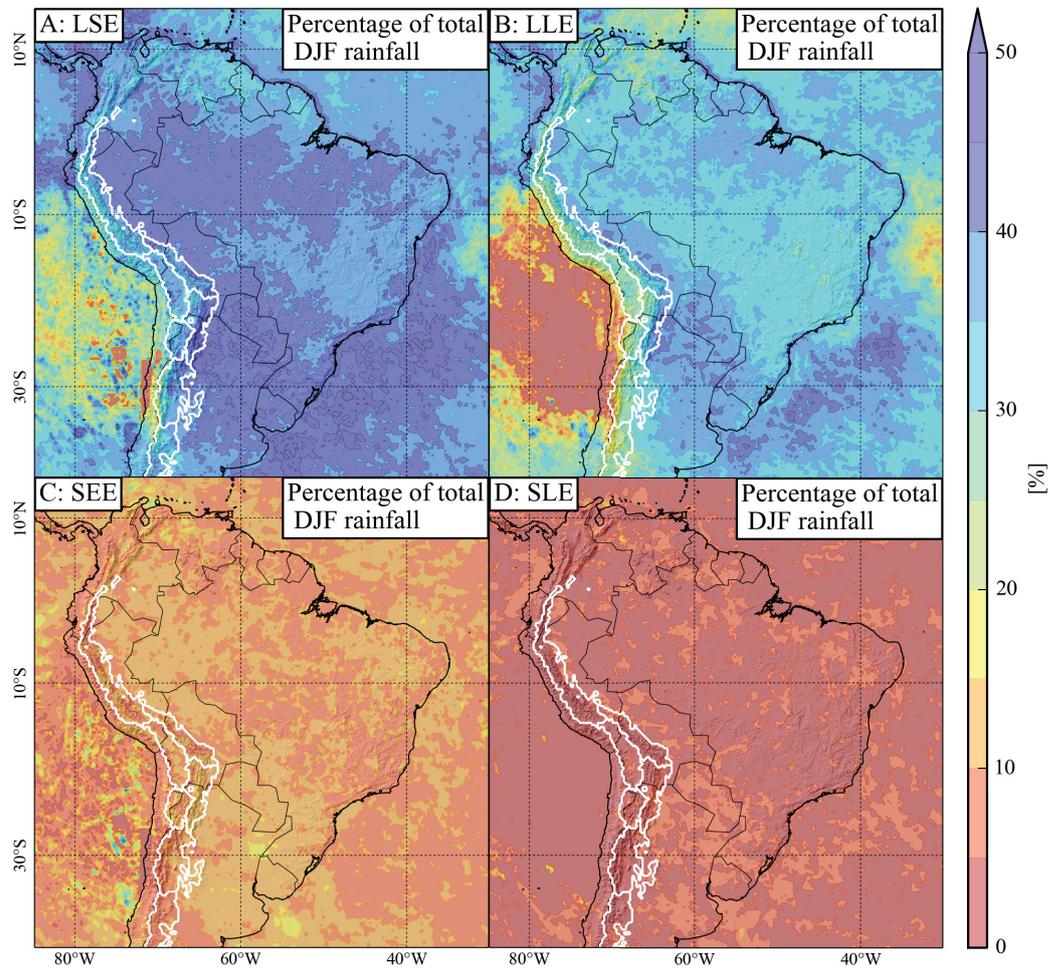
1. LSE events account for more than 40% of total seasonal rainfall in large parts of tropical South America, and even more than 50% in the some areas in the subtropics (Figure 6.4A).
2. The spatial distribution of total rainfall fractions for LLE is similar to the one obtained for LSE, but local values are about 10% lower (Figure 6.4B).
3. In contrast, SEE account for much lower fractions of total rainfall, with values ranging between 5% and 10% in most of Brazil, and between 10% and 20% in the subtropical plains (Figure 6.4C).
4. The lowest contributions were found for SLE, with fractions reaching 10% only in some parts of northern Argentina and Paraguay (Figure 6.4D).

Consecutive time steps with events will in the following be referred to as *bursts*. We emphasize that typical time scales of duration of the four events types (i.e., the number of events per bursts) vary by definition: bursts of LSE and SEE usually last 3 – 6 hours (roughly 1 to 2 events), while for LLE we find typical duration periods of 9 – 15 hours (3 – 5 events), and SLE last 6 – 12 hours (2 – 4 events) (see Figure C.1 in appendix C). In order to adjust for the varying event frequencies and burst durations when considering fractions of accounted DJF rainfall, we compute the average fraction of total DJF rainfall contributed by a single burst (Figure 6.5).

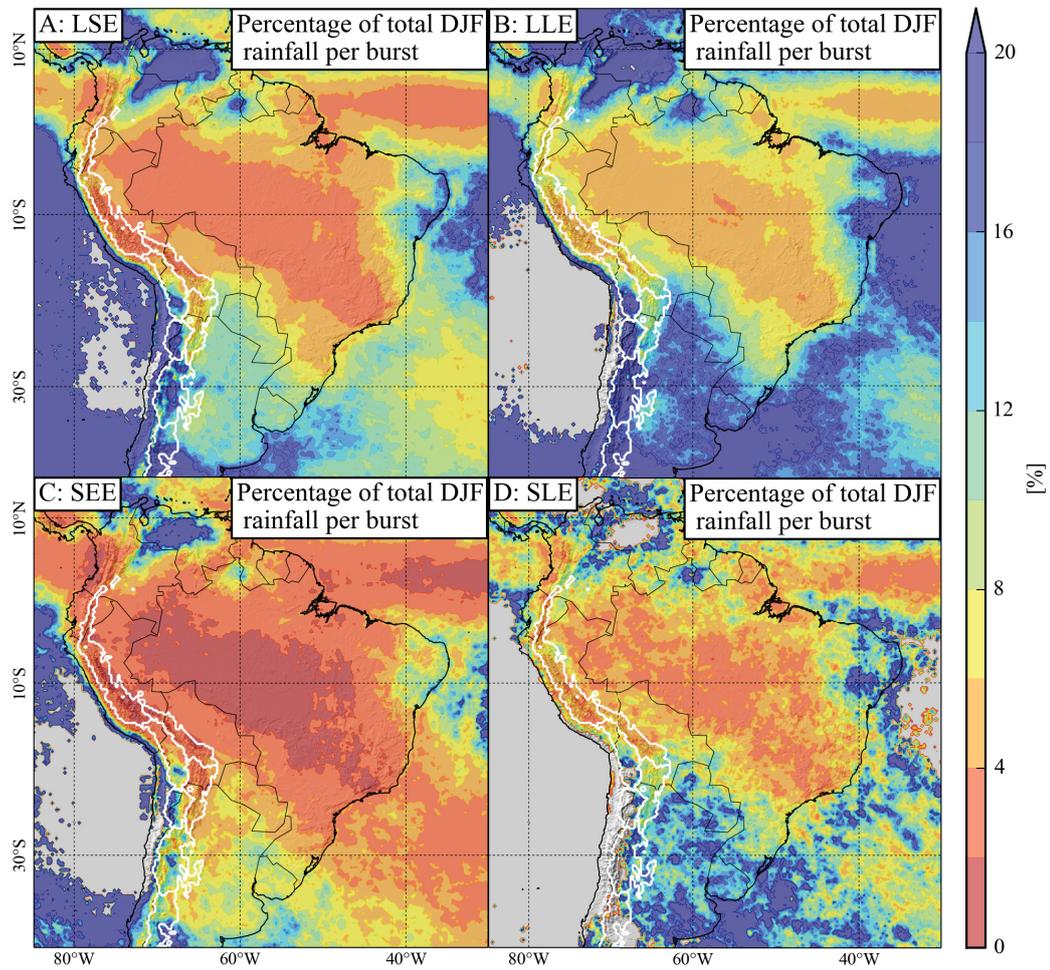
1. We find that each LSE burst accounts on average for 3% – 4% of total DJF rainfall in the Amazon Basin and the eastern slopes of the northern and northern Central Andes. In contrast, up to 10% of total DJF rainfall is accounted for by each burst in the subtropical plains (Figure 6.5A).
2. LLE bursts contribute more to total DJF rainfall, with corresponding fractions between 4% and 8% in the tropics and up to 20% in the subtropics (Figure 6.5B).
3. For SEE, we find contributions below 2% in the Central Amazon Basin and between 6% and 10% in northern Argentina (Figure 6.5C).
4. Finally, SLE contribute 2% to 6% to total DJF rainfall in the tropics, and up to 20% in the subtropics (Figure 6.5D).

We note that, by dividing these values by the typical numbers of events per burst, the fraction accounted for by each single *event* can be obtained. For an integration of these fractions with respect to the mountainous catchments C1 to C7 we refer to Figure C.1 in appendix C.

**Spatial extensions.** The two event types SEE and SLE involve a minimum spatial size of connected components of simultaneous events (their minimum "rainfall cluster size",  $S_i^q$ ). We show this size in units of  $km^2$ , which we obtain by computing  $(0.25^\circ \times 111km)^2 \times S_i^q \times \cos(\lambda)$ , where  $\lambda \in [-40^\circ, 15^\circ]$  denotes the latitudinal angle. We emphasize that the conversion from the number of grid cells to  $km^2$  is only approximately true, since the spatial distribution of the component size thresholds



**Figure 6.4.:** Percentage of total DJF rainfall contributed by local and short extreme events (A: LSE), local and long-lasting extreme events (B: LLE), spatially extensive events (C: SEE), and spatially extensive long-lasting extreme events (D: SLE).



**Figure 6.5.:** Average percentage of total DJF rainfall contributed by each single burst of consecutive events of type local and short (A: LSE), local and long-lasting (B: LLE), spatially extensive (C: SEE), and spatially extensive long-lasting (D: SLE).

is biased by the dependence of the grid cell size on the latitudinal position. We correct for this effect by weighting each threshold value with  $\cos(\lambda)$ . The component size threshold for SEE (Figure 6.6A) shows highest values north of the ITCZ over the tropical Atlantic Ocean and east of the Brazilian coast south of  $10^\circ\text{S}$ . Over southeastern South America values are smaller than in the latter regions, but still higher than over the remaining part of the continent. In contrast, SLE events (Figure 6.6B) exhibit highest component size thresholds over southeastern South America, centered over Uruguay.

### 6.5.2. Regional Connectivity of Andean Catchments

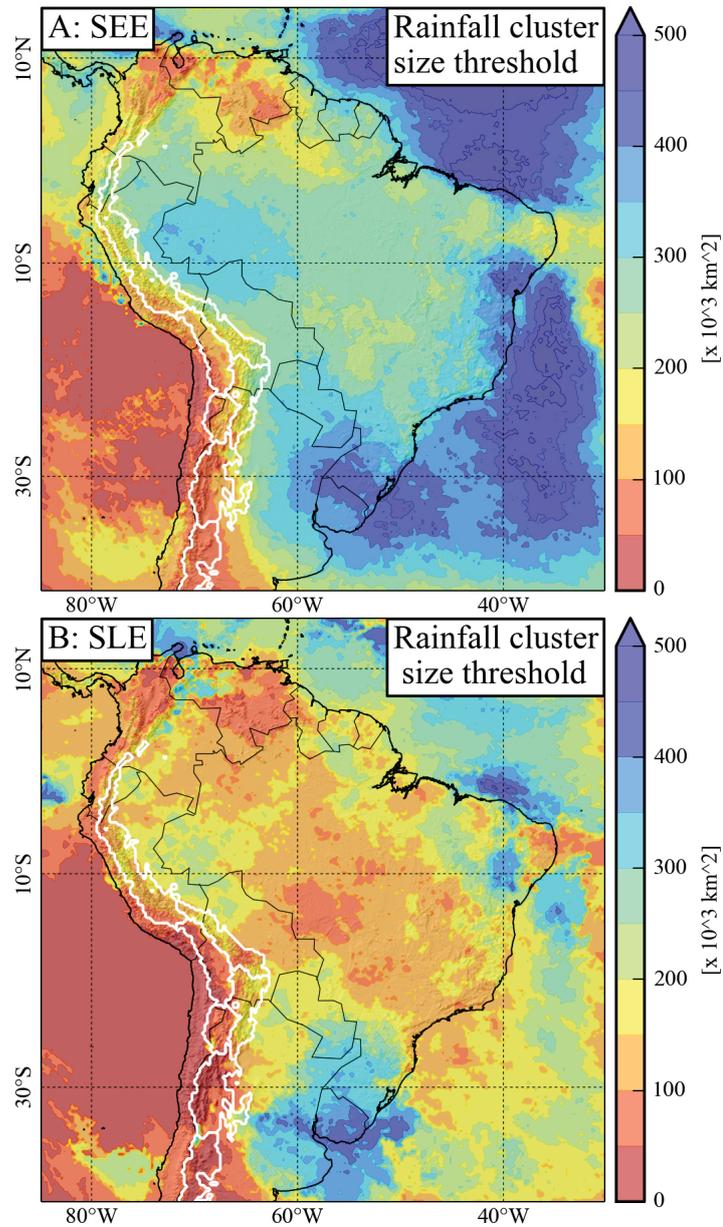
In the second step of this chapter, we investigate the origins of extreme rainfall on the Altiplano-Puna Plateau (C<sub>1</sub> and C<sub>2</sub>, Figure 6.7) in Bolivia and northern Argentina as well as at the eastern foothills of the entire Andean cordillera, which we resolved with respect to the various drainage basins located along the mountain range (C<sub>3</sub> to C<sub>7</sub>, Figure 6.8).

#### Altiplano-Puna Plateau (C<sub>1</sub> and C<sub>2</sub>)

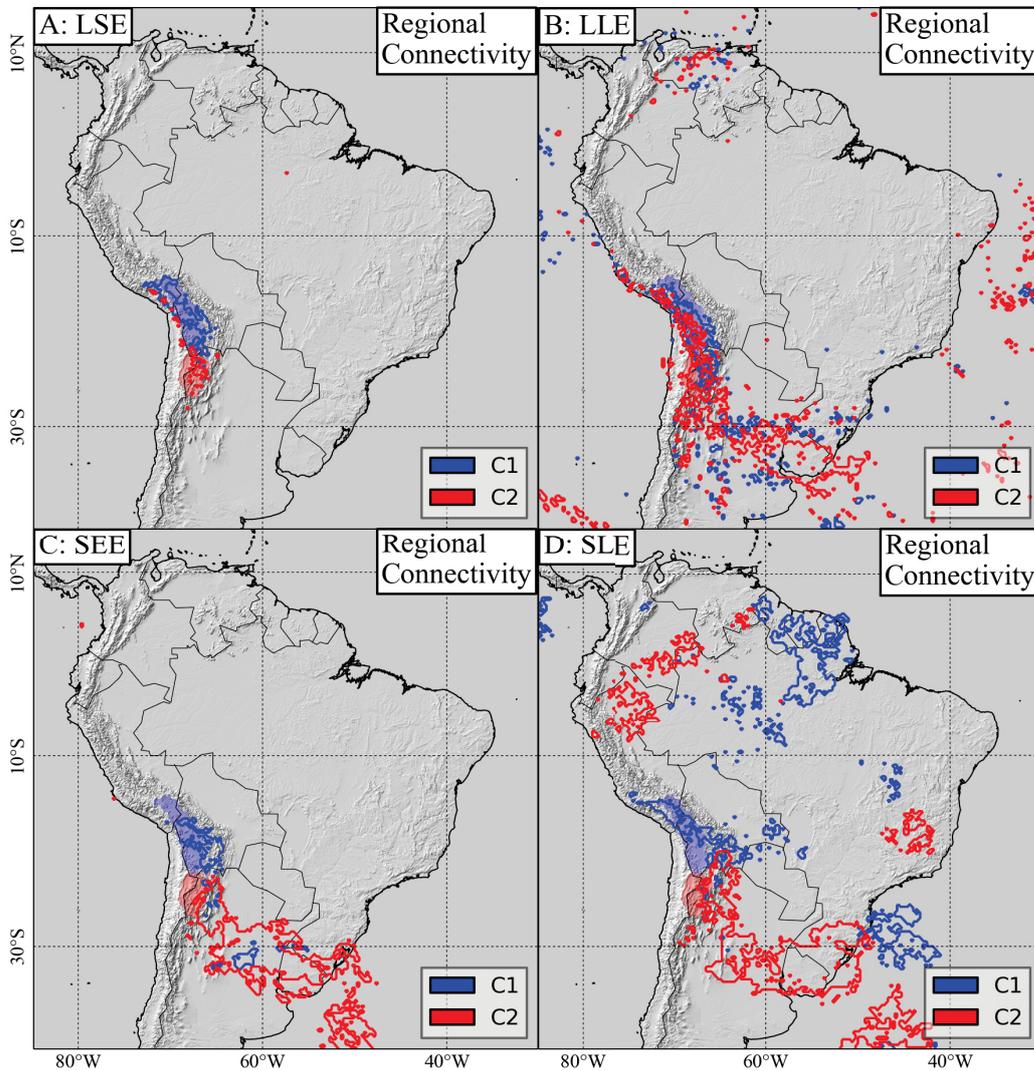
1. For LSE, we observe almost no linkages from other regions of South America for both C<sub>1</sub> and C<sub>2</sub> (Figure 6.7A).
2. For LLE, we observe few linkages from some scattered locations in northern Argentinean plains to C<sub>1</sub>. In contrast, we observe a large connected area over Uruguay and northern Argentina, which exhibits strong linkages into C<sub>2</sub>. Moreover, we observe linkages from C<sub>1</sub> to C<sub>2</sub> and vice versa (Figure 6.7B).
3. For SEE, there is only a small connected area in northeastern Argentina which is linked to C<sub>1</sub>, while a large connected area extending from Uruguay to the slopes of the northern Argentinean Andes exhibits linkages into C<sub>2</sub> (Figure 6.7C).
4. For SLE, we observe two large regions which have linkages into C<sub>1</sub>: The mouth of the Amazon in northeastern Brazil, as well as a region east of the southern Brazilian coast. In addition to the linkages from southeastern South America to C<sub>2</sub> which have already been found for LLE and SEE, for SLE there is also a region in northern Peru and southern Colombia which is strongly linked to C<sub>2</sub> (Figure 6.7D).

#### Andean foothills (C<sub>2</sub> to C<sub>7</sub>)

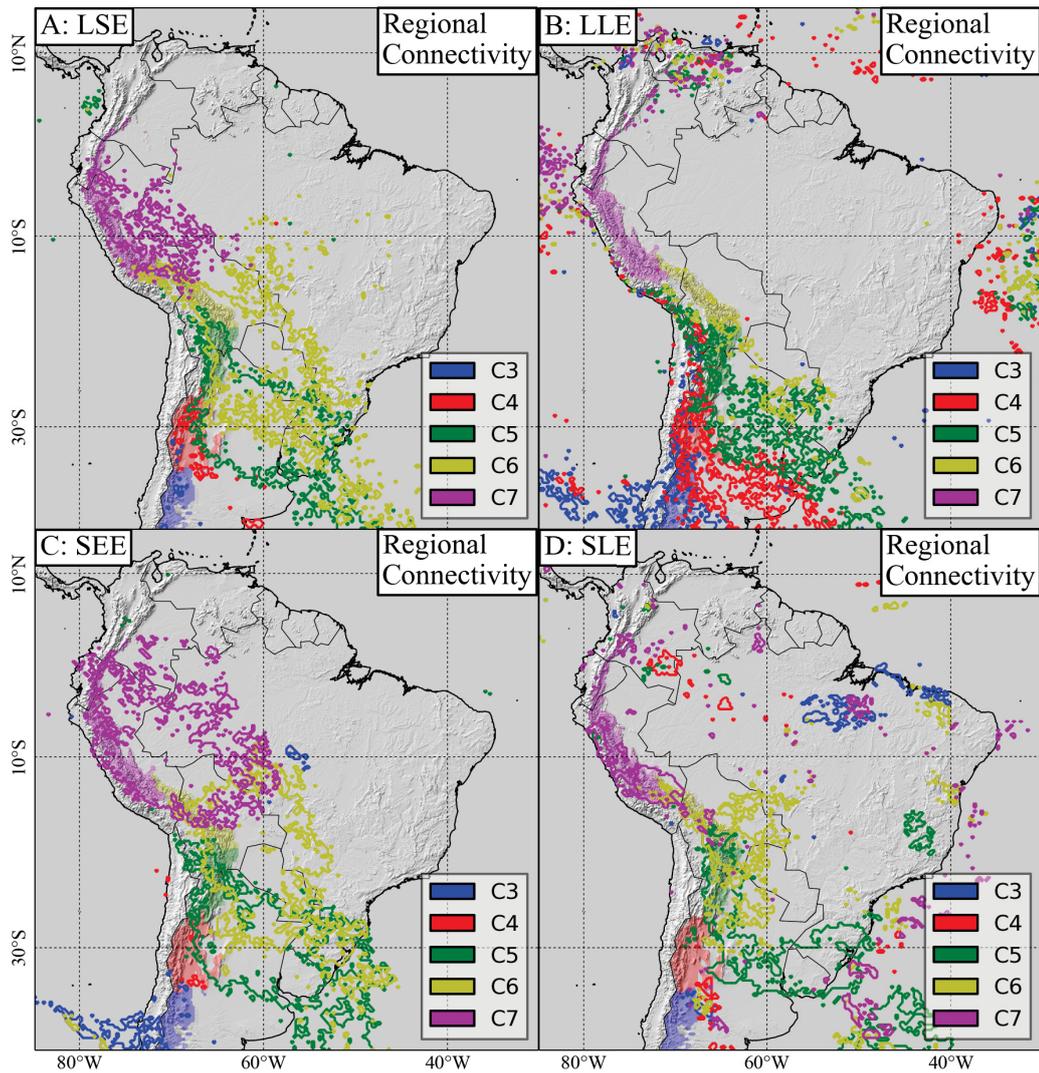
1. For LSE (Figure 6.8A), we do not find any significant linkages into C<sub>3</sub> and C<sub>4</sub>. C<sub>5</sub> shows strong linkages from a large connected area in northern Argentina, Uruguay and southern Brazil, while for C<sub>6</sub>, this area is shifted northward to include Paraguay, southern Bolivia and some parts of the Bolivian part of the Amazon Basin. For C<sub>7</sub>, we observe strong linkages from the western Amazon Basin in Peru and western Brazil.



**Figure 6.6.:** A. 90th percentiles of the sizes of connected components of simultaneous events ("cluster sizes") above the 50th percentile of local rainfall distributions confined to wet times (SEE). The size is measured in  $\text{km}^2$  and we corrected for the bias induced by the dependence of grid cell size on latitude. B. 90th percentile of the size of connected components of simultaneous events above the 90th percentile of the size of local 15-hour smoothed rainfall distribution confined to wet times (SLE).



**Figure 6.7.:** Regional Connectivity of the Altiplano (C1 in blue) and Puna Plateau (C2 in red) for local and short extreme events (A: LSE), local and long-lasting extreme events (B: LLE), spatially extensive events (C: SEE), and spatially extensive long-lasting extreme events (D: SLE). Spatially extensive, long-lasting extreme events (SLE) on the northern Altiplano Plateau (blue) are primarily connected to the Amazon Basin, while the southern Puna Plateau (red) is more strongly connected to the Argentinean plains and La Plata river basin.



**Figure 6.8.:** Regional Connectivity of the five mountainous catchments along the eastern Andean slopes (C3 to C7) for local and short extreme events (A: LSE), local and long-lasting extreme events (B: LLE), spatially extensive events (C: SEE), and spatially extensive long-lasting extreme events (D: SLE). In general, extreme rainfall in the northern Andean catchments (purple) are mostly locally connected, only SEE are influenced by the western Amazon basin. Tropical northern central Andean catchments (yellow) are impacted by rainfall events originating from southeastern South America and the southwestern Amazon Basin; Subtropical southern central Andean catchments (green) are dominated by rainfall originating from southeastern South America and propagating westward through the La Plata basin. Subtropical catchments to the south of the central Andes (red and blue) show only connections to extratropical regions.

2. For LLE (Figure 6.8B), there are strong linkages from central Chile and the southern Pacific ocean south of 30°S into C<sub>3</sub>, while a large connected area in central Argentina shows strong linkages into C<sub>4</sub>. C<sub>5</sub> has strong linkages from northern Argentina and Uruguay, while for C<sub>6</sub> and C<sub>7</sub>, we do not find significant linkages for LLE.
3. For SEE (Figure 6.8C), C<sub>3</sub> again exhibits linkages from central Chile and the adjacent Pacific Ocean, while C<sub>4</sub> shows no significant linkages. In contrast, there are strong linkages from a large area in northern Argentina and Uruguay to C<sub>5</sub>. Furthermore, C<sub>6</sub> shows strong linkages from a large area in Central South America, roughly extending between 10°S and 30°S and 75°W and 65°W. Strong linkages can be observed from the western Amazon Basin to C<sub>7</sub>.
4. For SLE (Figure 6.8D), there are no significant linkages into C<sub>3</sub> and C<sub>4</sub>. A large area comprised of southern Brazil, Uruguay, and northern Argentina is linked to C<sub>5</sub>, while a considerably smaller region in northern Argentina, western Paraguay, and Bolivia exhibits linkages into C<sub>6</sub>. We do not observe a large connected area with significant linkages into C<sub>7</sub>.

## 6.6. Discussion

### 6.6.1. Intensity, frequency, and spatial extent of rainfall events

By construction, the four proposed rainfall event types LSE, LLE, SEE, and SLE occur with varying intensities (Figure 6.2), frequencies (Figure 6.3), and spatial extents (Figures 6.6A and 6.5B). Therefore, they also play very different roles for the hydrological budgets of the different mountainous Andean catchments C<sub>1</sub> to C<sub>7</sub>. Our results indicate that LSE contribute the largest fraction of total DJF rainfall in the entire considered spatial domain, with percentages of up to 50% in the La Plata Basin. However, in view of the induced risk of natural hazards, it is crucial to consider the temporal duration of rainfall. For this purpose, we quantified the fraction of total DJF rainfall contributed by each single sequence of consecutive events (“bursts”). For all four event types (Figure 6.5), this revealed a pronounced latitudinal gradient, with lower contributions in the tropics, and higher contributions in the sub- and extratropics. Thus, there is a tendency towards low-frequency but high-magnitude events in subtropical South America. LLE and SLE contribute the largest fractions of total DJF rainfall per burst, with values close to 20% in northern Argentina, Uruguay, and Paraguay (Figures 6.5B and 6.5D).

Many studies have analyzed the spatial extensions of spatially connected components of simultaneous rainfall, and South America is affected by some of the largest of such rainfall clusters on Earth (Laing and Fritsch, 1997; Zipser et al., 2006). In particular, mesoscale convective complexes have been found to play a major role for total DJF rainfall in southeastern South America (Salio et al., 2007; Durkee and Mote, 2009), but are also important and have potentially disastrous impacts in other parts of the continent (Zipser et al., 2006). Nevertheless, they occur most frequently and

attain their largest size in the subtropical regions of northern Argentina and Paraguay (Durkee et al., 2009), which is consistent with our results for the 90th percentile scores of rainfall cluster sizes (Figures 6.6A and 6.6B). The hydrological role of the spatial extents of rainfall clusters can - by definition - not be directly quantified in terms of local (i.e., per grid cell) contributions to total DJF rainfall, as is also evident from the comparably low values in (Figure 6.5C). Nevertheless, the spatial extents are very important for the risk assessment of associated flooding and landsliding (Marengo et al., 1998; Grimm and Tedeschi, 2009).

### 6.6.2. Regional Connectivity of Andean Catchments

In view of natural hazards such as flooding and landsliding, we expect severe impacts of heavy rainfall on the high-elevation Altiplano-Puna Plateau as well as in the main river catchments along the eastern foothills of the Andean cordillera.

**Altiplano-Puna Plateau.** On the Altiplano ( $C_1$ ), up to 10% of total DJF rainfall is contributed by each burst of events of type SLE, which occur less than once per DJF season. Substantial connectivity of rainfall on the Altiplano to other geographical regions is only observed for this event type. This indicates that - typically - only these largest, longest-lasting and most intense rainfall clusters propagate up to the northern part of the plateau on more than 4km elevation, while the remaining three types of events do not occur in a spatially connected manner, but are controlled rather locally. The most pronounced geographical source region for SLE on the Altiplano is located in northeastern Brazil and extends with lower connectivity westward to the Amazon Basin. We associate this pattern with Amazonian squall lines (Cohen et al., 1995) and suggest that the largest of these systems, upon crossing the Amazon Basin, reach the Andes and propagate up to the Altiplano (Figure 6.7).

As for the Altiplano, LSE in the Puna de Atacama typically occur locally, without strongly synchronized events at other locations. In strong contrast, rainfall of type LLE, SEE, and SLE in the Puna de Atacama ( $C_2$ ) is mainly influenced by systems originating from southeastern South America. We associate this propagation of extreme, long lasting and spatially extensive rainfall from southeastern South America to the southern part of the Altiplano-Puna Plateau with frontal systems approaching from the south (P5; P6; Siqueira and Machado, 2004), caused by Rossby-wave activity in polar latitudes (Hoskins and Ambrizzi, 1993). This propagation pattern is also related to a subclass of mesoscale convective systems, which propagate in opposite direction of the low-level moisture flow towards the southern Central Andes (Anabor et al., 2008). Favorable atmospheric conditions for these propagation patterns are probably related to so-called *cold surges* (P5; P6; Garreaud and Wallace, 1998; Garreaud, 2000a). It is remarkable that the frontal systems exhibit such a strong impact on large areas at elevations above 4km asl. We also note that SLE on the Puna de Atacama are - in addition to the source region in southeastern South America - connected to preceding events in northern Peru and Colombia (Figure 6.7).

**Andean foothills.** Events of type LSE and SLE at the eastern foothills of the Andes south of  $25^{\circ}\text{S}$  ( $C_3$  and  $C_4$ ) do not show significant linkages to other geographical regions. In contrast, LLE and SEE at the eastern slopes of the southern Andes ( $C_3$ ) are associated with events over the adjacent Pacific Ocean. Extreme rainfall at the western slopes of the southern Andes can be explained by the interplay of frontal systems migrating eastward over the southern Pacific and the Andean orography (Garreaud, 2009). Our results suggest that the frontal influence extends eastward beyond the Andean mountain range and causes long-lasting (LLE) and spatially extensive (SEE) rainfall events in these regions (Figure 6.8).

LLE at the southern Central Andean foothills ( $C_4$ ) originate from the Argentinean lowlands and the adjacent Atlantic Ocean to the east. As in the case of the Puna de Atacama, we explain this pattern with frontal systems approaching from the south and migrating northward over South America east of the Andes. The Central Andean slopes ( $C_5$ ) in northern Argentina and southern Bolivia are strongly impacted by these frontal systems and associated cold surges, which also influence the formation and propagation of mesoscale convective systems (P5; P6; Anabor et al., 2008).

For all four event types, we observe large connected geographical source regions over northern Argentina, Uruguay, and southern Brazil. This is consistent with results found in the next chapter on the propagation of large rainfall clusters from southeastern South America towards the Central Andes (P5; P6).

For the northern Central Andean foothills ( $C_6$ ), the influence of frontal systems is substantially reduced for LLE and SLE, while for LSE and SEE, we still observe large connected source regions in Bolivia, Paraguay, and northern Argentina. However, these source regions are shifted northward when compared to the source regions of the Central Andes and also extend towards the Bolivian and Brazilian Amazon Basin. We suggest that this pattern may in fact be comprised of different meteorological signals, including northward migrating frontal systems and southwestward propagating squall-lines originating from the tropics. The fact that there are no strong linkages for LLE and SLE suggests that those events at the northern Central Andean foothills which are caused by these frontal systems and squall-lines on average only attain temporal durations below 12 hours (the typical time scale of LLE).

In strong contrast to the catchments in the southern and central Andes, the northern Andean foothills ( $C_7$ , north of  $15^{\circ}$ ) do not show any linkage with frontal systems propagating over subtropical South America. Instead, we find that LSE and SEE in this area originate from the western Amazon Basin, while LLE and SLE are not linked to other geographical locations. This suggests that long-ranged linkages of rainfall at the northern Andean slopes only occur for short-lived events. In contrast, long-lasting rainfall events in this region occur at more local spatial scales.

## 6.7. Conclusion

We have analyzed the frequency, duration, spatial extent, and spatial synchronization structure of 3-hourly rainfall events during the South American monsoon season. For

this purpose, we have defined four different types of rainfall: (1) events which are characterized by their high intensity alone; (2) long-lasting and intense events; (3) spatially extensive events; and (4) intense, long-lasting and spatially extensive events. We have focussed on the contributions of these events to total seasonal rainfall sums, but also - in view of potential predictability - their geographical origins over the South American continent using directed Event Synchronization and constructing directed networks from it.

Our main findings in this chapter can be summarized as follows: i) The overall contribution to total monsoon-seasonal rainfall is highest for events of type (1). However, taking into account the temporal duration of events reveals that large river catchments in the subtropical Argentinean plains are exposed to rare, long-lasting episodes of intense and spatially extensive rainfall, each of them contributing up to 20% of total monsoon seasonal rainfall. ii) The high-elevation Altiplano Plateau in Bolivia and southern Peru is only reached by the largest, longest-lasting and most intense thunderstorms originating from the eastern Amazon Basin. iii) For extreme events in the various river catchments along the mountainous Andes, we observe a clear transition regarding their geographical origin: While in the catchments south of 20°S, including the Puna Plateau, extreme rainfall originates from frontal systems approaching from the central Argentinean plains, the catchments north of 20°S, including the Altiplano Plateau, are mainly affected by squall lines originating from the Amazon Basin.

# Chapter 7.

## Prediction of Extreme Floods in the Eastern Central Andes

### 7.1. Summary

In the previous chapter, we have shown how the directed network approach on the basis of  $ES^{dir}$  can in principle be used to assess the predictability of extreme rainfall by determining typical geographical source regions for these events. Changing climatic conditions have led to a significant increase in magnitude and frequency of extreme rainfall events in the Central Andes of South America. These events are spatially extensive and often result in substantial natural hazards for population, economy, and ecology. In this chapter, we develop a general framework to predict extreme events by introducing the concept of network divergence on directed networks derived from measuring directed Event Synchronization. Using this framework, we reveal a linkage between polar and tropical regimes as the mechanism responsible for extreme rainfall in the Central Andes: the interplay of northward migrating frontal systems and the Andean topography leads to the opening of a low-level wind channel from the western Amazon to the subtropics, which provides the moisture for abundant rainfall. The northward movement of the frontal systems displaces the wind channel, causing the associated rainfall cluster to migrate from southeastern South America to the eastern slopes of the Central Andes. On the basis of these insights, we propose a simple forecast rule, which we apply to real-time satellite-derived rainfall data and show that it predicts more than 60% (90% during El Niño conditions) of rainfall events above the 99th percentile in the Central Andes. This chapter is based on the associated publications P5 and P6, and the following sections will closely follow these publications. Supplementary figures for this chapter can be found in appendix D.

### 7.2. Introduction

Prediction of extreme rainfall events is a challenging task and rainfall occurrence in the eastern Central Andes (ECA) can only be understood in the broader context of the South American monsoon system (SAMS). As described in section 1.2, a constant feature of the core monsoon season in South America (December through February, DJF) is the transport of moist air by low-level trade winds from the tropical Atlantic Ocean to the Amazon Basin along the Intertropical Convergence Zone (ITCZ, (Vera

et al., 2006)). However, the strength and direction of the subsequent moisture flow to the subtropics is subject to considerable variability, as was already noted in chapter 5 in the context of the South American rainfall dipole: Possible exit regions range from central Argentina to southeastern Brazil. A pronounced southward component towards the ECA is associated with the South American Low-Level Jet (SALLJ, (Marengo et al., 2004; Vera et al., 2006)) and a southward extension thereof, the Chaco Jet (Salio et al., 2002). These circulation regimes, which are partly controlled by the Northwestern Argentinean and the Chaco Low (Salio et al., 2002; Seluchi and Saulo, 2003; Saulo et al., 2004), have been put into relation with increased precipitation in southeastern South America (SESA, (Salio et al., 2007)). Southward-directed anomalies of the large-scale moisture flow are, however, also associated with enhanced rainfall in the ECA due to orographic lifting: Increased moisture flux is forced to rise at the Andean mountain front and leads to pronounced orographic rainfall (Bookhagen and Strecker, 2008; Romatschke and Houze, 2013).

The cause of the circulation variability and the corresponding rainfall anomalies has not yet been identified in a way that sufficiently resolves the temporal order of events (Kiladis and Weickmann, 1992; Lenters and Cook, 1999; Liebmann et al., 2004; Arraut and Barbosa, 2009). Since this is crucial for predicting associated extreme rainfall events, an early warning system for extreme rainfall in the ECA has been lacking. These events lead to severe infrastructural damage with large societal and economic ramification: For instance in early 2007, natural hazards associated with intense rainfall events in the ECA affected more than 133.000 households and produced estimated costs of 443 Mio. USD (Programa de las Naciones Unidas para el Desarrollo (PNUD), 2011).

In this chapter, we provide all theoretical information necessary to forecast spatially extensive extreme rainfall at the the ECA. For this purpose, we introduce the concept of network divergence on directed and weighted networks constructed from  $ES^{dir}$ . This measure is designed to assess the predictability of extreme events in significantly interrelated time series by determining sinks and sources (of extreme events) on directed and weighted networks. We present and apply the method with emphasis on extreme rainfall, but the methodology is more general and can be applied to a wide class of problems, ranging from climatic extreme event series to neuronal activity or data from financial markets.

### 7.3. Data

As in the previous chapter, we employ the remote-sensing derived and gauge-calibrated rainfall data TRMM 3B42 V7 (Huffman et al., 2007) in the spatial domain  $85^{\circ}\text{W}$  to  $30^{\circ}\text{W}$  and  $40^{\circ}\text{S}$  to  $15^{\circ}\text{N}$ , at horizontal resolution of  $0.25^{\circ} \times 0.25^{\circ}$ , and 3-hourly temporal resolution for the time period from 1998 to 2012. To test our forecast rule, we then use the (near) real-time satellite product TRMM 3B42V7 RT (Huffman et al., 2007) with identical temporal and spatial resolutions for the time period from 2001 to 2013. Geopotential height and wind fields at 850mb as well as Outgoing

Longwave Radiation were obtained from NASA's Modern-Era Retrospective Analysis for Research and Applications (MERRA, (Rienecker et al., 2011)).

Extreme rainfall events are defined as times with rainfall above the 99th percentile of all DJF seasons, which results in 108 (94) 3-hourly events at each grid cell for the 15 (13)-year rainfall time series of the gauge-calibrated (real-time) version of TRMM 3B42 V7.

## 7.4. Methods

### 7.4.1. Network construction

We employ  $ES^{dir}$  as defined in equation (2.7) to compute the synchronicity of extreme events at different grid cells using a maximum delay  $\tau_{max} = 16$  time steps, corresponding to two days. Furthermore,  $ES^{dir}$  can be used to compute the average strength of synchronization of extreme rainfall between geographic regions such as SESA and ECA as a function of time. This will in the following allow us to identify times of enhanced synchronization of events in SESA and ECA, which we will use to determine the responsible atmospheric conditions and, thereby, to formulate a forecast rule for extreme rainfall in the ECA.

Since all event rates are by construction identical in this chapter, we can determine a single significance threshold for  $ES^{sym}$  for all pairs of event time series: We construct independent surrogates which preserve the event rate as well as the block structure of subsequent events. From each original time series (48400 in total), we construct surrogate time series by uniformly randomly distributing original blocks of subsequent events in the same way as in chapter 6. Then, we compute  $ES^{dir}$  between all randomized time series and, from the histogram of all these values, determine the 95%-significance level. A directed network link from some node  $j$  to another node  $i$  will thus be placed if the corresponding entry of  $\mathbf{ES}^{dir}$  is above this significance threshold. In addition, we assign the respective value of  $\mathbf{ES}^{dir}$  to the corresponding network link as weight:  $A_{ij} = ES_{ij}^{dir}$ . Such a link from node  $j$  to node  $i$  indicates that, typically, extreme events at  $j$  occur shortly before they synchronously occur at  $i$ .

The strength of synchronizations into (out of) a grid cell is the sum of weights of all links pointing to (from) this grid cell, and in order to spatially resolve the temporal order of extreme events we introduce the measure network divergence  $\Delta\mathcal{S}$  (cf. equation (3.17) and Figure 7.2A). We recall from chapter 3 the definition of this measure as the difference of in-strength  $\mathcal{S}^{in}$  and out-strength  $\mathcal{S}^{out}$  at each grid cell:

$$\Delta\mathcal{S}_i := \mathcal{S}_i^{in} - \mathcal{S}_i^{out} := \sum_{j=1}^N A_{ij} - \sum_{j=1}^N A_{ji} . \quad (7.1)$$

Positive values of  $\Delta\mathcal{S}$  indicate sinks of the network: extreme events in these time series are preceded by extreme events in other time series; negative values indicate sources: extreme events there are followed by extreme events in other time series.

In addition to network divergence, in section 3.4.2 we defined the strength into and out of a region  $R$  (Figure 7.2A and 7.2B):

$$\mathcal{S}_i^{in}(R) = \frac{1}{|R|} \sum_{j \in R} A_{ji} \quad (7.2)$$

and

$$\mathcal{S}_i^{out}(R) = \frac{1}{|R|} \sum_{j \in R} A_{ij} . \quad (7.3)$$

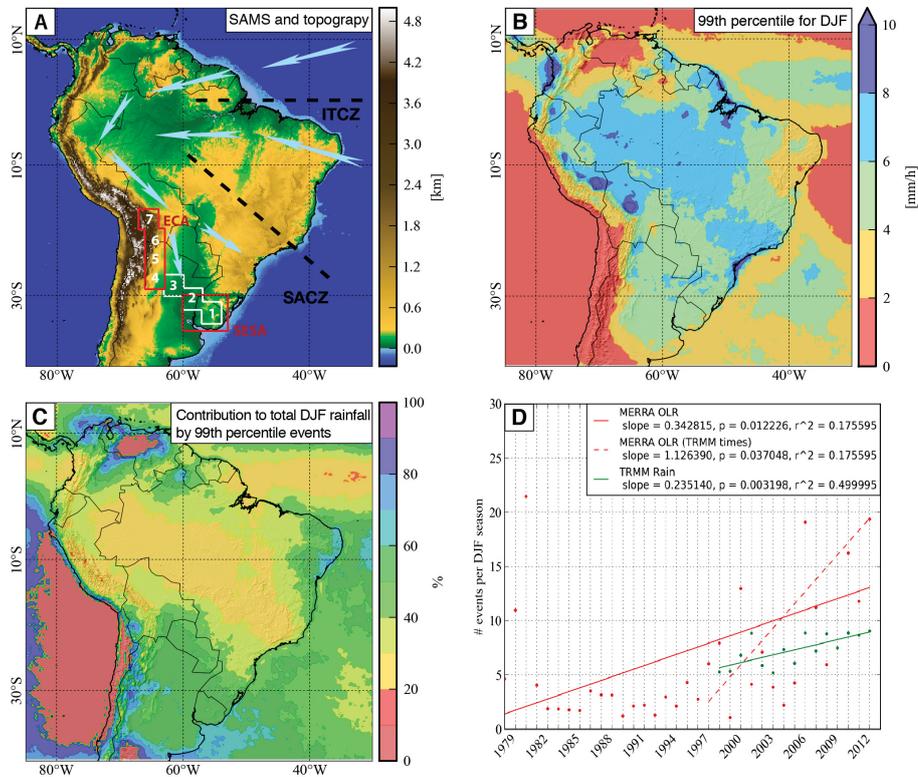
We recall from section 3.4.2 that the strength out of a region  $R$  to a node  $i$  is given by the average in-strength of  $i$  restricted to the links originating from  $R$  and vice versa. These measures will be used to determine where extremes go to from regions identified as source regions (negative values of  $\Delta\mathcal{S}$ ), or where extreme events in sink regions (positive values  $\Delta\mathcal{S}$ ) originate from.

## 7.5. Results

### 7.5.1. Climatic mechanism

During DJF, the spatial distribution of rainfall (see methods section 1 for data descriptions) is strongly influenced by the interplay of the southward shift of the ITCZ and the orographic barrier of the Andes (Figure 7.1A), leading to enhanced precipitation at the eastern Andean slopes, along the South Atlantic Convergence Zone (Carvalho et al., 2004), and in parts of SESA (Figure 7.1B). There exist strong spatial gradients in the amount of rainfall accounted for during events above the 99th percentile (Figure 7.1C). Most notably, very few extreme events (7 per season on average) account for more than 50% of total DJF rainfall in large parts of subtropical South America. We observe and corroborate earlier results (Marengo et al., 2009) that in the ECA, frequency as well as magnitudes of extreme events in DJF have increased substantially during the past decades (Figures 7.1D, D.1 to D.3).

In order to estimate the dynamics and temporal order of extreme rainfall in South America, we compute network divergence for the 3-hourly version of the satellite-derived and gauge-calibrated rainfall dataset TRMM 3B42 V7 (Figure 7.2A). The NW-to-SE stretching source regions over the Amazon Basin and over the equatorial Brazilian Atlantic coast can be attributed to Amazonian squall lines (P1; Cohen et al., 1995). Climatologically, the low-level flow from the Amazon towards the subtropics follows the band of sinks along the Bolivian Andes, which splits into two branches close to the Paraguayan border, corresponding to the SALLJ (Marengo et al., 2004) and the Chaco Jet (Salio et al., 2002), respectively. The most pronounced source region of the rainfall network is SESA, defined as the box ranging from  $35^\circ S$  to  $30^\circ S$  and  $60^\circ W$  to  $53^\circ W$  (Figure 7.1A). In order to investigate where synchronized extreme events occur within 2 days after extreme events occurred in SESA, we calculated the spatially averaged ES from SESA to each grid cell ( $\mathcal{S}^{out}(SESA)$ , Figure 7.2B) and,



**Figure 7.1.:** (A) Topography and simplified SAMS mechanisms. (B) 99th percentile of hourly rainfall during DJF derived from TRMM 3B42V7 (Huffman et al., 2007) in the spatial domain  $85^{\circ}\text{W}$  to  $30^{\circ}\text{W}$  and  $40^{\circ}\text{S}$  to  $15^{\circ}\text{N}$ , at horizontal resolution of  $0.25^{\circ} \times 0.25^{\circ}$ , and 3-hourly temporal resolution. (C) Fraction of total DJF rainfall accounted for by events above the 99th percentile. (D) Trend lines for the number of extreme events per DJF season averaged over boxes 6 and 7 in (A): for TRMM rainfall (108 events in total, green solid line) for the period from 1998 to 2012 and MERRA outgoing longwave radiation (OLR, (Rienecker et al., 2011)) for the period from 1979 to 2013 (252 events in total, red solid line) and for comparison for the period from 1998 to 2012 (red dashed line). Outgoing longwave radiation is used as a proxy for convective rainfall.

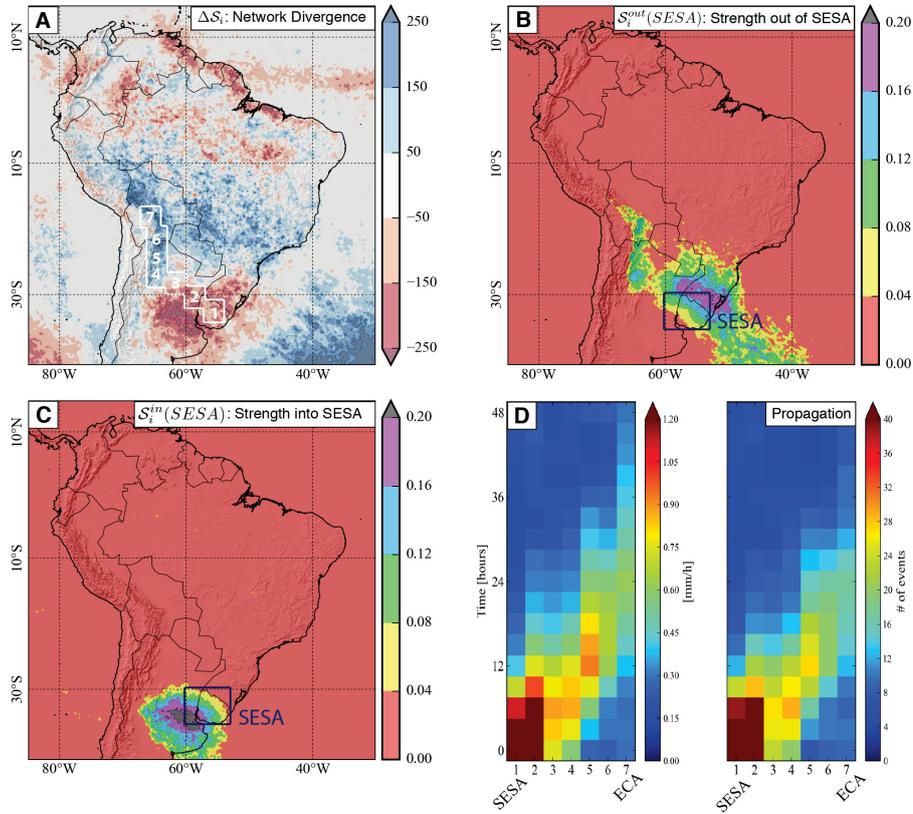
for comparison, from each grid cell to SESA ( $\mathcal{S}^{in}(SESA)$ , Figure 7.2C). This analysis reveals that extreme events in SESA are followed by extreme events along a narrow band following the eastern Andean slopes up to western Bolivia (Figure 7.2B), while they are only preceded by extreme events to the southwest (Figure 7.2C). These observations are consistent with results for  $\mathcal{S}^{in}(ECA)$ , showing that synchronized extreme events in the ECA occur within 2 days after they occurred in SESA (Figure D.4).

For certain atmospheric conditions, extreme rainfall in SESA is synchronized with extreme rainfall in the ECA within the subsequent 2 days. Since  $ES^{dir}$  can be used to identify times with high synchronization between these regions, we can determine the corresponding atmospheric conditions by constructing composites of geopotential height and wind fields for these times. We use the following framework to identify times of high synchronization between SESA and ECA: We refer to 3-hourly time steps for which at least 15 grid cells in SESA (corresponding to an area of  $\approx 11,000\text{km}^2$  or 2% of the SESA area as depicted, e.g., in Figure 7.1A) receive an extreme event as SESA times. This corresponds to time steps for which the number of extreme events at SESA is above the 60th percentile, computed on the set of time steps with at least one event. Furthermore, using the time series of synchronizations between SESA and ECA, we define SYNC times as time steps for which each grid cell in SESA receives an extreme event that synchronizes (within 2 days) with extreme events at more than 4 locations in the ECA. This corresponds to time steps for which the number of events at SESA that synchronize with one or more events at ECA is above the 80th percentile. Our results do not depend on small variations of the specific thresholds used to define SESA and SYNC times.

SESA times that are also SYNC times will be called propagation times, while SESA times that are not SYNC times will be referred to as non-propagation times (see Table 1). For the 15 DJF seasons considered here, we obtain 502 propagation times occurring during 136 connected storm periods of maximal length of 3 days (i.e., 9 per DJF season), while there are 582 non-propagation times during 164 storm periods. During propagation times, extreme events propagate along the sequence of a roughly SE-NW oriented swath profile (white boxes in Figure 7.1A and 7.2A) from SESA to ECA (Figure 7.2D), i.e., in the opposite direction of the low-level flow from the Amazon.

For the purpose of recognizing the conditions under which extreme events in SESA synchronize with extreme events in the ECA, we construct composite anomalies relative to DJF climatology of geopotential height and wind fields both at 850mb for propagation times and non-propagation times (Figure 7.3). Geopotential height and wind fields are derived from NASA's MERRA dataset (Rienecker et al., 2011).

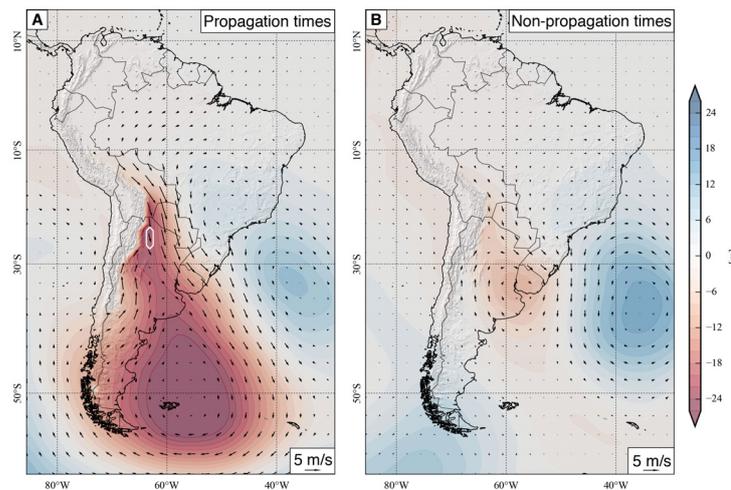
On the basis of these composites, we identify northward propagating frontal systems and the associated low-pressure anomalies as common drivers of extreme rainfall in SESA and the establishment of a low-level wind channel from the Amazon to the subtropics along the eastern Andean slopes: A low pressure anomaly originating from Rossby-wave activity propagates northwards, led by a cold front causing abundant rainfall in SESA through the uplifting of warmer air masses (Kiladis and Weickmann, 1992; Hoskins and Ambrizzi, 1993; Lenters and Cook, 1999; Seluchi and Garreaud,



**Figure 7.2.:** (A) Network divergence, defined as the difference of In-Strength and Out-Strength at each grid cell,  $\Delta S_i := S_i^{in} - S_i^{out}$ . Positive values indicate sinks of the directed and weighted network, which are interpreted as locations where synchronized extreme rainfall occurs within 2 days after it occurred at several other locations. On the other hand, negative values indicate sources, i.e. locations where synchronized rainfall occurs within 2 days before it occurs at several other locations. (B) Strength out of SESA,  $S_i^{out}(SESA)$ , which is the average in-strength restricted to SESA (see methods section). Note in particular the high values along ECA. (C) Strength into SESA,  $S_i^{in}(SESA)$ , which is the average out-strength restricted to SESA. Note in particular that there are no high values along ECA. (D) Temporal evolution of extreme rainfall events from SESA to ECA along the sequence of boxes indicated in (A). Composite rainfall amounts (left) and number of extreme events (right) in the respective boxes between SESA and ECA are displayed for propagation times and the subsequent 48 hours. Each box has an edge length of  $3^\circ$  ( $\approx 333\text{km}$ ), resulting in a total distance of  $\approx 2000\text{km}$ .

Times	Condition	Occurrences
SESA	# extreme events in $SESA \geq 60\text{th percentile}$	1084
SYNC	# synchronizations between $SESA \text{ and } ECA \geq 80\text{th percentile}$	518
Propagation	SESA & SYNC	502
Non-Propagation	SESA & NOT SYNC	582
Prediction	SESA & GPH	649

**Table 7.1.:** Different conditions used to determine the climatic mechanism and to formulate the forecast rule. GPH refers to the condition that the average geopotential height anomaly in the white polygon in Figure 7.3 is below  $-10\text{m}$ .



**Figure 7.3.:** (A) Composite anomalies relative to DJF climatology of 850mb geopotential height and wind fields from NASA's Modern-Era Retrospective Analysis for Research and Applications (MERRA, (Rienecker et al., 2011)) for propagation times. Temporal resolution is 3-hourly, spatial resolution is  $1.25^\circ \times 1.25^\circ$ . (B) The same composite anomalies as for (A), but for non-propagation times.

2006). When the frontal system propagates from SESA northeastward through the La Plata Basin in northeastern Argentina, the low-pressure anomaly extends to central Bolivia and merges with the Northwestern Argentinean Low (Seluchi and Saulo, 2003; Arraut and Barbosa, 2009) (Figure 7.3A). This leads to the opening of a geostrophic wind channel along the resulting isobars that was previously blocked by the Andes Cordillera. This channel acts as a conveyor belt and transports warm and moist air from the Amazon Basin along the eastern slopes of the Andes and collides with the cold air carried by the frontal system. In combination with orographic lifting effects, this leads to extreme rainfall in the ECA within 2 days of the initial rainfall in SESA. The enhanced moisture flow to SESA after the initiation of rainfall can be assumed to be further stabilized by the release of latent heat (Vera, 2002; Seluchi and Saulo, 2003). With the cold front moving north, the flow will change its direction accordingly. A comparison with (Anabor et al., 2008) suggests that this climatic regime may be associated with Mesoscale Convective Systems (Salio et al., 2007), which are formed over SESA and propagate upstream. Similar features have also been described in the context of so-called cold surges: northward incursions of cold air from midlatitudes (Garreaud and Wallace, 1998; Garreaud, 2000a).

### 7.5.2. Extreme event forecast

Typically, rainfall events propagate from SESA to the ECA within the first day after the initial event in SESA (Figure 7.2D), with an average speed of  $\approx 80\text{km/h}$ . These results can be used to establish an operational warning system of floods in the ECA. We employ the 3-hourly real-time satellite product TRMM 3B42V7 RT (Huffman et al., 2007) for the time period from 2001 to 2013. In order to forecast extreme rainfall events in the ECA, we define prediction times as SESA-times with a low-pressure anomaly in northwestern Argentina (geopotential height anomalies less than  $-10\text{m}$  in white polygon in Figure 7.3A; this condition is abbreviated as GPH in table 1). There are in total 649 such prediction times, occurring during 139 connected periods, resulting in an average of 10 such periods per season. The rainstorms associated with these events are likely to lead to severe floods and landslides downstream (Programa de las Naciones Unidas para el Desarrollo (PNUD), 2011; Bookhagen and Strecker, 2012) because of their large spatial extent combined with little to no rainfall infiltration at high elevations: During the 2 days following prediction times, about 1/4 of each of the 4 boxes comprising ECA (boxes 4 to 7 in Figure 7.1A) receives an extreme event, corresponding to about  $28,000\text{km}^2$  (Figure D.5). In particular, in the northern part of ECA (box 7 in Figure 7.1), extreme events propagate to high elevations: In the northernmost box 7, at altitudes higher than 3000m above sea level, still about 60% (80% during positive El Niño Southern Oscillation (ENSO) phases) of all extreme events occur during prediction times (Figures D.6 and D.7).

For the TRMM3B42V7 RT dataset, more than 60% of all extreme events and of total DJF rainfall occur in the ECA during the 48 hours following prediction times (Figures D.8 and D.9). During positive ENSO phases, they account for more than 90% of extreme rainfall events and more than 80% of total DJF rainfall in the

	observed	not observed	marginal
forecasted	a	b	a+b
not forecasted	c	d	c+d
marginal	a+c	b+d	n=a+b+c+d

**Table 7.2.:** Contingency table used for computing the Heidke-Skill-Score (HSS).

Value	all years	positive ENSO
a	318	98
b	331	104
c	292	18
d	8419	1220

**Table 7.3.:** The specific values of a, b, c, and d used to compute the HSS of the forecast rule.

northern parts of the ECA as well as on parts of the Bolivian Altiplano (Figures D.10 and D.11). In order to take into account the spatial extension of extreme rainfall, we formulate our forecast rule as follows: Whenever the conditions of prediction times are fulfilled, there will be at least 100 events above the 99th percentile during the following 2 days in at least one of the ECA boxes (white boxes 4 to 7 in Figure 7.2A). Note that the corresponding average number of extreme events within such two-day periods is 50.

### 7.5.3. Prediction Skill

In order to assess the skill of this simple forecast rule, we employ the Heidke-Skill-Score (HSS, (Wilks, 2006)). Given the separations between forecasted and observed events indicated in Table 7.5.3, it is defined as

$$HSS = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)} \quad (7.4)$$

for a skill-comparison versus randomness. This score yields  $HSS = 0$  for a uniformly random forecast, and  $HSS = 1$  for a perfect forecast. Applying our forecast rule to the 3-hourly forecast dataset (TRMM 3B42V7 RT), we find the values summarized in Table 7.5.3 for the time period 2001 to 2013.

For our forecast rule, we obtain  $HSS = 0.47$  when computed for all times during the DJF seasons between 2001 and 2013. We recall, however, that the considered climatic regime is only responsible for 60% of extreme events in the ECA. This implies that the remaining 40% can by construction not be predicted by our forecast rule, and the HSS is accordingly reduced. Moreover, the forecast skill certainly depends on the specific choice of the spatial boxes 4 to 7 and may change by adjusting their position. For positive ENSO conditions, we obtain  $HSS = 0.57$ . The HSS is rather insensitive to variations of the condition on the number of extreme events in SESA and the exact geopotential height anomaly in northwestern Argentina, while it decreases

rapidly for more events to be predicted in the ECA (Figures D.12 and D.13). We note that, while the mechanism responsible for these extreme rainfall events in the ECA was uncovered using network divergence, the conditions used for the forecast rule can be determined directly and with little computational efforts by spatially averaging rainfall and geopotential height data. We emphasize that we did not train the proposed forecast rule in the sense of parameter optimization. Instead, the rule is derived directly from the results of the network divergence analysis and we show that its forecast skill does not change rapidly when changing the conditions used to define prediction times.

## 7.6. Discussion

Our results provide all information necessary to implement an operational forecast system of extreme rainfall events in the ECA. It is very unlikely that previous state-of-the-art weather forecast models could predict these events: First, the propagation pattern only appears for very high event thresholds (97th percentile or higher, see Figure D.14), and this “heavy tail” of the rainfall distribution is not well implemented in current weather forecast models (see, e.g., (Jones et al., 2011) and citations therein). Second, for the regional climate model ETA, which is used at the Center for Weather Forecasting and Climate Research (CPTEC) for operational weather forecast in South America, we compared the synchronization strength of SESA to the pattern found for TRMM and conclude that this model does not reproduce the propagation of extreme events from SESA to ECA (Figure D.15). Furthermore, while the climatological phenomenon of cold surges has already been described in other studies (e.g. by Garreaud and Wallace, 1998; Garreaud, 2000a), only the usage of the high-spatiotemporal satellite product TRMM 3B42 allows to uncover the propagation of extreme events from SESA to ECA. This mechanism could not be found on the basis of reanalysis data such as the European Centre for Medium-Range Weather Forecasts Interim Reanalysis or NASA’s MERRA precipitation product (Figure D.15).

In appendix D, we show composites of rainfall, geopotential height and wind fields for times when extreme rainfall propagates from SESA to ECA (Figure D.16), from 12 hours before to 30 hours after rainfall peaks in SESA.

## 7.7. Conclusion

We have defined the measure network divergence on directed and weighted graphs, and have proposed a general framework to predict extreme events in large, interactive systems. Applying network divergence to high-spatiotemporal resolution rainfall data identified a climatic mechanism that allows to predict more than 60% (90% during positive ENSO conditions) of rainfall events above the 99th percentile in the ECA from two conditions: Preceding extreme rainfall at SESA and the presence of a low-pressure anomaly in northwestern Argentina. The forecast rule we developed

can thus be implemented with little computational efforts on the basis of spatially averaged rainfall and pressure measurements.

# Chapter 8.

## Dataset and Model Intercomparison

### 8.1. Summary

Based on the methodological and climatological insights gained in the previous chapters, we now compare six different rainfall datasets for South America with a focus on their representation of extreme rainfall during the monsoon season (December to February): The gauge-calibrated satellite product TRMM 3B42 V7, the (near) real-time version TRMM 3B42 V7 RT, the satellite-gauge combination product GPCP 1DD V1.2, the ECMWF model-derived reanalysis product ERA-interim, as well as output of a high spatial resolution run of the ECHAM6 global circulation model, and output of the regional climate model ETA. For the latter three, this comparison can also be understood as a model evaluation. In addition to statistical values of local rainfall distributions, we focus on the spatial characteristics of extreme rainfall co-variability using symmetric Event Synchronization ( $ES^{sym}$ ) and several of the network measures introduced in the previous chapters. This way, we uncover substantial differences in extreme rainfall patterns between the different datasets: i) The three model-derived datasets yield very different results compared to the satellite-gauge combinations regarding the main climatological propagation pathways of extreme events as well as the main convergence zones of the monsoon system. ii) Large discrepancies are found for the development of mesoscale convective systems in southeastern South America. iii) Both TRMM datasets and ECHAM6 indicate a linkage of extreme rainfall events between the central Amazon Basin and the eastern slopes of the central Andes, but this pattern is not reproduced by the remaining datasets. Our results suggest that none of the three model-derived datasets adequately capture extreme rainfall patterns in South America. This chapter is based on the associated publication P7, and some of the following sections will closely follow parts of this publication.

### 8.2. Introduction

Extreme rainfall plays a crucial role for the hydrological cycle in large parts of South America. In the subtropics, events above the 90th percentile account for more than 50% of total rainfall during the core monsoon season from December to February (see Figure 7.1 in the previous chapter, but also Figure 8.5 in this chapter). As shown in chapter 6, considerable contributions are made by Mesoscale Convective Systems

(MCS), in particular in northern Argentina, Paraguay, and southern Brazil (Zipser et al., 2006; Salio et al., 2007). But apart from their role for the overall water budget, understanding extreme rainfall events and their synchronization is important because of the associated natural hazards. For example, extreme rainfall frequently triggers landslides and flash floods, in particular along the eastern slopes of the Andes (see chapters 6 and 7, but also (P<sub>4</sub>; P<sub>5</sub>; P<sub>6</sub>; Schuster et al., 2002; Coppus and Imeson, 2002; O’Hare and Rivas, 2005; Moreiras, 2005b)), but as well in urban areas in southeastern South America (Marengo et al., 2013b).

The far-reaching impacts of extreme rainfall events call for a better understanding of how their frequency, magnitude, and spatial co-variability are represented by different datasets. Furthermore, in order to obtain a sound assessment of future extreme rainfall development, it is crucial to evaluate climate models with respect to their performance in reproducing observed spatiotemporal characteristics of extreme rainfall. The representation of extreme events in a dataset or model is usually only assessed by investigating local rainfall distributions and in particular the behavior of their respective tails. Because of their outstanding hydrological, but also societal relevance, a detailed investigation of the implementation of extreme events in different datasets and climate models, with particular focus on their spatiotemporal interrelations, is needed.

In this chapter, we compare the representation of extreme rainfall events between three gridded observational, as well as three gridded model-derived datasets for South America: the satellite-gauge combined TRMM 3B42 V7, the corresponding real-time product TRMM 3B42 V7 RT, which is not gauge adjusted, the gauge-satellite combined product GPCP, the model-derived reanalysis data from the European Centre for Medium-Range Weather Forecasts (ERA-interim), as well as high-spatial resolution data obtained from the global circulation model ECHAM6, and from the ETA regional climate model.

Several previous studies have analyzed rainfall variability in the South American monsoon system (SAMS), but only a few studies have compared different rainfall datasets in South America: Matsuyama et al. (2002) have analyzed the Climate Prediction Center merged analysis of precipitation, and more recently Negrón Juárez et al. (2009) have compared satellite and gauge products over the Amazon Basin, while Silva et al. (2011) evaluated different reanalysis products over the entire continent. Carvalho et al. (2012) provide a comparison of station based, satellite derived, and reanalysis data with a focus on daily gridded precipitation (see also references therein). A specific analysis of the implementation of rainfall processes over tropical South America in CMIP<sub>5</sub> was recently carried out by Yin et al. (2013). Typically, these studies use principal component analysis (PCA) in order to investigate the spatial characteristics of rainfall co-variability in terms of empirical orthogonal functions (EOFs). However, for the reasons explained in the introductory chapter 1, PCA-based techniques are not suitable to analyze the spatial patterns of co-variability of extreme rainfall.

In chapters 4 and 5, as well as the associated publication P<sub>1</sub> and P<sub>2</sub>, the combination of  $ES^{sym}$  and undirected networks was applied to the TRMM 3B42 V7 dataset

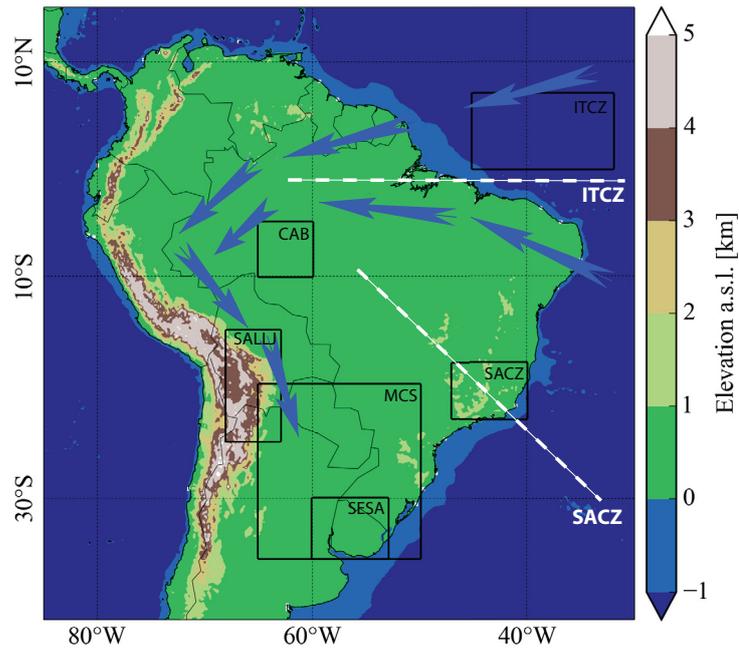
in order to investigate the spatial structure of synchronicity of extreme rainfall events in the SAMS. This dataset has been concluded to provide a reasonably good representation of rainfall variability in different regions (Xue et al., 2013; Chen et al., 2013), but particularly in South America (Carvalho et al., 2012; Zulkaffi et al., 2014). Here, we apply the same methodology and compare results for TRMM 3B42 V7 to corresponding results for the two other observational as well as for the three model-derived datasets. TRMM 3B42 V7 will thus serve as a reference dataset in this chapter.

### 8.3. Data

We employ six different daily rainfall datasets:

1. TRMM: the research grade Tropical Rainfall Measurement Mission gauge-calibrated satellite product (TRMM 3B42 V7) at 3-hourly temporal and  $0.25^\circ \times 0.25^\circ$  spatial resolutions (Huffman et al., 2007), available from 1998 to 2013.
2. TRMM RT: the (near) real-time Tropical Rainfall Measurement Mission satellite product (TRMM 3B42 V7 RT) at 3-hourly temporal and  $0.25^\circ \times 0.25^\circ$  spatial resolutions (Huffman et al., 2007), available from 2001 to 2013.
3. GPCP: the Global Precipitation Climatology Project (GPCP 1DD V1.2), a satellite-gauge combination at daily temporal and  $1.0^\circ \times 1.0^\circ$  spatial resolutions (Huffman et al., 2001), available from 1996 to 2013.
4. ERA: the European Centre for Medium-Range Weather Forecasts Interim Reanalysis (ERA-interim) product with daily temporal and  $0.75^\circ \times 0.75^\circ$  spatial resolutions, available from 1979 to 2013 (Dee et al., 2011).
5. ECHAM6: A standard AMIP simulation for the time period 1998 to 2008 (forced by the boundary conditions specified for CMIP5) carried out with the ECHAM6 model (Stevens et al., 2013) at 6-hourly temporal resolution and horizontal resolution of T255 (ca. 50 km) with 95 vertical levels. This simulation, which is carried out within the German consortium project STORM, is described in (“Effect of horizontal resolution on ECHAM6-AMIP performance”).
6. ETA: output from the regional climate model ETA (Mesinger et al., 2012) provided by the Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) driven by ERA-interim. The original temporal resolution of this model run is 6-hourly, and native horizontal resolution is  $50\text{km} \times 50\text{km}$  with 38 vertical levels. The run covers the period from 1990 to 2008 (Solman et al., 2013; Marengo et al., 2013a).

For all datasets, we focussed on daily values of the DJF seasons confined to the spatial domain from  $85^\circ\text{W}$  to  $30^\circ\text{W}$  and  $40^\circ\text{S}$  to  $5^\circ\text{S}$  (see Figure 8.1). TRMM and TRMM RT are first compared on their native  $0.25^\circ \times 0.25^\circ$  grid for the time period



**Figure 8.1.:** Topography of South America and important atmospheric features of the South American Monsoon System (SAMS).

2001 to 2012. The GPCP product is compared to TRMM for the period 1998 to 2008 at its native resolution of  $1.0^\circ \times 1.0^\circ$  after a bilinear interpolation of the TRMM data to this resolution. All other comparisons are carried out for the common time period from 1998 to 2008, with all datasets bi-linearly interpolated to a common spatial resolution of  $0.75^\circ \times 0.75^\circ$ .

## 8.4. Methods

Our dataset comparison consists of two steps: First, we analyze and compare local rainfall distributions and several traditional statistical values with a focus on the tail behavior of the distributions. Second, we investigate the spatial synchronicity structure of strong, extreme, and most extreme rainfall events. For the time period 1998 to 2008, we define these events as the top 200, top 100, and top 50 events at each location, corresponding to events above the 80th, 90th, and 95th percentile of the 11 DJF seasons, respectively: For strong events (above 80th percentile) we have:  $0.2 \times 11\text{years} \times 92\text{days} = 202$  events, for extreme events (above 90th percentile) we have  $0.1 \times 11\text{years} \times 92\text{days} = 101$  events, and for the most extreme events (above the 95th percentile) we obtain  $0.05 \times 11\text{years} \times 92\text{days} = 51$  events. For the comparison between TRMM and TRMM RT, we have 12 DJF seasons (2001 to 2012), which results in the top 216, 108, and 54 events at each grid cell. Locations with less than

the respective number of events are discarded from the analysis. Previous studies have demonstrated that TRMM 3B42 V7 performs very well in reproducing the main spatial patterns of the SAMS (see also chapter 4) and found overall good agreement between TRMM and GPCP (Carvalho et al., 2012).

To measure synchronicity of extreme events at different grid cells we employ  $ES^{sym}$ , here with a maximum delay of  $\tau_{max} = \pm 5$  days. It should be emphasized that, despite the relatively large  $\tau_{max}$ ,  $ES^{sym}$  only counts uniquely associable events within this time range. All pairs of grid cells for which the value of  $ES^{sym}$  is among the top 2% of all values will be represented by a network link. This link density of 2% is chosen such that all links correspond to significant values of  $ES^{sym}$  at a significance level of 5%. Significance is tested against a null model based on a uniformly random placement of events as in chapter 4. We will employ the following network measures in this chapter:

- Degree (DG): By simply counting the number of other grid cells with synchronous extreme events, DG assesses the importance of a grid cell for the distribution of extreme events to other locations.
- Betweenness Centrality (BC): This measure estimates the importance of a grid cell for the long-ranged, directed propagation of extreme events. For TRMM, regions with high BC values have been found in the ITCZ, over the Amazon Basin as well as along the eastern slopes of the Andes in the vicinity of the SALLJ (cf. chapter 4 and the associated publication P1).
- Clustering coefficient (CC): This measure estimates the spatial coherence of the occurrence of extreme events. High values have been observed in regions with frequent development of MCS (cf. chapter 4 and the associated publication P1).
- Regional Connectivity (RC): This measure can be used to assess where extreme events occur synchronously with events in a given region under consideration. We will focus on the central Amazon Basin as well as SESA and SACZ (Figure 8.1).

## 8.5. Results and Discussion

As noted above, several studies have concluded that TRMM 3B42 V7 provides reliable estimates of rainfall variability (Xue et al., 2013; Chen et al., 2013; Carvalho et al., 2012; Zulkaffi et al., 2014). We will therefore consider the results for the TRMM dataset as a reference and discuss deviations from these results for the other five datasets.

### 8.5.1. Statistical values of rainfall distributions

#### Mean daily rainfall

Mean daily rainfall rates (Figure 8.2) show considerable differences between the spatial distribution exhibited by TRMM and the remaining five datasets. For TRMM, we observe high values at the ITCZ, over the Amazon Basin, along the SACZ, near the Colombian Pacific coast as well as at the eastern slopes of the Peruvian and Bolivian Andes.

The TRMM RT dataset shows a very similar spatial distribution, but at the eastern slopes of the Peruvian and Bolivian Andes, mean daily rainfall values are up to 5mm/day higher. The GPCP dataset shows slightly lower values than TRMM in the ITCZ as well as at the eastern slopes of the Peruvian and Bolivian Andes, but is otherwise in good agreement with TRMM.

In contrast, the model-derived datasets (ERA, ECHAM6, and ETA) deviate substantially from TRMM. ERA mean daily rainfall values are higher in the eastern Amazon Basin, but lower in the western part of the basin as well as along the SACZ. Furthermore, rainfall values at the eastern slopes of the Argentinean Andes are much higher than for TRMM.

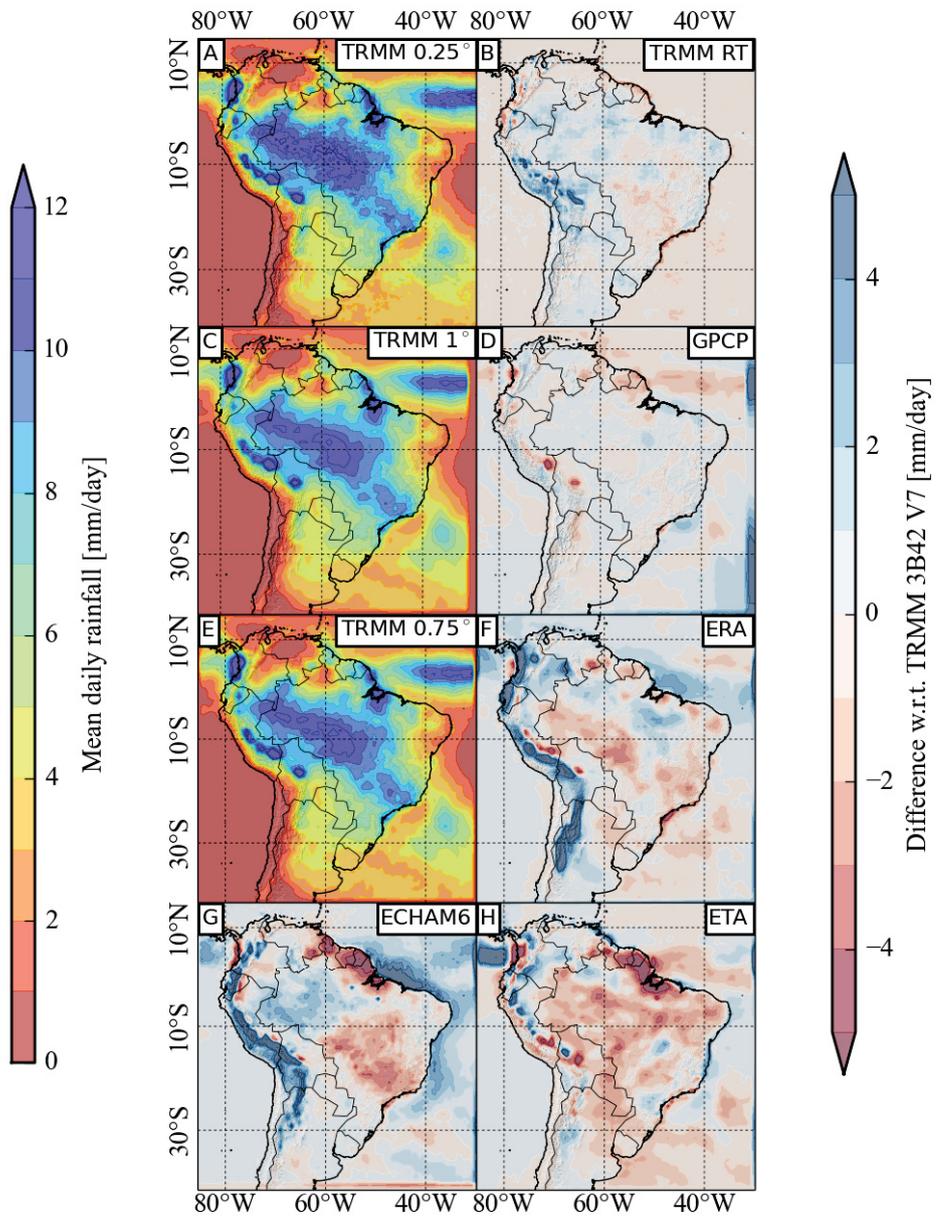
The mean daily values of ECHAM6 are higher than TRMM in the western Amazon Basin and along the eastern slopes of the Andes in northern Argentina and southern Bolivia, but diminished at the climatological position of the SACZ. East of 40°W, over the subtropical Atlantic Ocean, there is a band of slightly higher values than for TRMM in this region.

ETA mean daily values are lower than for TRMM in the entire Amazon Basin, along the SACZ, in most of SESA, in western Colombia as well as in northeastern Brazil. On the other hand, values are higher than for TRMM at the eastern slopes of the Ecuadorian Andes.

#### 90th and 95th percentiles of daily rainfall

For TRMM, scores at the 90th percentile (Figure 8.3) follow a very similar spatial distribution as the mean values. For all five remaining datasets, the deviations from TRMM's spatial distribution at the 90th percentile scores are qualitatively similar to the deviations observed for mean daily rainfall, however with higher differences in absolute terms. TRMM RT overestimates 90th percentiles at the eastern Andean slopes of Bolivia and Peru, while GPCP underestimates 90th percentiles in this region. Apart from additional discrepancies between TRMM and GPCP over the ITCZ, the observational datasets (TRMM RT and GPCP) are still the ones with best agreement with TRMM.

Compared to TRMM, the model-derived datasets show substantial deviations: ERA exhibits lower values than TRMM in the ITCZ, the entire Amazon Basin, and the SACZ. Meanwhile, we observe strongly overestimated rainfall scores at the eastern slopes and high-elevation regions of the central and southern Andes. The ECHAM6 model performs well in reproducing TRMM's spatial distribution of extreme scores in



**Figure 8.2.:** Mean daily rainfall during DJF for TRMM 3B42 V7 at 0.25° (A), 1° (C), and 0.75° (E) spatial resolution; Differences with respect to TRMM 3B42 V7 for TRMM 3B42 V7 RT (B), GPCP 1DD V1.2 (D), ERA-interim (F), ECHAM6 (G), and ETA (H).

the western and central Amazon Basin, while values are too high along the ITCZ and too low in the eastern Amazon Basin and along the continental part of the SACZ. Rainfall scores are strongly overestimated along the entire Andes mountain range. In contrast, the ETA model reproduces the values obtained for TRMM reasonably well in the SACZ, but not in the Amazon Basin, northeastern Brazil, and in SESA, where values are too low. For all six datasets, the spatial distribution (but not the absolute values) of 95th percentile scores (Figure 8.4) resembles that of the 90th percentile, the exception being locally higher scores in northeastern Argentina and southern Brazil for TRMM, TRMM RT, GPCP, and ECHAM6, but not for ERA and ETA.

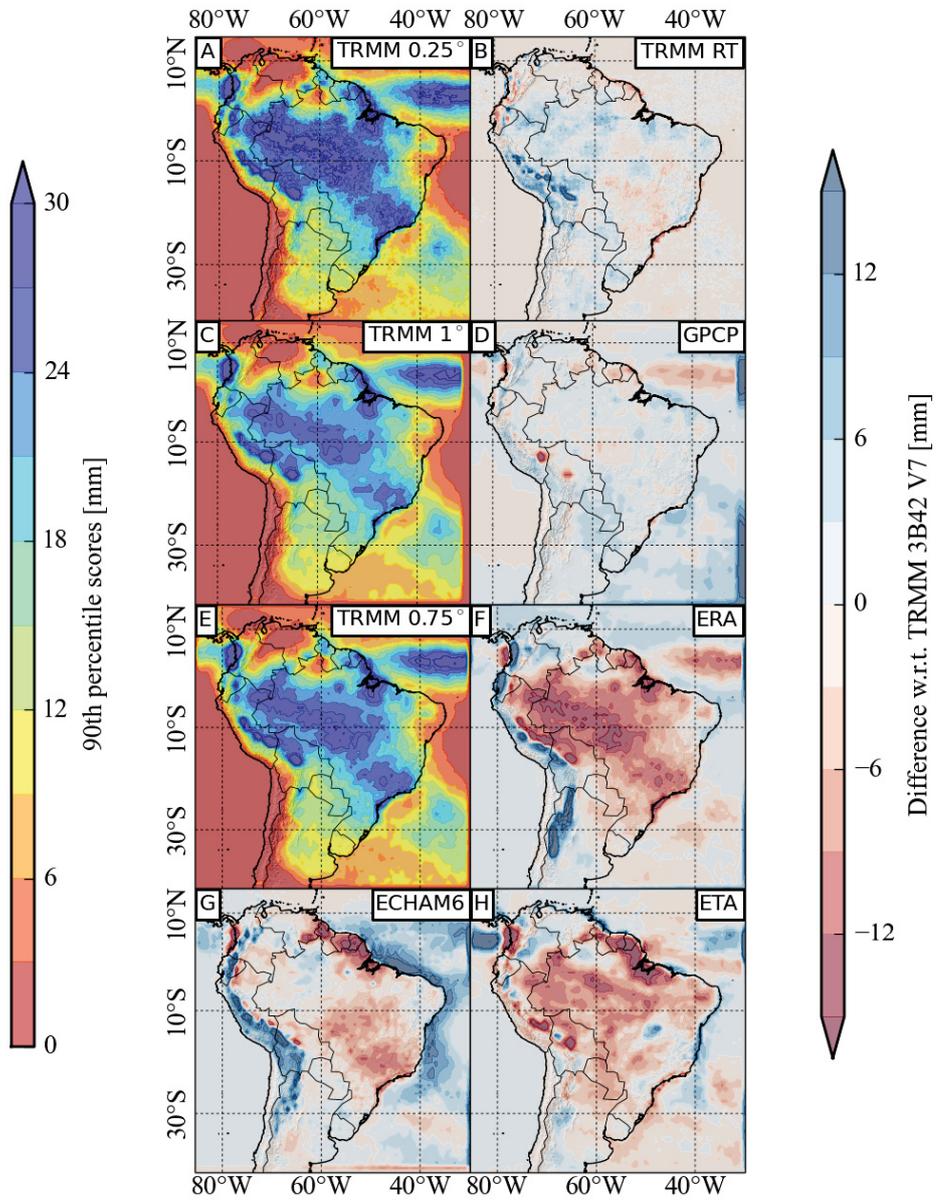
### Fractions of total seasonal rainfall

The fraction of total DJF rainfall accounted for by events above the 90th percentile (Figure 8.5) shows a pronounced gradient from the tropics to the subtropics for all six datasets. For TRMM and TRMM RT, more than 70% of total DJF rainfall is caused by these events in large parts of South America south of 20°S, while they account for about 50% in the Amazon Basin and in the vicinity of the SACZ. For GPCP, we find only very small deviations from TRMM over the South American continent, while values are considerably lower over the Atlantic Ocean north of the equator and south of 20°. These fractions are smaller for the ERA dataset, with 40 – 80% in southern South America and less than 30% in the Amazon Basin and along the SACZ. For ECHAM6, fractions are closer to the fractions found for TRMM, with 30 – 60% in the Amazon Basin and at the SACZ and 60 – 80% further south. The ETA model produces values between the fractions found for ERA and ECHAM6. The spatial pattern for all six datasets is remarkably similar. We note that fractions accounted for by events above the 95th percentile have a very similar spatial distribution as fractions for events above the 90th percentile, with, of course, reduced values (not shown).

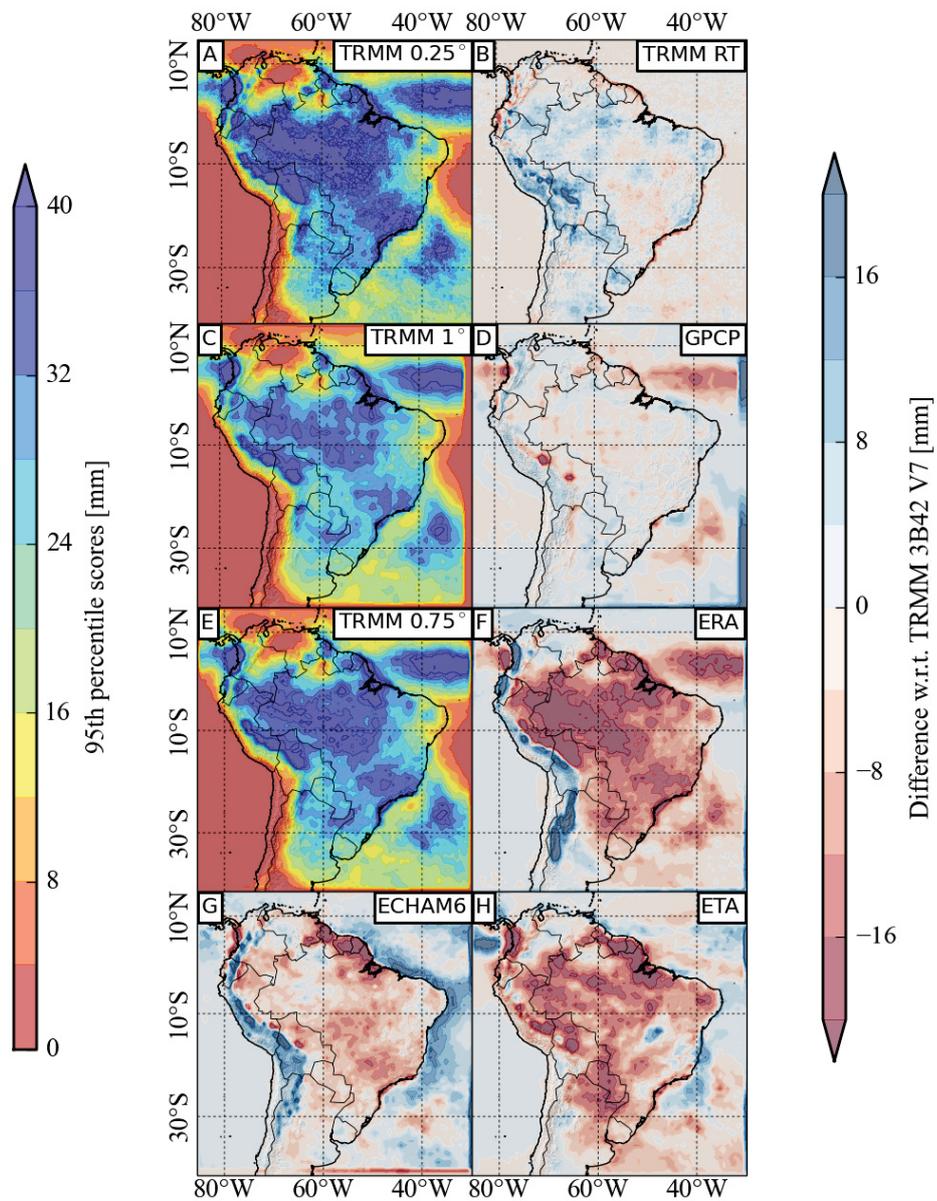
The fact that the model-derived datasets (ERA, ECHAM6, and ETA) show lower values in SESA indicates that the models fail to reproduce the low-frequency and high-magnitude characteristics found for the observational datasets (TRMM, TRMM RT, and GPCP) in this region. Furthermore, the above mentioned latitudinal gradient is less pronounced for the three model-derived datasets.

### Difference between 90th and 50th percentiles

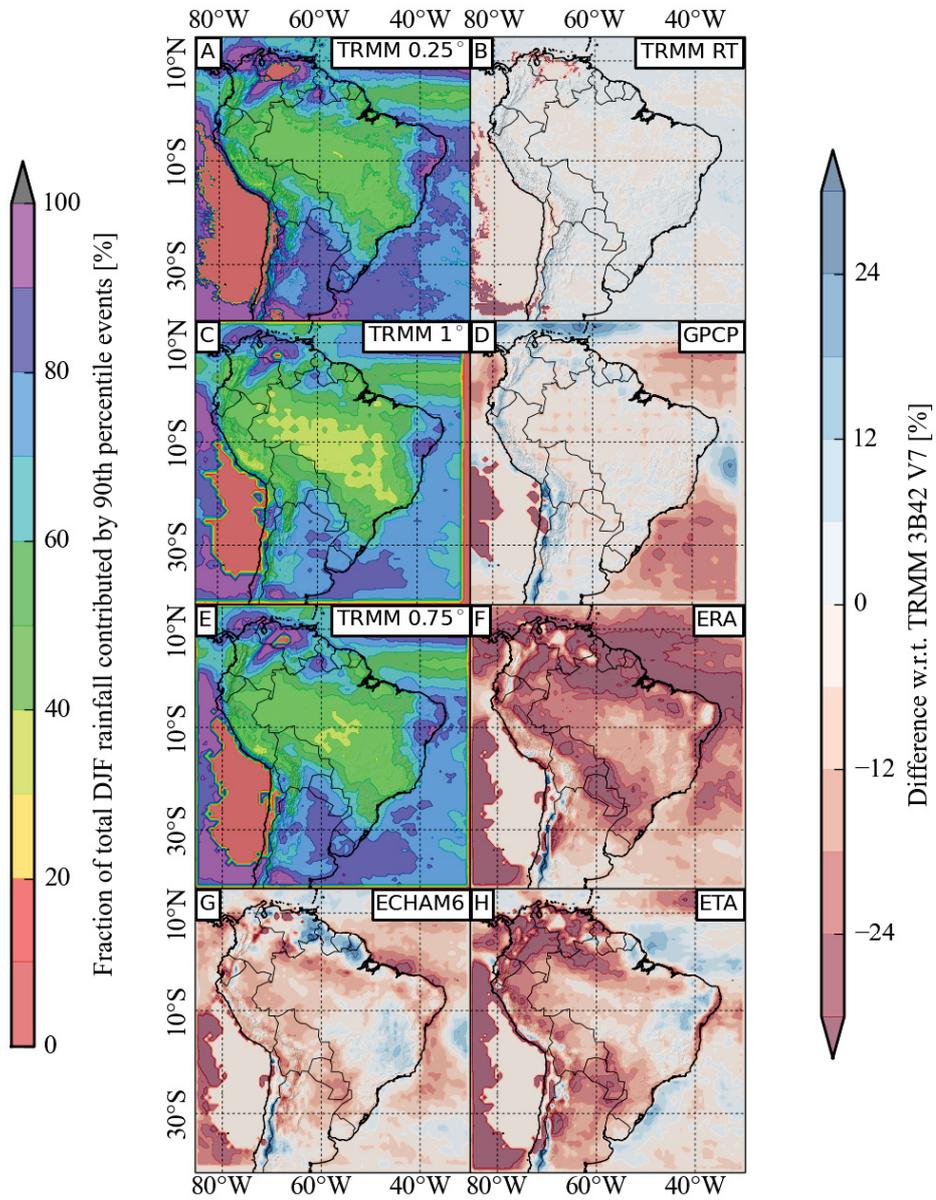
The difference between local scores at the 90th and at the 50th percentile (Figure 8.6) yield a relatively simple estimate of the tail behavior of the local rainfall distributions at each grid point. For TRMM, we observe high values in the ITCZ and the northern Brazilian coast, at the Colombian Pacific coast as well as at the eastern slopes of the Central Andes in Peru and Bolivia. Intermediate values can be found in the Amazon Basin and the adjacent SACZ. Again, TRMM RT has higher values than TRMM at the eastern slopes of the central Andes, but is otherwise in good agreement with



**Figure 8.3.:** Scores at the 90th percentile of daily rainfall during DJF for TRMM 3B42 V7 at 0.25° (A), 1° (C), and 0.75° (E) spatial resolution; Differences with respect to TRMM 3B42 V7 for TRMM 3B42 V7 RT (B), GPCP 1DD V1.2 (D), ERA-interim (F), ECHAM6 (G), and ETA (H).



**Figure 8.4.:** Scores at the 95th percentile of daily rainfall during DJF for TRMM 3B42 V7 (at 0.25° (A), 1° (C), and 0.75° (E) spatial resolution; Differences with respect to TRMM 3B42 V7 for TRMM 3B42 V7 RT (B), GPCP 1DD V1.2 (D), ERA-interim (F), ECHAM6 (G), and ETA (H).



**Figure 8.5.:** Fraction of total DJF rainfall accounted for by events above the 90th percentile of daily rainfall for TRMM 3B42 V7 at 0.25° (A), 1° (C), and 0.75° (E) spatial resolution; Differences with respect to TRMM 3B42 V7 for TRMM 3B42 V7 RT (B), GPCP 1DD V1.2 (D), ERA-interim (F), ECHAM6 (G), and ETA (H).

TRMM. The GPCP dataset exhibits lower values than TRMM in the ITCZ and at the eastern central Andean slopes, but is otherwise also in good agreement.

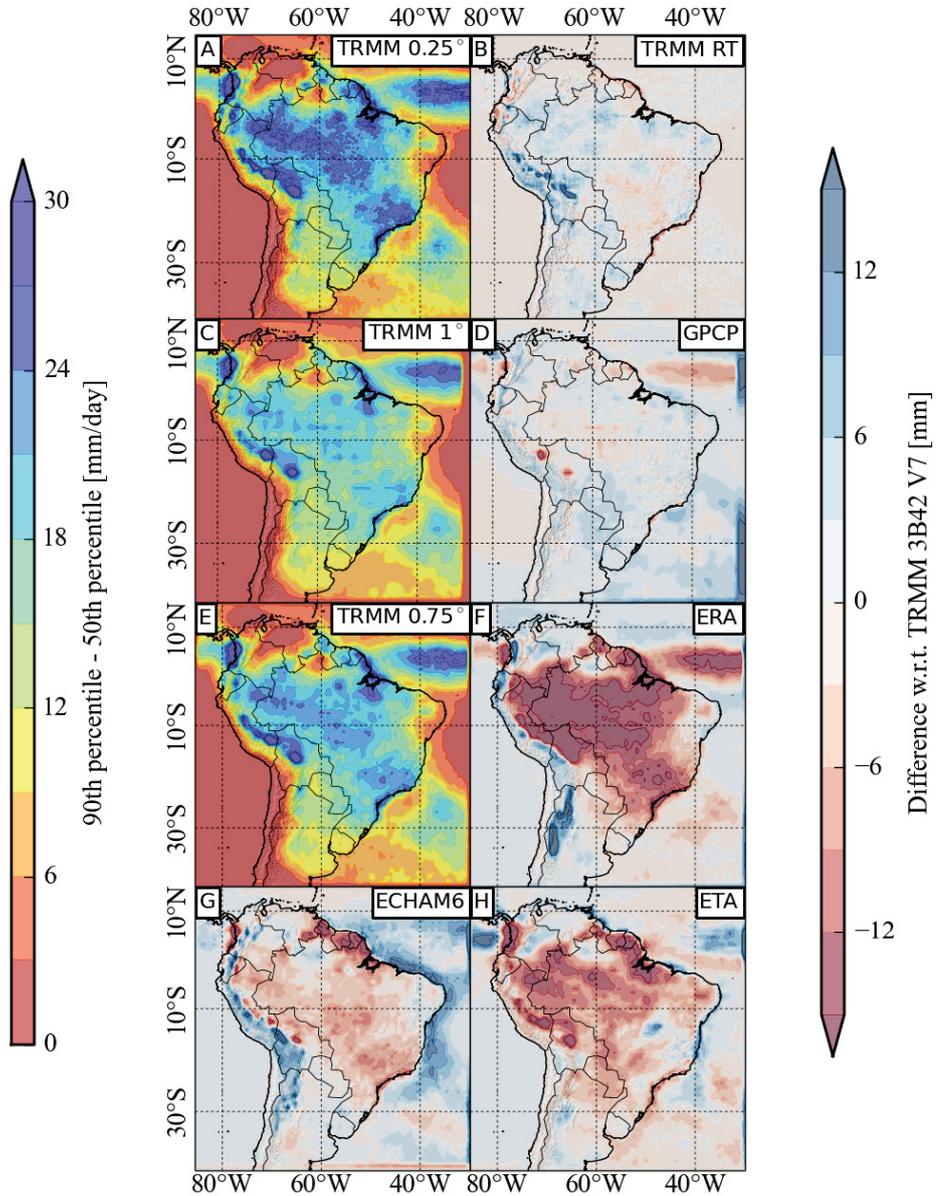
The model-derived datasets perform rather poorly in reproducing the values found for TRMM. Most notably, they strongly underestimate the distributions' tails in the entire Amazon Basin. For ERA, values are lower than for TRMM, with locally higher values only at the Colombian Pacific coast and the eastern slopes of the Argentinean Andes. In the entire Amazon Basin and the SACZ, values are strongly reduced as compared to TRMM (values  $< 10\text{mm/day}$  instead of between  $15\text{mm/day}$  and  $25\text{mm/day}$ ). For ECHAM6, values are higher than for ERA but still lower than for TRMM in the Amazon Basin and the continental part of the SACZ. At the ITCZ and the adjacent coast, but also at the eastern slopes of the Peruvian and Bolivian Andes, values are close to the values found for TRMM. However, values are too high at the northern Argentinean Andes. The ETA dataset exhibits lower values than TRMM in the Amazon Basin, but higher values in some eastern parts of the continental SACZ.

### Differences between 95th and 90th percentiles

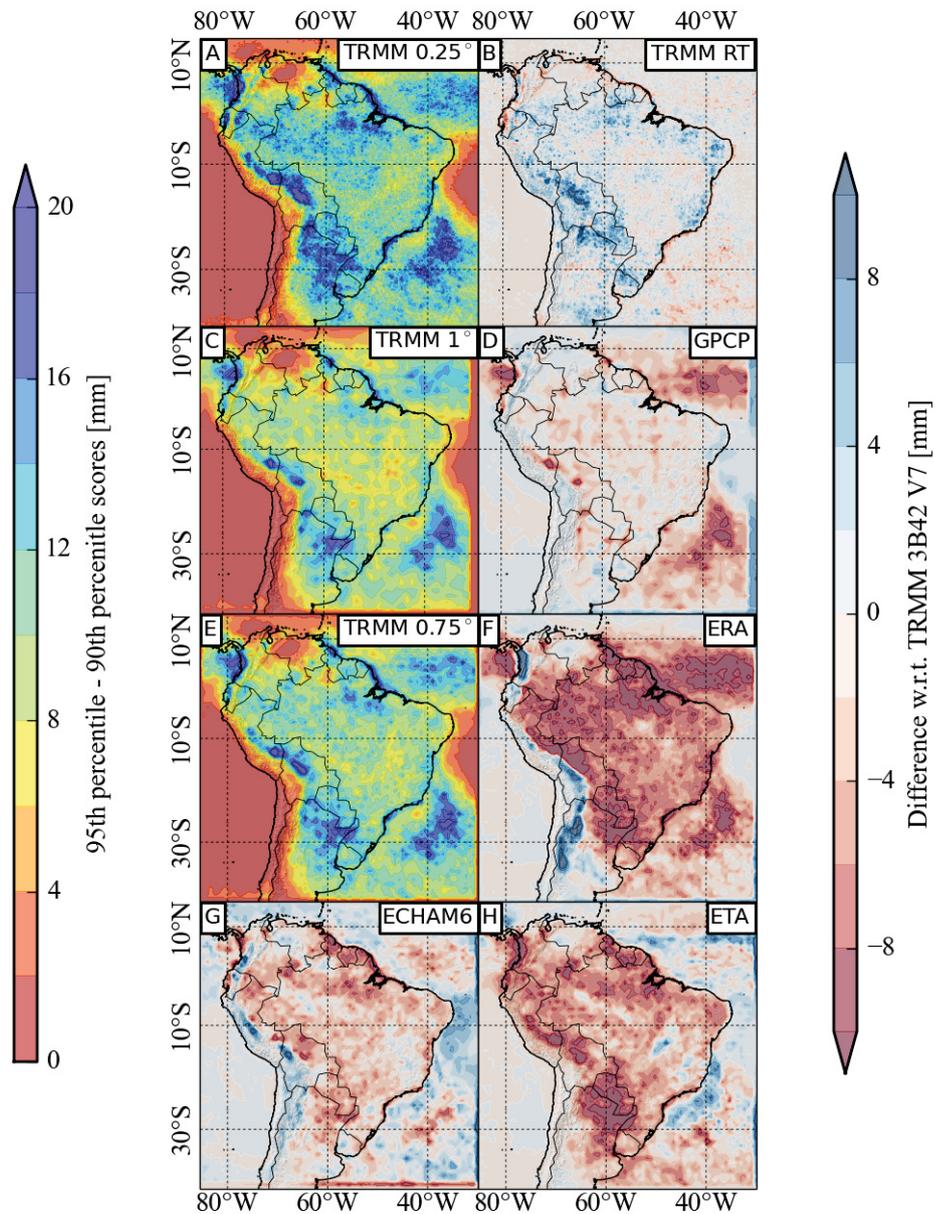
In order to assess the different datasets' behavior for the most extreme events, we computed the difference between scores at the 95 and the 90th percentile (Figure 8.7). For the TRMM dataset, the difference between scores at the 95th and 90th percentile exhibits high values in the ITCZ and the coastal areas around the equator, at the Colombian Pacific coast, and at the eastern slopes of the Peruvian and Bolivian Andes. Furthermore, high values can be observed in northern and northeastern Argentina as well as in the oceanic part of the SACZ. Meanwhile, the continental part of the SACZ does not show a particularly high difference between the 95th and the 90th percentile. In large parts of the continent, TRMM RT shows higher values than TRMM, in particular in the Bolivian Andes, in northern Argentina, Paraguay, and southern Brazil. GPCP is in fair agreement with TRMM, apart from underestimated values at the ITCZ and the eastern central Andes. However, rather large differences from TRMM can be observed over the ocean (in particular over the Atlantic Ocean south of  $20^\circ$ ), where no rain gauge data are available.

For the ERA dataset, difference between scores at the 95th and 90th percentile exhibit lower values than for the TRMM dataset over the entire continent, except for the western slopes of the Colombian Andes and the eastern slopes of the eastern slopes of the Bolivian and Argentinean Andes, where values are locally high. The ECHAM6 model data show high values over the ITCZ, the Colombian Pacific coast, the eastern Andes from Peru to Argentina as well as over the subtropical Atlantic Ocean east of  $40^\circ\text{W}$ . For ETA, we observe high values over the ITCZ and the oceanic part of the SACZ comparable to TRMM. However, in the remainder of the continent values are lower than for TRMM, in particular, the high values in northwestern Argentina are not well reproduced by the ETA model.

High values for the observational datasets (TRMM, TRMM RT, and GPCP) in northeastern Argentina can be identified with the development of MCS in this region (Durkee et al., 2009), which are thus responsible for the most extreme rainfall events.



**Figure 8.6.:** Difference between scores at the 90th and the 50th percentile (90th percentile - 50th percentile) of daily rainfall during DJF for TRMM 3B42 V7 at 0.25° (A), 1° (C), and 0.75° (E) spatial resolution; Differences with respect to TRMM 3B42 V7 for TRMM 3B42 V7 RT (B), GPCP 1DD V1.2 (D), ERA-interim (F), ECHAM6 (G), and ETA (H).



**Figure 8.7.:** Difference between scores at the 95th and the 90th percentile (95th percentile - 90th percentile) of daily rainfall during DJF for TRMM 3B42 V7 at 0.25° (A), 1° (C), and 0.75° (E) spatial resolution; Differences with respect to TRMM 3B42 V7 for TRMM 3B42 V7 RT (B), GPCP 1DD V1.2 (D), ERA-interim (F), ECHAM6 (G), and ETA (H).

None of the three model-derived datasets feature a particularly heavy tail of the rainfall distribution in this area, suggesting that all of them fail to reproduce the most extreme rainfall events originating from the MCS.

The strongest orographic rainfall peaks at the eastern flanks of the Peruvian and Bolivian Andes (Bookhagen and Strecker, 2008) are overestimated by TRMM RT, ERA, and ECHAM6, but underestimated by GPCP and ETA.

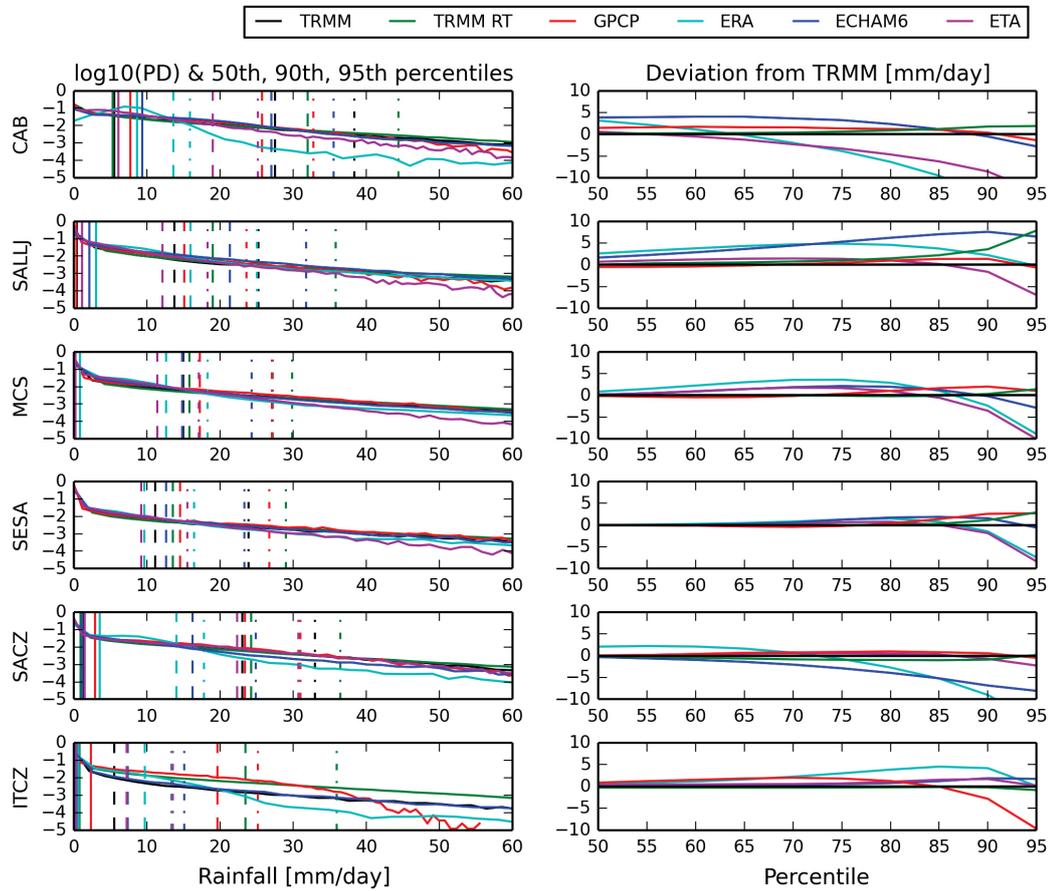
### Regional characteristics of rainfall distributions

For the spatial boxes denoted by *CAB*, *SALLJ*, *MCS*, *SESA*, *SACZ*, and *ITCZ* in Figure 8.1, we computed the 50th to 95th percentiles (in steps of 5) of the respective rainfall distributions (Figure 8.8, left panel) and the difference of each dataset from the reference TRMM dataset (Figure 8.8, right panel).

**CAB.** In the CAB, the TRMM RT dataset shows small positive deviations for the highest percentiles, while we observe small positive deviations for GPCP in particular for lower percentiles. In contrast, with increasing percentiles, ERA and ETA show increasing negative deviations. Underestimation of rainfall over the central Amazon by a collection of regional models (including ETA) was also observed by Solman et al. (2013). The ECHAM6 model, in contrast, overestimates daily rainfall scores up to the 80th percentile in this region, but underestimates the 95th percentile score.

**SALLJ.** In the vicinity of the SALLJ east of the central Andes, TRMM RT shows considerable positive deviations only for the 90th and 95th percentile, and overall good agreement with TRMM for the remaining percentiles. GPCP scores are close to TRMM scores on the entire domain from the 50th to the 95th percentile. ERA and ECHAM6 show strong positive deviations from TRMM for the entire domain. However, deviations for ERA become small for high percentiles, while for ECHAM6, differences increase up to the 90th percentile. This overestimation of orographic rainfall, in particular at the eastern slopes of the Andes, is a well-known problem shared by most global and regional circulation models (Urrutia and Vuille, 2009; Solman et al., 2013). In contrast, the ETA model produces values close to TRMM up to the 85th percentile; for higher percentiles the scores are lower than for TRMM. The ETA model thus performs better than ERA and ECHAM6 over this mountainous region for the lower percentiles of the distribution, probably due to the step-like representation of mountains in its vertical coordinate system (the so-called *eta coordinate*) (Mesinger, 1984; Mesinger and Black, 1992; Pesquero et al., 2009). Still, the 90th and 95th percentiles are underestimated by ETA when compared to TRMM.

**MCS.** In the MCS region in subtropical South America, TRMM RT and GPCP are in good agreement with TRMM for all percentiles in this region. ERA, ECHAM6, and ETA exhibit higher scores than TRMM up to the 85th percentile, and negative deviations for the 90th and 95th percentile. However, ECHAM6 remains close to TRMM, while ERA and ETA deviate strongly for the 95th percentile. Problems to implement the particularly strong cyclogenetic activity in this region have been observed for a number of global and regional models (Solman et al., 2007; Chou et al.,



**Figure 8.8.:** Left panel: Logarithmic rainfall distributions averaged for the Amazon, the SALLJ region, the MCS region, SESA, SACZ, and the ITCZ (see Figure 8.1 for definitions of these regions). Vertical lines denote median (solid), 90th percentile (dashed), and 95th percentile (dot-dashed) scores of the respective distributions. Right panel: Percentile score differences with respect to TRMM 3B42 V7 for the same regions.

2011; Solman et al., 2013). Our results here suggest that these deficiencies are mainly due to the underestimation of the most extreme events.

**SESA.** Due to considerable spatial overlap, it is little surprising that for SESA, we observe a similar behavior as for the MCS region. However, discrepancies are smaller up to the 85th percentile and larger for the 90th and 95th percentile than for the latter region.

**SACZ.** In the SACZ region, values agree fairly well between TRMM, TRMM RT, GPCP, and ETA. The good performance of the ETA model in reproducing rainfall in this area was also observed by Solman et al. (2013). For higher percentiles, ERA and ECHAM6 show considerable negative deviations from TRMM, indicating that these two model-derived datasets fail to reproduce strong and extreme precipitation in the SACZ region.

**ITCZ.** Over the Atlantic ITCZ, TRMM RT has almost no deviations from TRMM. This is not surprising, since there are no gauge stations that can be used for post-processing adjustment. GPCP exhibits considerably lower 90th and 95th percentile scores than TRMM. ERA shows moderate positive deviations for low percentiles, which however increase to about 5mm at the 90th percentile. ECHAM6 and ETA are in good agreement with TRMM on the entire percentile domain.

The results described so far all concern statistical properties of local rainfall distributions. In order to investigate the spatial structure of synchronization between extreme events at different locations, we will now turn to the results of our CN approach.

### 8.5.2. Complex network measures

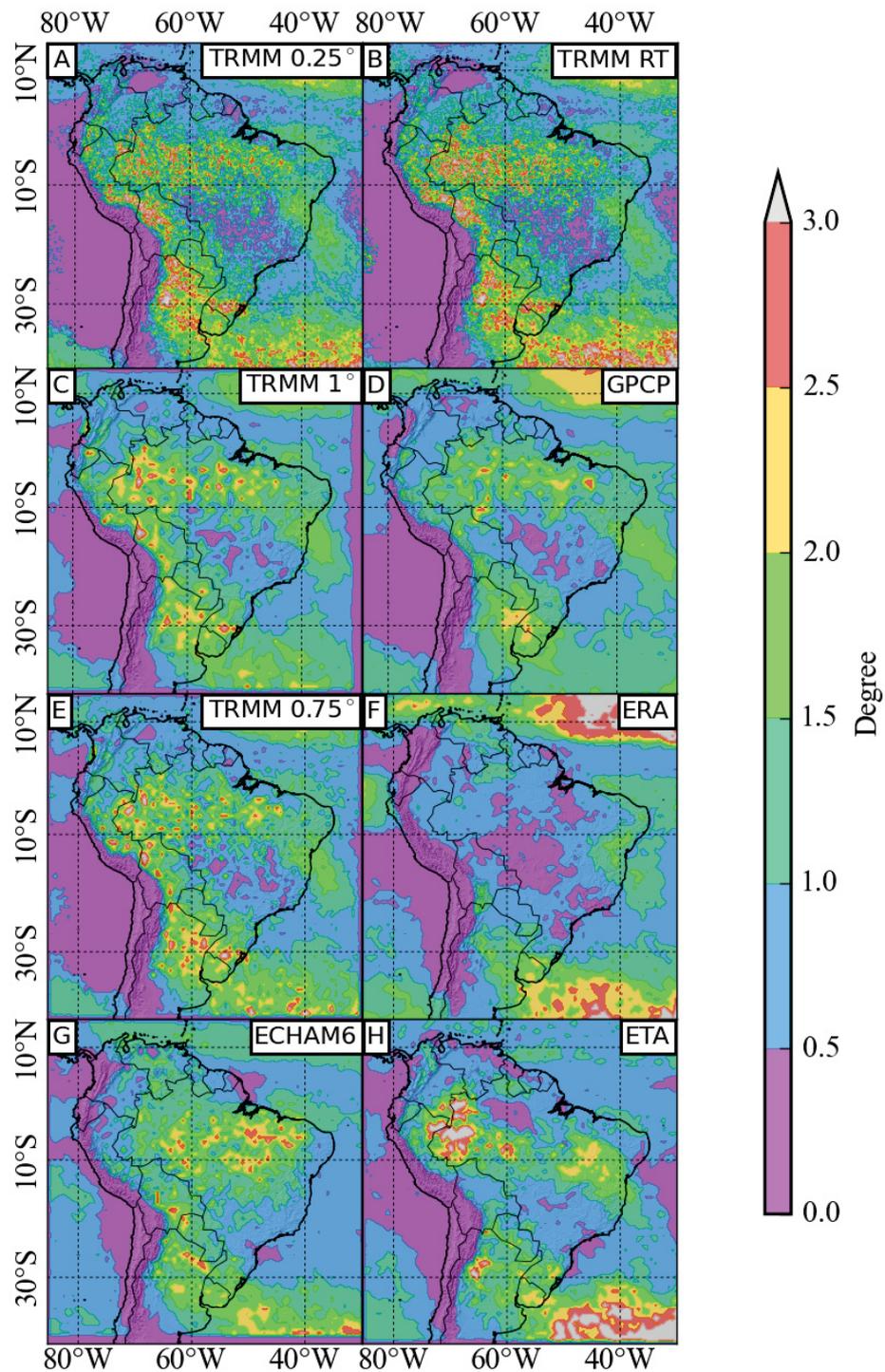
In the following, we will describe the spatial characteristics of extreme rainfall synchronicity, which are reflected by the network measures degree (DG), betweenness centrality (BC), clustering (CC), and regional connectivity (RC), introduced in section 3.4.1. All measures are computed for network measures derived for a maximum delay of  $\tau_{max} = \pm 5$  days.

For all six datasets, these network measures are obtained by constructing networks as described in section 8.4 for rainfall events above the 80th, 90th, and 95th percentile, respectively.

Instead of presenting differences from TRMM as for local statistical values in the last section, we will show absolute values for each dataset, because rather than specific local values, the overall spatial patterns are important for this kind of analysis. For the three network measures DG, BC, and CC, we will also compare spatial averages between the different datasets for the areas Amazon, SALLJ, MCS, and ITCZ (Figure 8.15).

#### Degree (DG)

As explained in section 3.4.1, we expect the DG to be high at locations which are particularly important for the direct distribution of extreme rainfall over the continent:



**Figure 8.g.:** Network measure degree calculated for the 90th percentile of all DJF seasons for TRMM 3B42 V7 at 0.25° (A), 1° (C), and 0.75° (E) spatial resolution, as well as for TRMM 3B42 V7 RT (B), GPCP 1DD V1.2 (D), ERA-interim (F), ECHAM6 (G), and ETA (H).

extreme events at these grid cells are strongly synchronized with extreme events at many other locations.

DG for 90th percentile events of the TRMM dataset (Figure 8.9 and left column of Figure 8.15) exhibits a connected area of high values from northeastern Brazil westward across the Amazon Basin, and along the eastern slopes of the central Andes towards central Argentina, SESA and the adjacent subtropical Atlantic Ocean. The SACZ is characterized by low values of DG. This spatial pattern indicates the main and well-known climatological moisture pathways along which extreme rainfall events synchronize (P1; Vera et al., 2006; Marengo et al., 2012): These pathways lead from the mouth of the Amazon River across the Amazon Basin and continue farther west to the Andes, where the mountain range blocks the low-level winds and channels them southwards. Extreme events follow this wind channel (the SALLJ), cause abundant rainfall at the eastern slope of the Andes through orographic lifting and continue further towards northern Argentina and SESA.

For the TRMM RT dataset, we observe a very similar spatial distribution of DG as for the gauge-calibrated TRMM version. Also for GPCP, the spatial pattern is similar to the one found for TRMM. However, lower values along the eastern slopes of the Andes can be observed, which may be due to low numbers of measurement stations in these regions, as well as interpolations to a regular  $1^\circ$  grid.

In contrast, for ERA we get high values only north of the ITCZ over the Atlantic Ocean, in Uruguay and eastern Argentina as well as over the adjacent subtropical Atlantic Ocean. All other parts of the continent exhibit quite continuously low DG values and the pattern which was observed for TRMM is absent. ERA thus fails to reproduce the main synchronization pathway of extreme events exhibited by TRMM.

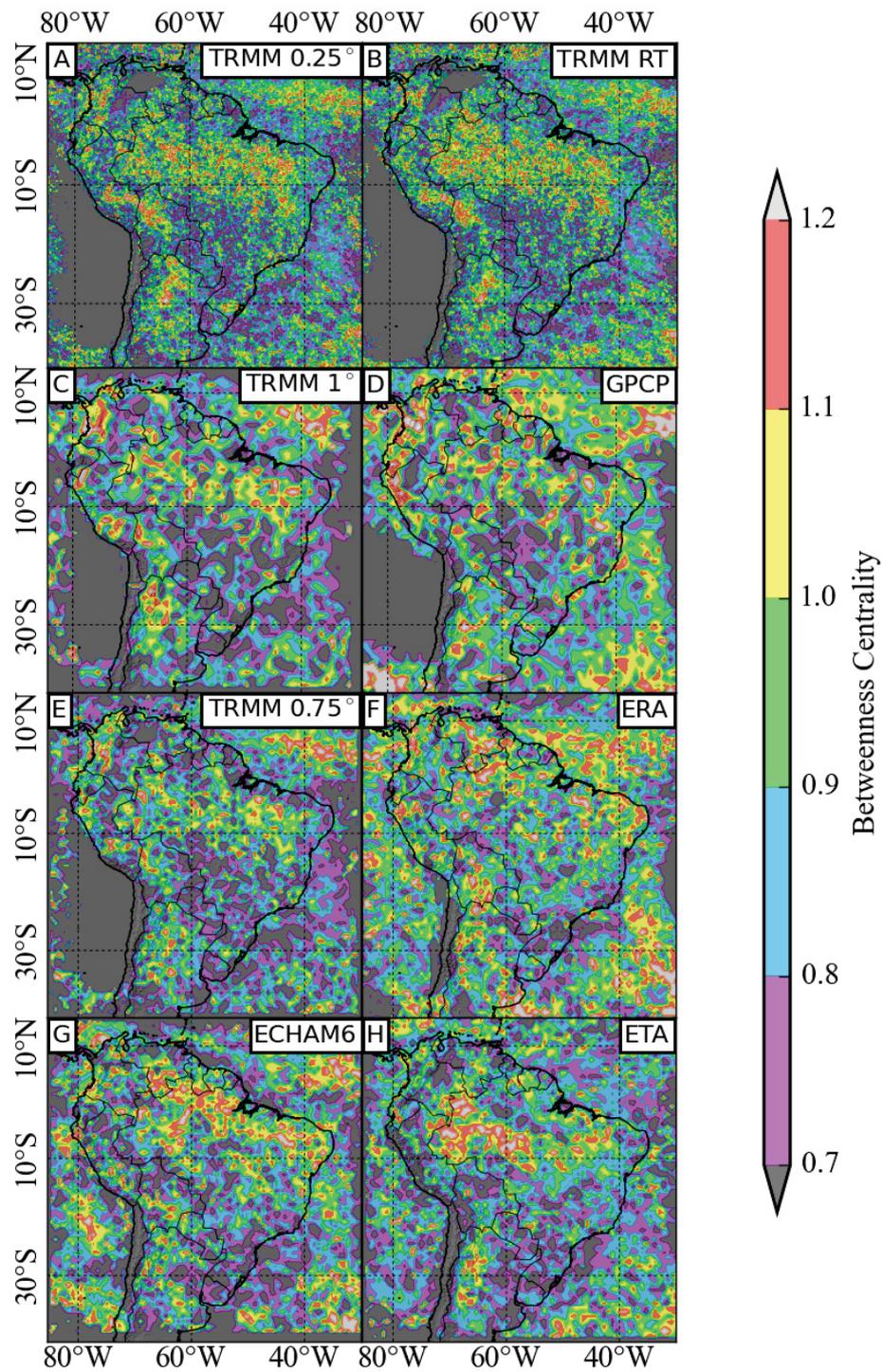
DG for the ECHAM6 model data is high in northeastern Brazil, over the Amazon Basin as well as at the eastern slopes of the Bolivian Andes, in most of SESA and over the adjacent subtropical Atlantic Ocean. However, values for ECHAM6 are lower as compared to TRMM over the western Amazon Basin and at the slopes of the Peruvian and northern Bolivian Andes, indicating that this model does not accurately represent the southwestward propagation of extreme rainfall from the Amazon.

The ETA model data yield high DG over the western Amazon Basin, in northern Argentina, and over the subtropical Atlantic Ocean south of  $30^\circ\text{S}$ . No high values can be found along the eastern slopes of the Andes. Despite these substantial discrepancies, we argue that the ECHAM6 as well as the ETA model reproduce the large-scale DG pattern to a reasonable extent.

### **Betweenness centrality (BC)**

For BC (Figure 8.10 and middle column of Figure 8.15), we expect a strong emphasis on long-ranged, directed synchronization pathways, because this measure is defined on the basis of shortest paths in the network. In contrast, DG only measures the connectivity to direct CN neighbors of a grid cell.

For the 90th percentile events of TRMM, BC exhibits high values over the ITCZ, from northeastern Brazil across the Amazon Basin and along the eastern slopes of



**Figure 8.10.:** Network measure betweenness centrality calculated for the 90th percentile of all DJF seasons for TRMM 3B42 V7 at 0.25° (A), 1° (C), and 0.75° (E) spatial resolution, as well as for TRMM 3B42 V7 RT (B), GPCP 1DD V1.2 (D), ERA-interim (F), ECHAM6 (G), and ETA (H).

the Andes southwards to central Argentina. In addition, relatively high values can be observed over the SACZ. This spatial distribution of BC has been found to be consistent with the interpretation of large scale, directed moisture pathways in chapter 4 and the associated publication P1. Apart from being less homogeneous, the spatial pattern is similar to the one for DG over eastern Brazil, the Amazon Basin, along the Andes and over SESA. On the other hand, BC is high in the vicinity of the ITCZ and, to weaker extent, in the vicinity of the SACZ, where DG is particularly low. This highlights the role of these convergence zones for large-scale propagation of extreme events over the entire continent, since their high connectivity only becomes apparent when taking into account network paths incorporating several other locations, as opposed to only local next neighbor connections.

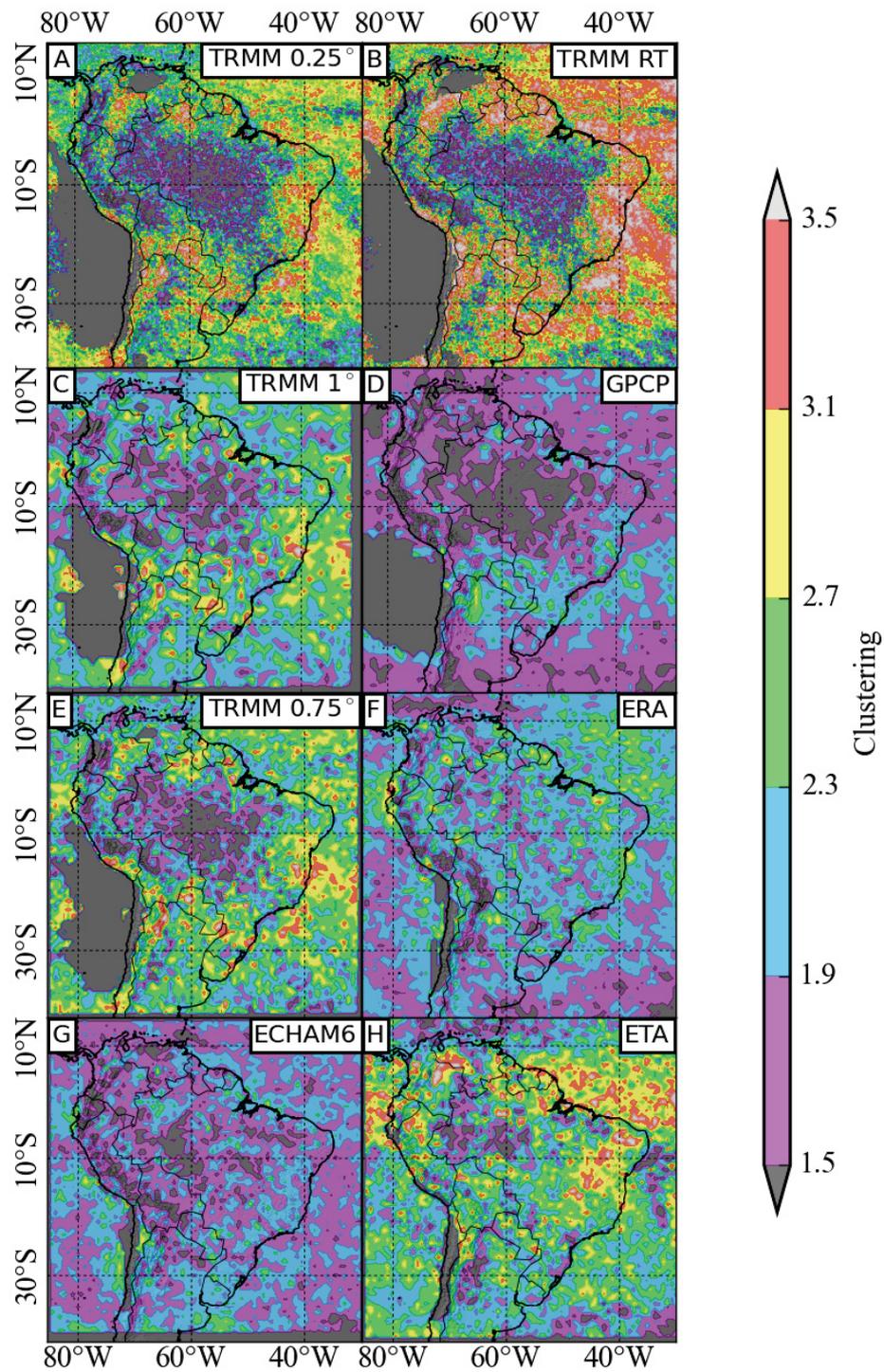
The spatial pattern of BC for the TRMM RT dataset is in very good agreement with the results for TRMM, indicating that the large-scale propagation of extreme events is already captured well by the satellite-only product.

Similarly, results for the GPCP dataset are in fair agreement with TRMM's spatial pattern, with only small discrepancies at the eastern slopes of the southern Central Andes.

In strong contrast to DG, the spatial pattern of BC exhibited by ERA does quite accurately resemble the spatial pattern found for TRMM, with high values marking the large-scale propagation pathway from the ITCZ towards the Andes and southwards to northern Argentina (Figure 8.10). Since BC is the more sophisticated measure for the long-range, directed propagation pathways, we conclude that ERA reproduces the climatological synchronization pathway, although it does not perform well in estimating the local connectivities along the pathway.

For the ECHAM6 Model, there are high values of BC over the ITCZ as well, but also over the entire coast of northeastern South America as well as the CAB. Furthermore, high values can be observed in Bolivia and northern Argentina east of the Andes. Hence, this model accurately incorporates the large-scale propagation paths.

The ETA model yields high BC in the central and western Amazon Basin. Values are also relatively high at the eastern slopes of the Southern Andes of Argentina. However, over the ITCZ and in the vicinity of the SALLJ in Bolivia they are not as high as for the other five datasets. Hence, the ETA model does not perform well in reproducing the large-scale transport route from the ITCZ across the Amazon and southward along the Andes to the subtropics. In particular, problems with implementing the SALLJ's impact on extreme rainfall and corresponding orographic effects at the eastern slopes of the Central Andes are apparent. The latter observation, which is corroborated by the results obtained for DG, is in agreement with the underestimation of 90th and 95th percentile scores by the ETA model at the eastern Andean slopes described in section 8.5.1.



**Figure 8.11.:** Network measure clustering calculated for the 90th percentile of all DJF seasons for TRMM 3B42 V7 at 0.25° (A), 1° (C), and 0.75° (E) spatial resolution, as well as for TRMM 3B42 V7 RT (B), GPCP 1DD V1.2 (D), ERA-interim (F), ECHAM6 (G), and ETA (H).

### Clustering (CC)

Extreme rainfall in regions with high CC values (Figure 8.11 and right column of Figure 8.15) can be expected to occur in a spatially coherent manner; in particular, this is the case for large thunderstorms, squall lines, and MCS.

For events above the 90th percentile (Figure 8.11 and right column of Figure 8.15), the TRMM dataset exhibits high CC at the mouth of the Amazon river, in eastern Brazil, at the slopes of the northern Argentinean Andes as well as in most of SESA including most of Paraguay. The Amazon Basin and the SACZ are characterized by low CC. In chapter 4, we attributed high CC values in northern Argentina, Paraguay, and southern Brazil to the frequent development of MCS in this region. Similarly, high values in northern Brazil close the Atlantic coast may correspond to the so-called Amazonian squall lines (Cohen et al., 1995). Furthermore, rainfall on the Altiplano and Puna de Atacama plateaus in northwestern Argentina and southwestern Bolivia has been found to typically occur in a way that it either rains on the entire plateau for several days in a row, or it does not rain on the entire plateau (Garreaud, 2000b). This spatial coherence of rainfall events is expressed by high CC values over the plateau for the TRMM dataset.

While the relative spatial pattern of CC for TRMM RT looks similar to the one for TRMM, we observe substantially higher values for the satellite-only product in all regions and for all three event types.

Although the GPCP dataset yields too low CC values in the entire spatial domain, it has to be noted that the only region of relatively enhanced values is the subtropical region around Paraguay. Thus, GPCP does in this sense still discover this area of frequent MCS development, although values are much lower than for TRMM for all three percentiles.

ERA as well as ECHAM6 do not show any concise spatial pattern of CC, suggesting that these models have problems to reproduce large convective systems (e.g., MCS or large thunderstorms).

In contrast, the ETA model shows a clear spatial signature for CC, with relatively low values over the CAB, although they are still much higher than for TRMM. However, while some similarities to the pattern obtained for TRMM are apparent, values are not particularly high in Paraguay and northern Argentina.

### Regional Connectivity (RC)

Regional connectivity of a given region yields an estimate of where extreme events typically occur synchronously with events in that region.

**CAB.** Events above the 90th percentile in the CAB (Figure 8.12) are synchronized with events in the entire Amazon Basin, but also with events at the eastern slopes of the Peruvian and northern Bolivian Andes for TRMM. Furthermore, there are weak linkages to northern Argentina. TRMM RT exhibits very similar regional connectivity for the central Amazon. For both TRMM and TRMM RT, weak connections to the eastern Amazon Basin can be interpreted as a consequence of the dominant easterly

direction of low-level winds in this area, driving extreme rainfall events from the eastern part of the Amazon to the west. However, we also observe connection to the west until the eastern slopes of the Central Andes in northern Bolivia. This shows that extreme events in the Amazon Basin are climatologically synchronized with extreme events at the eastern slopes of the Andes. On longer spatial scales, we observe a linkage between extreme rainfall in the Amazon and a small area in northern Argentina at the foothills of the Andes. These results support the hypotheses that the Amazon plays a crucial role for distributing and propagating extreme rainfall events to large parts of the remaining continent, as it was suggested by high DG and BC values.

For GPCP, the CAB is only connected to the larger area of the Amazon Basin and the slopes of the Peruvian Andes, but not of the Bolivian Andes, possibly due to a lack of gauge stations in this area.

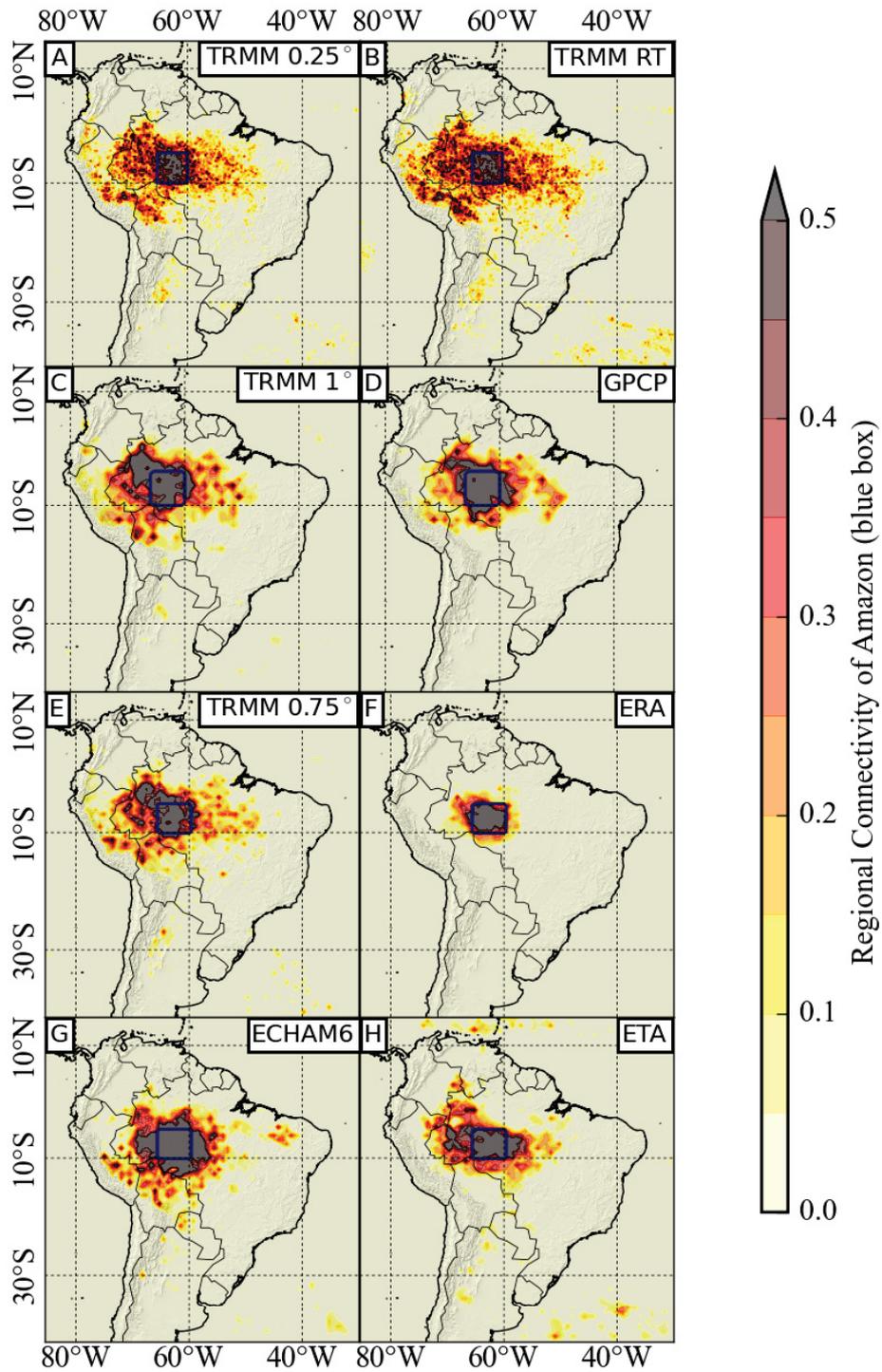
In comparison, for ERA the connectivity of the central Amazon is more localized, with almost no links pointing out of the region. In particular, the pattern of long-ranged connection between the Amazon Basin and the eastern slopes of the Peruvian and Bolivian Andes is not reproduced by ERA.

In contrast, the ECHAM6 model produces strong connections to the eastern part of the Amazon Basin as well as to the Peruvian and Bolivian Andes to the west in a similar way as observed for TRMM. However, the teleconnection from the Amazon Basin to northern Argentina is not visible.

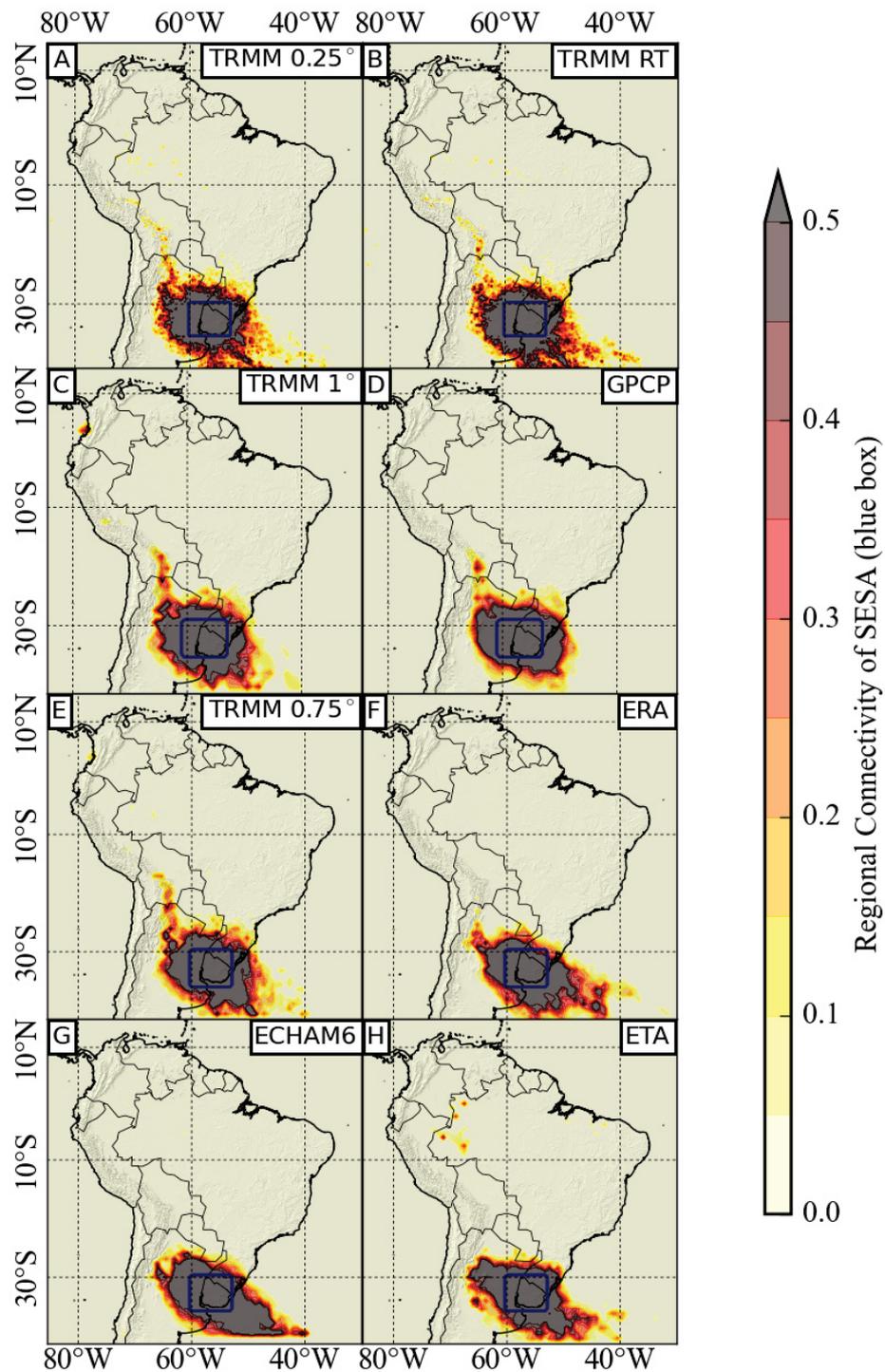
For the ETA model, events in the central Amazon are less synchronized to the eastern parts of the basin as compared to TRMM, but still present. No connectivity to the eastern slopes of the Andes and the adjacent lowlands can be observed. Thus, the ETA model reasonably resembles the connectivity of the central Amazon to the eastern and northern parts of the basin, but it does not show any connectivity to the eastern slopes of the Peruvian and Bolivian Andes. This suggests that the impact of the low-level flow from the Amazon Basin towards the Andes and relevant orographic lifting effects on extreme rainfall are not correctly implemented in this model.

**SESA.** The regional connectivity of SESA (Figure 8.13) is represented quite similarly by all six datasets. The main discrepancy is that only TRMM and TRMM RT show a connection to the eastern slopes of the southern central Andes in northwestern Argentina and southern Bolivia. This signature can be assigned to a certain class of MCS, which form over SESA and then migrate towards the Bolivian Andes as described in the previous chapter 7. While the GPCP dataset seems to incorporate them (although less concise), none of the remaining three datasets yields a signature corresponding to these special MCS.

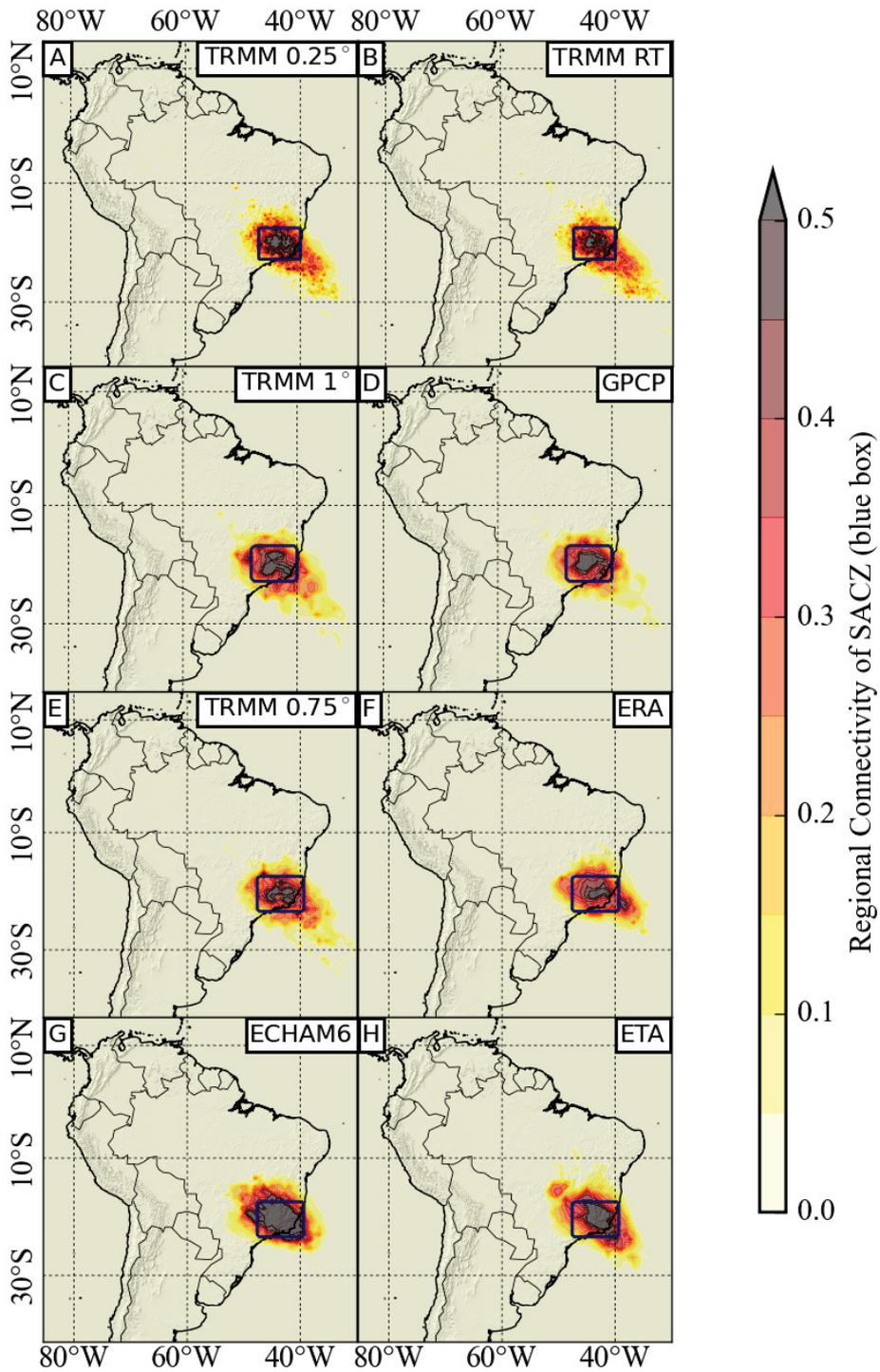
**SACZ.** The SACZ region is connected only locally with very few links pointing out of it for TRMM (Figure 8.14). All five remaining datasets quite accurately reproduce this rather localized signature.



**Figure 8.12.:** Network measure regional connectivity for the central Amazon Basin (blue box) calculated for the 90th percentile of all DJF seasons for TRMM 3B42 V7 at 0.25° (A), 1° (C), and 0.75° (E) spatial resolution, as well as for TRMM 3B42 V7 RT (B), GPCP 1DD V1.2 (D), ERA-interim (F), ECHAM6 (G), and ETA (H).



**Figure 8.13.:** Network measure regional connectivity for southeastern South America (SESA, blue box) calculated for the 90th percentile of all DJF seasons for TRMM 3B42 V7 at 0.25° (A), 1° (C), and 0.75° (E) spatial resolution, as well as for TRMM 3B42 V7 RT (B), GPCP 1DD V1.2 (D), ERA-interim (F), ECHAM6 (G), and ETA (H).



**Figure 8.14.:** Network measure regional connectivity for the South Atlantic Convergence Zone (SACZ, blue box) calculated for the 90th percentile of all DJF seasons for TRMM 3B42 V7 at 0.25° (A), 1° (C), and 0.75° (E) spatial resolution, as well as for TRMM 3B42 V7 RT (B), GPCP 1DD V1.2 (D), ERA-interim (F), ECHAM6 (G), and ETA (H).

### Spatial averages of network measures

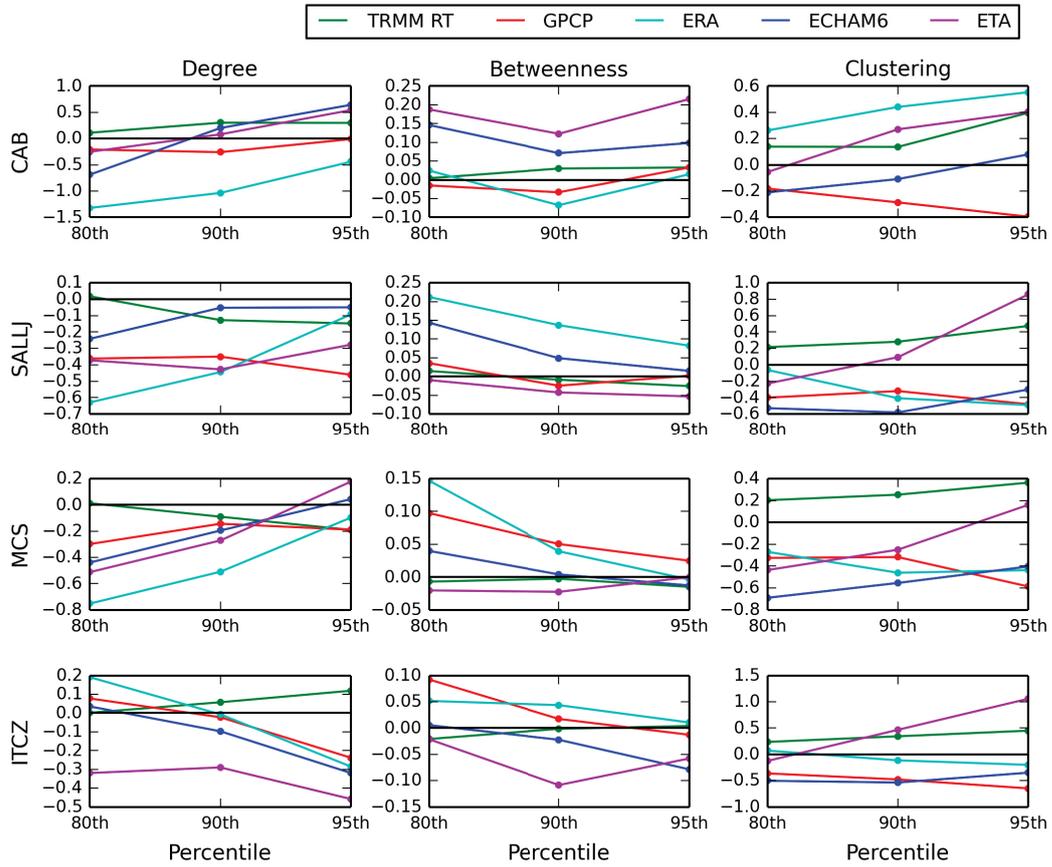
The four regions indicated as CAB, SALLJ, MCS, and ITCZ in Figure 8.1 are of particular relevance for some key features of the SAMS: As explained in section 1.2, the ITCZ, the Amazon Basin, as well as the area east of the Central Andes in the vicinity of the SALLJ play important roles for the large-scale propagation of extreme events. On the other hand, the MCS region is frequently exposed to exceptionally large convective complexes, contributing large fractions of total seasonal rainfall. In addition to the pattern analysis carried out in the previous sections, we therefore compute spatial averages of the network measures DG, BC, and CC for these four regions (Figure 8.15), computed for networks derived from events above the 80th, 90th, respectively 95th percentiles. As above, we shall consider TRMM as reference, and discuss deviations from the values found for TRMM for the remaining five datasets.

**CAB.** Due to its central role for the propagation of extreme rainfall over the entire continent, the CAB is expected to exhibit high DG and BC (first row in Figure 8.15). For TRMM RT and GPCP, the local connectivity (as expressed by DG) as well as the long-ranged centrality (as expressed by BC) of the CAB are reproduced well for 80th, 90th, and 95th percentile events. For ERA, DG is strongly suppressed for all three event thresholds, but BC is reproduced reasonably well. Both ECHAM6 and ETA show too low DG for 80th percentile events, and too high DG for 95th percentile events, while DG for 90th percentile events is accurate. BC of the latter two datasets is too high for all three event types.

The crucial role of the Amazon Basin for the distribution and propagation of extreme rainfall is thus captured by all datasets except ERA, with the ETA and ECHAM6 models even overestimating the long-ranged centrality as expressed by BC.

**SALLJ.** Both the local connectivity (DG) and the long-range centrality (BC) of extreme events in the SALLJ region are accurately represented by the TRMM RT dataset. GPCP strongly underestimates DG for all three event types, possibly due to lack of station data at the Andean slopes. Interestingly, however, BC for the GPCP data is in good agreement with TRMM, indicating that the long-range character of extreme event propagation in this region is still captured well, despite the shortcomings in local connectivity. For the most extreme events, ERA performs well with respect to the SALLJ, but for 80th and 90th percentile events, DG is strongly underestimated. In contrast, BC is substantially higher than for TRMM for all three event thresholds. ECHAM6 shows too low DG and too high BC for events above the 80th percentile, but both measures are in good agreement with TRMM for higher percentiles. The ETA model strongly underestimates DG in the SALLJ region, and also BC is too low for all three percentiles.

Together with the spatial patterns of DG and BC described above, our results indicate that the influence of the SALLJ on the propagation of extreme rainfall, as well as associated orographic lifting effects, are neither adequately represented by the station-based GPCP, nor by the model-derived ERA and ETA datasets. In particular for the ETA model, this is somewhat surprising, since the different vertical coordinate



**Figure 8.15.:** Differences of network measures degree, betweenness centrality, and clustering to the TRMM reference dataset for the spatial boxes Amazon, SALLJ, MCS, and ITCZ. (see Figure 8.1 for definitions of boxes). Values are shown for rainfall events above the 80th, 90th, and 95th percentiles respectively.

scheme is expected to improve the representation of rainfall over complex and steeply ascending terrain. As noted above, however, this only yields a better representation of mean daily rainfall when compared to ERA and ECHAM6, while the tail of the distribution at the eastern Andean slopes is underestimated by ETA. In view of the shortcomings with respect to DG and BC, we suggest that ETA does not only have problems with the magnitudes, but also with the synchronization structure of these events.

**MCS.** As mentioned in section 1.2, the region labeled MCS in Figure 8.1 is exposed to exceptionally large Mesoscale Convective Systems. According to our interpretation of CC, this should lead to high values of this measure.

The TRMM RT dataset exhibits too high CC values over the MCS region for all three percentile thresholds. While the absolute values of CC are too low for GPCP, its spatial pattern described above (Figure 8.11) does exhibit relatively high values over the MCS region. In contrast, ERA, and ECHAM6 underestimate the clustering of events above all three percentile thresholds in this region. This suggests that these models fail to reproduce the clustering in the spatial synchronization structure caused by Mesoscale Convective Systems in this area. The problems of global and regional models to implement the amplitude and spatiotemporal patterns of large thunderstorms in this region are well-known (Solman et al., 2007; Chou et al., 2011; Solman et al., 2013). It should therefore be emphasized that, while it underestimates CC for events above the 80th and 90th percentiles, the ETA dataset is the only one for which we observe CC values comparable to TRMM for the most extreme events (above the 95th percentile).

**ITCZ.** The ITCZ plays an important role as source of extreme rainfall on the South American continent, with emphasis on long-ranged propagation. We therefore expect it to exhibit high BC, as is found for TRMM.

For the TRMM RT data, BC over the ITCZ region is in excellent agreement with TRMM. GPCP overestimates BC for events above the 80th percentile but also yields accurate values for events above the 90th and 95th percentiles. ERA overestimates BC over the ITCZ for all three event thresholds, while ECHAM6 is accurate for events above the 80th and 90th percentiles. The ETA model strongly underestimates the ITCZ's role for the long-range connectivity of all three event thresholds.

## 8.6. Conclusion

The remote-sensing derived, gauge-calibrated TRMM 3B42 V7 has been previously found to provide a reasonable dataset for delineating South American rainfall (Carvalho et al., 2012), in particular with respect to the spatiotemporal patterns of extreme events (see chapter 4). In this chapter, we used TRMM 3B42 V7 as a reference data set, to which we compared the real-time satellite-product TRMM 3B42 V7 RT, the remote-sensing and station-based product GPCP 1DD V1.2, the model-derived reanalysis product ERA-interim, as well as data output from the global

circulation model ECHAM6 and the regional climate model ETA, which is driven by ERA-interim.

We performed two comparisons: First, we evaluate dataset differences with a classical statistical approach; second, we rely on complex network theory to analyze spatial patterns of extreme event co-variability in the six datasets.

In the first part, we discussed classical statistical values. Overall best agreement with TRMM is found for the TRMM RT and the GPCP product. This is little surprising, because GPCP partly uses the same calibration scheme as TRMM. Most notably in this context, while mean daily rainfall values agree well with some discrepancies over the Amazon Basin and the South Atlantic Convergence Zone, all three model-derived datasets underestimate the scores of the 90th and 95th percentiles when compared to TRMM 3B42 V7. Depending on the geographic area and the specific dataset, percentile scores remain in good agreement with TRMM up to the 65th to 85th percentile. Above these percentiles, ERA-interim, the ETA, but also the ECHAM6 model (in the South Atlantic Convergence Zone) underestimate the frequency of extreme rainfall events. The right-hand tail of the daily rainfall distribution, representing pronounced low-frequency but high-magnitude events, is thus not reproduced well by any of the model-derived datasets.

In the second part, we have applied a recently introduced methodology based on complex network theory to analyze the spatial characteristics of extreme rainfall synchronicity.

We have employed several complex network measures in order to quantify different aspects of this internal spatial structure of extreme rainfall synchronicity and found substantial differences between the six datasets analyzed: Our results suggests that TRMM RT's and GPCP's representation of the large-scale propagation patterns are the ones closest to the patterns found by TRMM. We observe that the model-derived ERA-interim reanalysis data do not correctly reproduce key features of the South American monsoon system from the perspective of local synchronizations (degree). These include the effects of deep convection over the Amazon Basin or the orographic barrier of the eastern Andes and their role for large scale moisture transport. However, the ERA-interim dataset does incorporate these features accurately when taking into account long-ranged connections over several steps in the network (betweenness centrality). Still, the global model ECHAM6 and the regional model ETA perform better in reproducing these features. The propagation pathway from the western Amazon towards the subtropics along the eastern slopes of the Andes, which is associated with the South American Low-Level Jet, is well represented by the ECHAM6 model, while the ETA model fails to produce this feature.

Concerning large convective systems, in contrast to TRMM and TRMM RT, ERA as well as ECHAM6 fail to reproduce any spatially organized rainfall clusters in regions where they are known to frequently occur, such as in southeastern South America. ETA exhibits some spatial patterns possibly corresponding to large organized systems, but also not in this specific area where these would be most relevant. While the overall spatial organization for GPCP differs considerably from TRMM, Mesoscale Convective Systems in southeastern South America are detected reasonably well.

Linkages of synchronous extreme rainfall from the central Amazon Basin to the eastern slopes of the Andes observed for TRMM and TRMM RT are only resembled by the ECHAM6 model, but not by the ERA, ETA, and GPCP datasets.

Regarding teleconnections from two regions important for the aforementioned rainfall dipole (southeastern South America and the South Atlantic Convergence Zone), all six datasets coherently find a rather localized pattern with little connections to other regions. However, a linkage from southeastern South America to the eastern slopes of the southern central Andes is only present for the TRMM, TRMM RT, and the GPCP datasets.

## Chapter 9.

# Impacts of the El Niño Southern Oscillation on Extreme Moisture Divergence

### 9.1. Summary

The purpose of this chapter is to investigate the temporal evolution of moisture divergence and its spatial clustering properties over South America, with particular focus on dependencies on the phase of the El Niño Southern Oscillation (ENSO). Moisture divergence is computed from daily reanalysis data of vertically integrated moisture flux. A sliding-window approach is used to construct a sequence of complex networks, each obtained from synchronization of events of strong positive (negative) moisture divergence, which we interpret as strong evapotranspiration (precipitation) events. The main results can be summarized as follows: i) Moisture divergence values over the Amazon rainforest are typically higher during positive ENSO periods (El Niño events). ii) The spatial coherence of strong positive (upwelling) events assumes a characteristic pattern of reduced coherence in this area during El Niño conditions. This influence of ENSO on moisture divergence and its spatial coherence is dominated by the El Niño events of 1982, 1987, and 1997. iii) The clustering characteristics of the obtained climate networks qualitatively agree with the spatial distribution of connected regions with simultaneous events (i.e., events that occur at the same time), but provide a more detailed view on the spatial organization of strong atmospheric upwelling events. Interestingly, no comparable results are found for negative extremes of moisture divergence (strong precipitation events). This chapter is based on the associated publication P7, and some of its sections will closely follow the presentation in this publication.

### 9.2. Introduction

Recent climate network studies have provided novel insights into the global impact and spatiotemporal organization of the El Niño Southern Oscillation (ENSO) (Yamasaki et al., 2008; Tsonis and Swanson, 2008). In particular, this approach has led to a novel strategy for anticipating positive ENSO (El Niño) events (Ludescher et al.,

2013) as well as discriminating between different types of positive and possibly also negative ENSO (La Niña) events (Radebach et al., 2013).

As described in chapter 4, the spatial patterns exhibited by various complex network measures based on the temporal synchronization between extreme rainfall events reveal the most important features of the South American Monsoon System. These include the main moisture pathways, their interplay with orography (orographic rainfall, (Bookhagen and Strecker, 2008)), the main convergence zones, and areas with frequent thunderstorm development. Specifically, it was documented that frequent occurrences of squall lines (Garreaud and Wallace, 1997) and Mesoscale Convective System (Durkee et al., 2009; Durkee and Mote, 2009) lead to high values of the local clustering coefficient in the resulting networks. Since the probability of two grid cells to exhibit a high degree of interdependence typically decays with their distance (Donges et al., 2009b; Radebach et al., 2013), high local clustering coefficients in networks based on the synchronization of events indicate that extreme events in the corresponding region occur in a spatially coherent manner.

In this chapter, we analyze strong positive and negative values (above the 90th percentile and below the 10th percentile) of the daily divergence of vertically integrated moisture flux. Positive moisture divergence, in simple terms, can be viewed as upward moisture transport from the surface to the atmosphere, for example caused by evapotranspiration; in contrast, negative moisture divergence corresponds to downward movement of moisture from the atmosphere to the surface in form of precipitation. We investigate the dependence of the associated spatial patterns on the ENSO phase, with a focus on the spatial coherence and large-scale organization of events. For this purpose, we use a sliding-window approach and construct networks based on symmetric event synchronization ( $ES^{sym}$ , cf. equation (2.4)) for time intervals of 365 days width in steps of 60 days. We analyze the temporal evolution of local and global clustering coefficients of these networks, and compare it to ENSO variability.

For comparison and in order to look at the resulting spatial patterns from a different viewpoint, we also compute the average size of connected regions of simultaneous events. Here, *simultaneous* means that events occur at the very same day, as opposed to *synchronized*, where we here allow for a maximum delay of  $\pm 5$  days between associated events. This approach provides a rather intuitive and direct way to analyze the spatial coherence of strong vertical moisture fluxes and adds quantitative information on the typical spatial extensions of simultaneous events in a given region. It can be understood as a more traditional approach to analyze spatial coherence, complementary to the climate network approach.

### 9.3. Data

We employ daily data of vertically integrated moisture flux from NASA’s Modern-Era Retrospective Analysis for Research and Applications (MERRA, see (Rienecker et al., 2011)), with temporal coverage from 1979 to 2010 for the region between 15°N

and 40°S, and 30°W and 85°W, at a latitudinal resolution of 1/2° and longitudinal resolution of 2/3°.

Divergence of this flux was calculated using finite differences. Denoting vertically integrated atmospheric moisture content (i.e., precipitable water) by  $A$ , divergence of vertically integrated moisture flux by  $M$ , precipitation by  $P$ , and evapotranspiration by  $E$ , the water balance equation reads

$$\partial_t A + M = E - P. \quad (9.1)$$

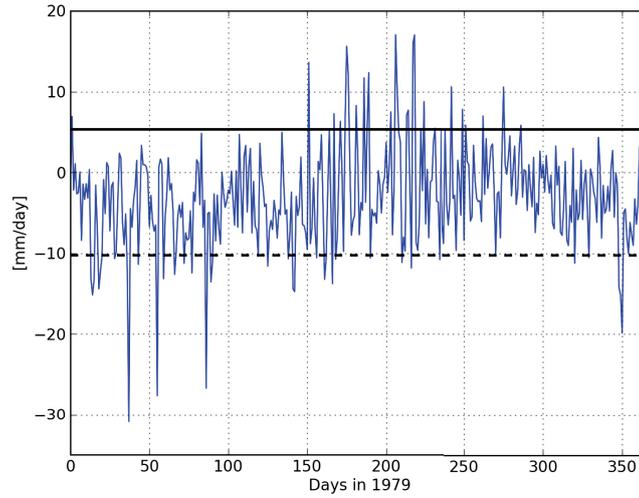
We emphasize that  $E$  and  $P$  are never negative. A positive extreme of  $M$  in a given grid cell can in principle be caused by two different effects: A sudden decrease of moisture inflow from neighboring grid cells, or an extreme event of  $E$ . In the former case, the neighboring grid cells cannot have a positive extreme event of  $M$  at the same time step, while in the latter case extreme events should typically occur spatially homogeneously. Therefore, if positive extreme events of  $M$  occur in a spatially homogeneous way, i.e. as part of large connected components of simultaneous positive extreme events, the former effect can be dismissed. As we will demonstrate in the course of this work (see Figure 9.11 below), positive extreme events of  $M$  occur in fact spatially homogeneously, which allows to interpret these events as extreme events of  $E$ . On the other hand, by the same rationale, negative extrema of  $M$  typically lead to extreme events of  $P$ .

While spatial clustering of strong rainfall events in South America has a straightforward interpretation in terms of thunderstorms and large convective systems and has been discussed in chapter 4 and the associated publication P1, the spatial coherence of strong evapotranspiration events has not been discussed in the literature so far. This may be due to the fact that a climatological interpretation is not as simple as in the case of precipitation, since evapotranspiration is considered to be controlled by regional and local conditions. However, factors influencing evapotranspiration rates, such as solar radiation, temperature, and wind, can indeed act coherently on rather large spatial scales and thus lead to interrelations of evapotranspiration time series at remote locations. We observe that evapotranspiration exhibits an event-like structure, i.e., short time periods during which values are much higher than during the remaining times (Fig. 9.1). A particular focus of the following analysis will be on the Amazon Basin. In this area, evapotranspiration by dense vegetation cover and high biomass amounts – in combination with southwestward and westward propagating convective storm systems from the Atlantic Ocean – yield an important contribution to overall positive moisture divergence (Lean and Warrilow, 1989; Shukla et al., 1990; Eltahir and Bras, 1993).

For the ENSO signal, the monthly multivariate ENSO index (MEI<sup>5</sup>, cf. (Wolter and Timlin, 1993; Wolter and Timlin, 1998)) is used. It is based on the six variables sea-level pressure, zonal and meridional surface wind components, sea and air surface temperature, and total cloud fraction. MEI is therefore considered a more reliable

---

<sup>5</sup>retrieved from <https://climatedataguide.ucar.edu/climate-data/multivariate-enso-index>



**Figure 9.1.:** Example of a daily time series of moisture divergence for 1979 at  $5^{\circ}S$ ,  $60^{\circ}W$ , and the corresponding thresholds at the 10th (dashed) and the 90th percentile (solid). Positive values indicate net evapotranspiration, while negative values indicate net precipitation.

estimator of the ENSO state than indices like NINO 3.4 (based solely on sea surface temperatures) or SOI (based on sea-level pressure) (Rasmusson and Carpenter, 1982; Trenberth, 1997; Trenberth and Stepaniak, 2001). Time intervals for which MEI is larger than +1 (smaller than  $-1$ ) are considered to be warm, i.e. El Niño (cold, i.e. La Niña) episodes, while the remaining periods will be referred to as neutral.

## 9.4. Methods

### 9.4.1. Network construction

In order to investigate the temporal evolution of the clustering characteristics of the obtained networks, we choose a sliding window approach. We construct networks from time slices with a length of 365 days in steps of 60 days ( $\sim 2$  months), resulting in 195 time steps for the entire period between 1979 and 2010. Out of these, 38 fall into El Niño phases and 21 into La Niña phases. We first transform the data for each of these time slices and at each grid point into an evapotranspiration (precipitation) event series by considering those days as events for which  $M$  is among the highest (lowest) 10% of all values for that location and time interval, resulting in 36 events for each time slice. We note that, since the percentiles are computed for each grid cell and time step, the corresponding event thresholds vary in space and time.

In order to construct a network from the obtained event series, we employ symmetric Event Synchronization ( $ES^{sym}$ ) to decide whether or not two time series (i.e., network

nodes) are sufficiently similar to be connected by a network link. In this case, we consider  $l = 36$  events for each grid point and each time interval. To exclude unreasonably long delays between events at different locations, we here choose a maximum delay of  $\tau_{max} = 5$  days.

A network is obtained by thresholding  $ES^{sym}$  at the 95th percentile, resulting in a link density of 5%. The network's adjacency matrix is thus given by

$$A_{ij} = \Theta(ES_{ij}^{sym} - T^{0.95}) - \delta_{ij}, \quad (9.2)$$

where it is recalled that  $\Theta$  is the Heaviside “function”,  $T^{0.95}$  denotes the 95th percentile of all entries of  $\mathbf{ES}^{sym}$  for a given time window, and Kronecker's  $\delta$  appears in order to exclude self loops. In this way, we construct  $t = 195$  networks for time spans of 365 days, in steps of 60 days. We find that all event synchronization values that are represented by network links through this procedure are significant ( $p$ -value  $< 0.05$ ) with respect to a null model based on uniformly placing 36 events (highest respectively lowest 10%) in a time series of 365 days at random.

#### 9.4.2. Network analysis

By construction, clustering characteristics of synchronized evapotranspiration and precipitation events are encoded in the topology of the networks derived in the manner described above. As outlined in chapter 3, topological properties of networks can be quantified by suitable network measures. The first measure we consider here is the local clustering coefficient (LC):

$$LC_i := \frac{\sum_{j < k} A_{ij} A_{jk} A_{ik}}{\sum_{j < k} A_{ij} A_{ik}}. \quad (9.3)$$

This measure is identical with the clustering coefficient (CC) defined in section 3.4.1. However, we repeated its definition here and renamed it to *local* clustering coefficient in order to distinguish it from the *global* clustering coefficient defined below.  $LC_i$  gives the relative frequency of network neighbors of  $i$  that are themselves connected. Since the probability of a network link between two grid cells decays with their geographical distance (Donges et al., 2009b), a connected region with high local clustering coefficients will be a region of large spatial coherence: Strong events in regions of continuously high local clustering coefficients are expected to typically all occur close in time (within  $\tau_{max} = 5$  days), while low local clustering coefficients indicate more erratic and spatially incoherent behavior. This is in agreement with the results of chapter 4 and the associated publication P1.

The global clustering coefficient (GC) is defined as the arithmetic mean of local clustering coefficients, taken over the entire network:

$$GC := \frac{1}{N} \sum_{i=1}^N \frac{\sum_{j < k} A_{ij} A_{jk} A_{ik}}{\sum_{j < k} A_{ij} A_{ik}} = \frac{1}{N} \sum_{i=1}^N LC_i, \quad (9.4)$$

where  $N$  is the number of nodes. Therefore, it gives an estimate of the overall tendency towards spatially coherent evapotranspiration or precipitation regimes.

### 9.4.3. Connected regions of simultaneous extremes

Complementarily to the network approach, we also investigate the spatial patterns of the average size (measured as the number of grid cells) of spatially connected regions of simultaneous events (SC). Here, two grid cells are considered spatially connected if they are longitudinally, latitudinally, or diagonally adjacent in space. We first look for each day  $\mu$  in a given time window (of length 365 days) for the  $N_\mu$  spatially connected components  $\{\mathcal{C}_m^\mu\}_{1 \leq m \leq N_\mu}$  of grid cells at which events occur at that day. Then, we assign to each grid cell the value of the size of the component it belongs to:

$$S_i^\mu = \sum_{\{\mathcal{C}_m^\mu\}} \delta_{\mathcal{C}_m^\mu}(i) |\mathcal{C}_m^\mu|, \quad (9.5)$$

where  $|\mathcal{C}_m^\mu|$  denotes the cardinality of  $\mathcal{C}_m^\mu$  and  $\delta_{\mathcal{C}_m^\mu}(i) = 1$  if  $i \in \mathcal{C}_m^\mu$  and  $\delta_{\mathcal{C}_m^\mu}(i) = 0$  otherwise. Finally, we average  $S_i^\mu$  over all days  $\mu$  for which  $S_i^\mu > 0$ ,

$$SC_i = \frac{1}{t'} \sum_{\{\mu | S_i^\mu \neq 0\}} S_i^\mu, \quad (9.6)$$

where in our case  $t' = 36$  days. Since the actual size of horizontal grid cells depends on the latitudinal position, we correct the values of SC by weighting them with  $\cos(\lambda)$ , where  $\lambda$  denotes the latitudinal angle ranging from 40°S to 15°N.

### 9.4.4. Pattern analysis

The time evolution of the network's global clustering properties (GC) can be directly compared to ENSO variability. For LC, we obtain a vector of dimension  $N$  (the number of grid points) for each time interval. But apart from the temporal evolution of the overall mean of this vector (i.e., GC), we are interested in a more detailed analysis of how the spatial patterns change during time evolution. For this purpose, we first compute Spearman's rank correlation coefficients between the ENSO signal and the time series of LC at each location.

Complementarily, we investigate the time evolution of spatial patterns of LC by means of pattern similarity analysis. For this purpose, we first compute the *spatial ranks* RLC of all LC values for each time interval: For each time step  $\mu$ , we have an  $N$ -dimensional vector  $LC^\mu$ , containing the values  $LC_i^\mu$  of the local clustering coefficient for the  $N$  geographical positions  $i$ . This array can be sorted with respect to these values. By "spatial ranks" ( $RLC$ ), we refer to the position  $RLC_i^\mu$  (a number between 1 and  $N$ ) of a given entry  $LC_i^\mu$  in that sorted array. We then calculate the

$L^1$ -distance between the respective vectors  $RLC^\mu$  and  $RLC^\nu$  obtained for all pairs of time intervals  $(\mu, \nu)$ :

$$L^{\mu\nu} = \sum_{i=1}^N |RLC_i^\mu - RLC_i^\nu| \quad (9.7)$$

with space indices  $1 \leq i \leq N$  for the number of grid cells (i.e., network nodes),  $N = 9324$ , and time indices  $1 \leq \mu, \nu \leq t$  with  $t = 195$ . The column- (or row-) wise mean,

$$ML^\mu = \frac{1}{t} \sum_{\nu=1}^t L^{\mu\nu} \quad (9.8)$$

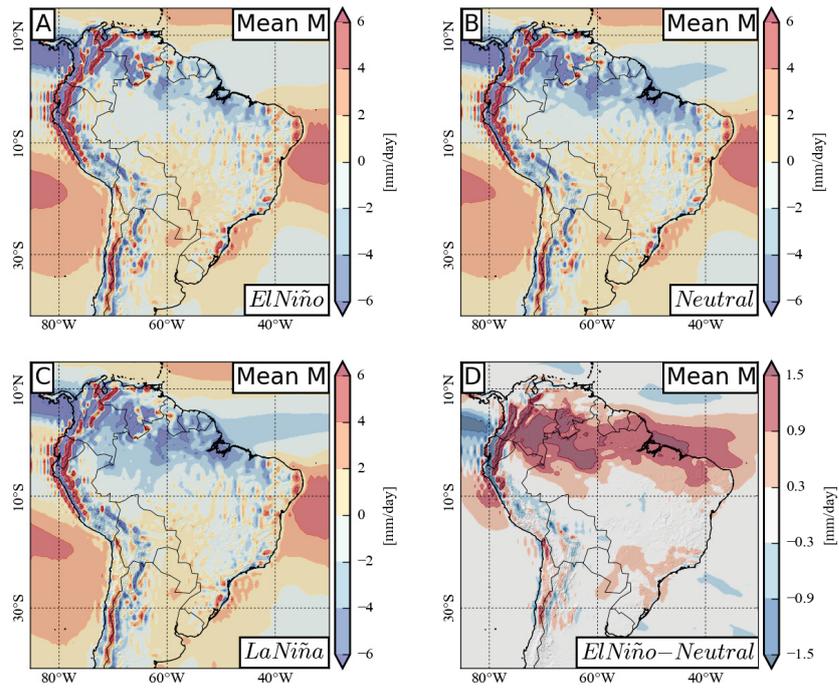
of the symmetric distance matrix  $\mathbf{L} = (L^{\mu\nu})_{1 \leq \mu, \nu \leq t}$  provides information on the dissimilarity of a spatial pattern observed in a given time interval in comparison with the patterns attained during all other intervals. Low values indicate that the spatial patterns assume some characteristic structure at these times. Calculating  $ML$  for all times results in another time series that can be compared to ENSO variability.

Finally, for LC and SC, we construct composites, i.e. we average the values at each location separately for warm (El Niño), neutral and cold (La Niña) stages of ENSO in order to obtain typical spatial patterns of these two measures for the three ENSO phases.

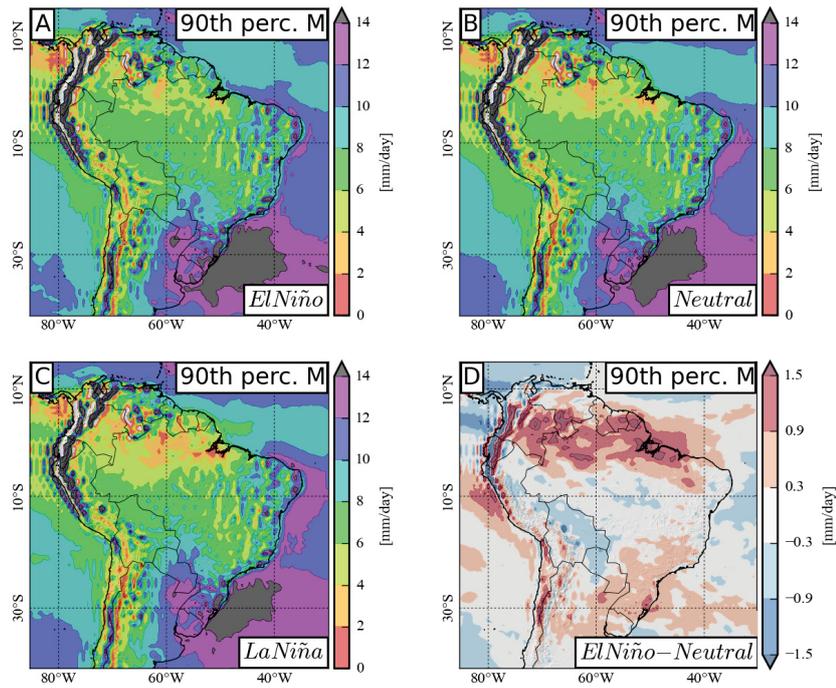
## 9.5. Results

### 9.5.1. Spatial patterns of moisture divergence

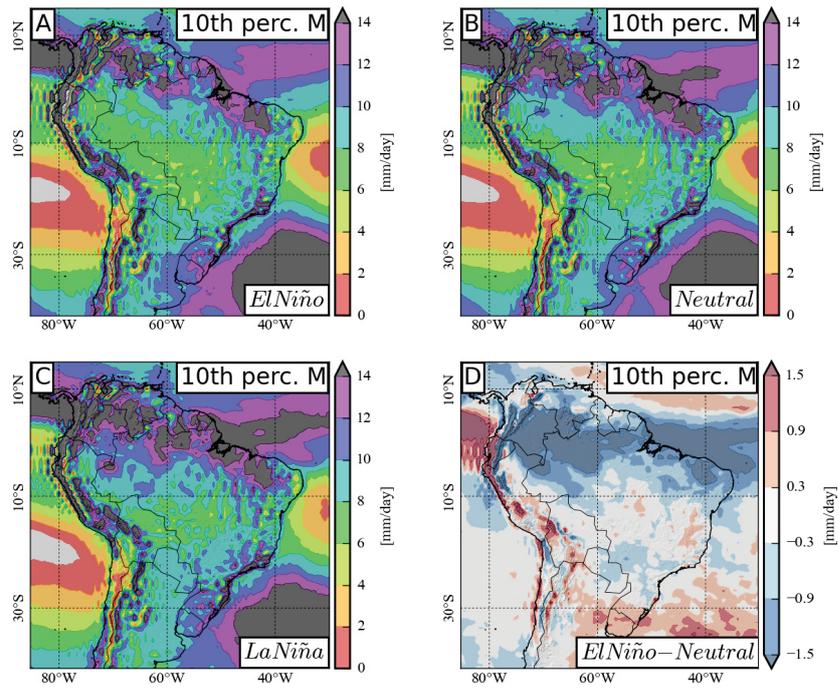
Regarding composites (i.e., spatial fields averaged over the respective time periods), we find that the strongest influence of ENSO on mean daily values of  $M$  (Fig. 9.2), as well as on the thresholds defining strong evapotranspiration (above the 90th percentile, Fig. 9.3) and precipitation (below the 10th percentile, Fig. 9.4) events appears in a belt over the eastern and northern Amazon Basin and the adjacent coastal regions near the equator. Mean as well as strong evapotranspiration event thresholds increase in these regions during El Niño conditions, while strong precipitation event thresholds show the opposite behavior: values decrease over the tropical Atlantic Ocean and the entire Amazon Basin. The La Plata Basin exhibits higher 90th percentiles and lower (i.e., more pronounced negative) 10th percentiles during El Niño conditions, whereas average values are not markedly different in this area during the different ENSO phases. For a more detailed quantification of the impact of ENSO on  $M$ , we employ Spearman's rank correlation coefficient (SR) to analyze the interdependence between ENSO based on the MEI and each local value of mean, 90th (strong evapotranspiration) and 10th (strong precipitation) percentiles of  $M$ . We find correlation values between +0.3 and +0.5 over the Amazon Basin and the tropical Atlantic Ocean close to the Brazilian coast for both mean values



**Figure 9.2.:** Composites of mean daily moisture divergence  $M$  for warm (A), neutral (B), and cold (C) ENSO conditions, and difference of composites for warm and neutral ENSO conditions (D). Note the increased values in northern South America and along the equator during El Niño conditions.



**Figure 9.3.:** Composites of 90th percentiles of moisture divergence  $M$  (corresponding to strong evapotranspiration events) for warm (A), neutral (B), and cold (C) ENSO conditions, and difference of composites for warm and neutral ENSO conditions (D). Note the increased values in northern South America and along the equator during El Niño conditions.



**Figure 9.4.:** Composites of 10th percentiles of moisture divergence  $M$  (corresponding to strong precipitation events) for warm (A), neutral (B), and cold (C) ENSO conditions, and difference of composites for warm and neutral ENSO conditions (D). Note the decreased values in northern South America and along the equator during El Niño conditions.

(Fig. 9.5A) and strong evapotranspiration events (Fig. 9.5B). In contrast, thresholds for strong precipitation events are negatively correlated with MEI (between  $-0.3$  and  $-0.5$ ) in this area (Fig. 9.5C). Correlations between MEI and daily means of  $M$  are positive throughout eastern Brazil and negative in Uruguay and northeastern Argentina ( $58^\circ\text{W}$ ,  $32^\circ\text{S}$ ). In contrast, for strong evapotranspiration event thresholds, there are negative correlations in easternmost Brazil, while positive correlations reach from southeastern Brazil to northern Argentina, covering the entire La Plata Basin. Correlations between MEI and the 10th percentiles are positive with values between  $+0.3$  and  $+0.5$  in the southern La Plata basin.

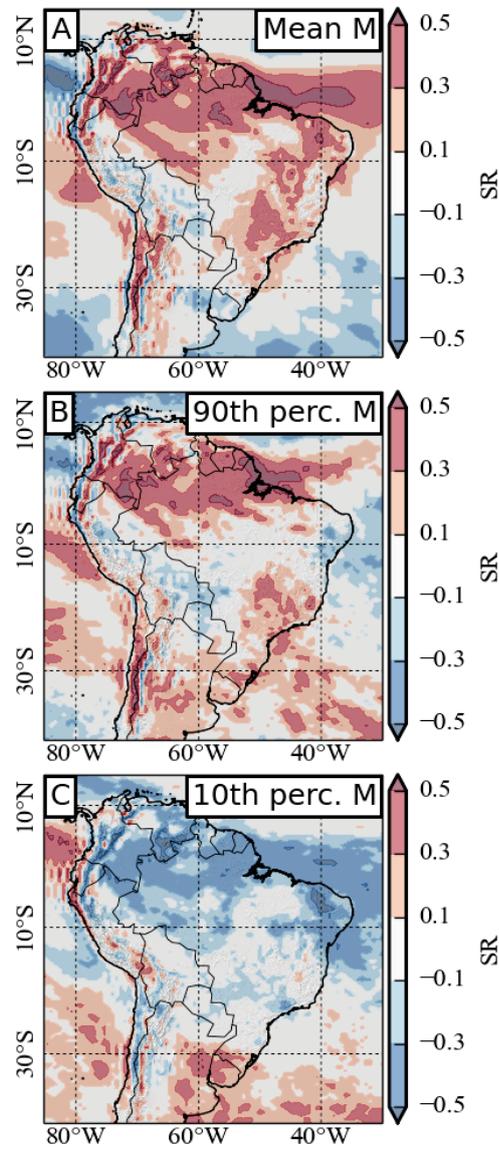
In addition, we observe positive correlation values in the hyper-arid Atacama desert in northern Chile as well as on the adjacent Puna Plateau in northwestern Chile and southern Bolivia for both daily means and strong evapotranspiration event thresholds. In contrast, strong precipitation event thresholds are negatively correlated with ENSO in these regions.

### 9.5.2. Clustering of extreme moisture divergence: time-dependence

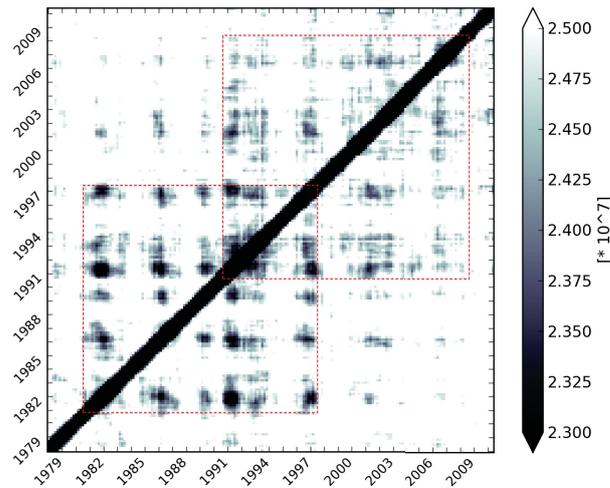
For strong evapotranspiration events, the distance matrix  $\mathbf{L}$  (Fig. 9.6) obtained from the procedure described above exhibits reduced  $L^1$ -distance values for a variety of time intervals. Most notably, there appear two blocks in  $\mathbf{L}$  which have shorter mean distances among themselves as compared to patterns obtained from other time intervals. These two blocks overlap between 1992 and 1998. The mean of  $\mathbf{L}$  restricted to the time from 1979 to 1998 is  $2.45 \cdot 10^7$  with standard deviation  $0.32 \cdot 10^7$ , and  $2.44 \cdot 10^7$  with standard deviation  $0.32 \cdot 10^7$  when restricted to the time from 1992 to 2012. For comparison, the mean of  $\mathbf{L}$  between these two time periods (i.e., the mean of the block of  $\mathbf{L}$  defined by  $1998 \leq \mu \leq 2012$  and  $1979 \leq \nu \leq 1991$ ) is  $2.57 \cdot 10^7$  with standard deviation  $0.06 \cdot 10^7$ . For all times together, we obtain  $2.50 \cdot 10^7$  for the mean of  $\mathbf{L}$  and  $0.25 \cdot 10^7$  for the standard deviation. For strong precipitation events,  $\mathbf{L}$  does not show such a clear pattern (Fig. 9.7).

From 1979 to 2005, for strong evapotranspiration events the row-wise mean of  $\mathbf{L}$  (ML, Fig. 9.8B) tends to have its minima at times of positive ENSO conditions. Here, minima correspond to the sequence of low values of  $\mathbf{L}$  in Fig. 9.6. The row-wise mean of  $\mathbf{L}$  is anti-correlated with MEI (Fig. 9.8A) with  $\text{SR}(\text{GC}, \text{MEI}) = -0.47$  ( $p$ -value of the order of  $10^{-12}$ ). In contrast, for strong precipitation events, the row-wise mean of  $\mathbf{L}$  (Fig. 9.9B) does not show visual similarity to MEI, corresponding to low and non-significant correlation values.

The global clustering coefficient of the obtained networks responds negatively to MEI, with  $\text{SR}(\text{GC}, \text{MEI}) = -0.48$ , for strong evapotranspiration events (Fig. 9.8C). The corresponding  $p$ -value, obtained from a two-sided  $t$ -test, is of the order of  $10^{-13}$ . For strong precipitation events, no visual interdependence between MEI and GC can be observed (Fig. 9.9C). In accordance, the correlation is much weaker in this case ( $\text{SR}(\text{GC}, \text{MEI}) = -0.15$  with  $p$ -value  $p = 0.04$ ).



**Figure 9.5.:** Spearman's rank correlation coefficient (SR) between the ENSO index MEI and mean (A), 90th percentiles (B), and 10th percentiles (C) of moisture divergence M. Note in particular the positive correlations between mean and 90th percentiles of moisture divergence and MEI in northern South America, as well as the negative correlations in the same area for 10th percentiles.



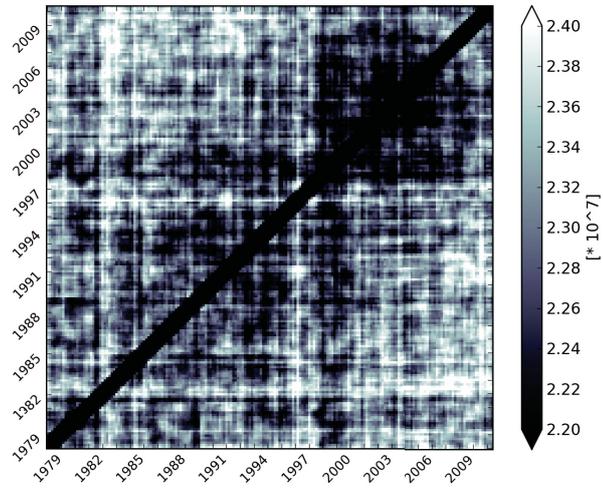
**Figure 9.6.:**  $L^1$ -distance matrix  $\mathbf{L}$  between the ranks of the local clustering coefficients (RLC) obtained for events above the 90th percentile (strong evapotranspiration events) for all time windows. Note the two time periods (1981 to 1998 and 1992 to 2009) with lower distances among themselves and higher distances to the respective other time period. The corresponding two blocks of  $\mathbf{L}$  are indicated by dashed red lines.

Due to the weak interrelations between MEI and clustering properties obtained for strong precipitation events, we will focus on strong evapotranspiration events in the following sections.

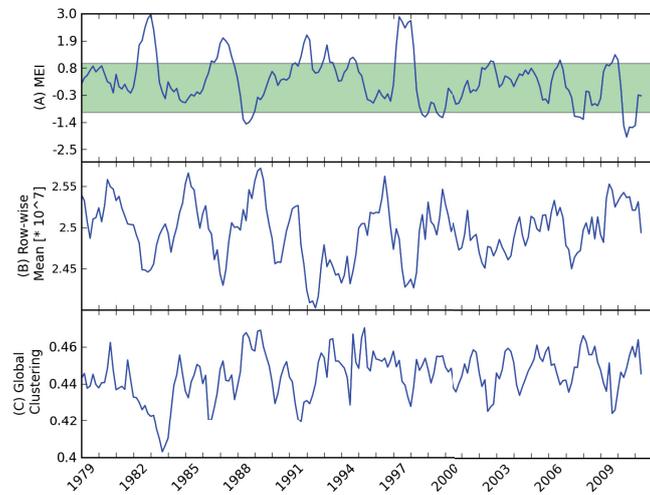
### 9.5.3. Spatial patterns of clustering of extreme evapotranspiration events

Mean composites of LC (Fig. 9.10) show relatively high values over the Pacific Ocean between  $10^\circ\text{S}$  and  $30^\circ\text{S}$ , over the tropical Atlantic Ocean and the adjacent coast of northeastern Brazil, and over most of subtropical South America for all three ENSO stages. Relatively low values can be seen over the central Andes and eastern central Brazil. Comparing the three different ENSO stages reveals that El Niño times are most notably characterized by decreased LC values over the entire Amazon Basin as compared to neutral and La Niña times.

For both positive and negative ENSO phases as well as neutral ENSO conditions, mean composites of the average size of connected regions of simultaneous events (SC, Fig. 9.11) show high values over the subtropical Atlantic Ocean and the adjacent southeastern South American continent. However, during La Niña times, these high values reach farther north towards the western Amazon Basin. In contrast, during El Niño episodes the values of SC decrease in these regions, in particular over the Amazon Basin north of  $10^\circ\text{S}$ , when compared to the neutral ENSO stages.



**Figure 9.7.:**  $L^1$ -distance matrix  $\mathbf{L}$  between the ranks of the local clustering coefficients (RLC) obtained for events below the 10th percentile (strong precipitation events) for all time windows.



**Figure 9.8.:** (A) ENSO index MEI, (B) row-wise mean (ML) of the  $L^1$ -distance matrix  $\mathbf{L}$  between the ranks of the local clustering coefficients (RLC), and (C) time evolution of global clustering (GC) obtained for strong evapotranspiration events.

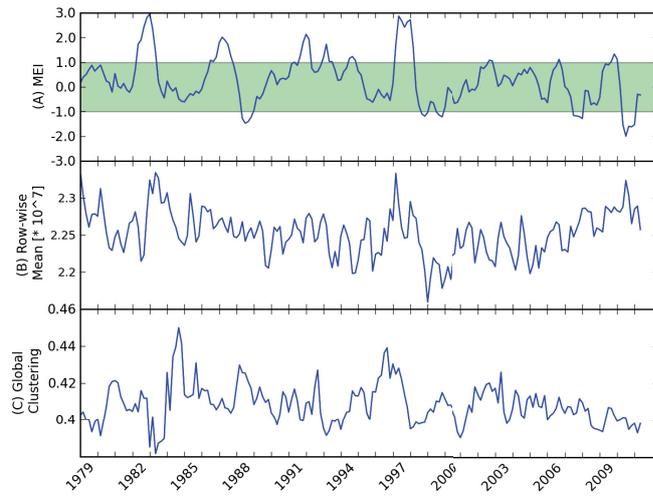


Figure 9.9.: As Fig. 9.8 for strong precipitation events.

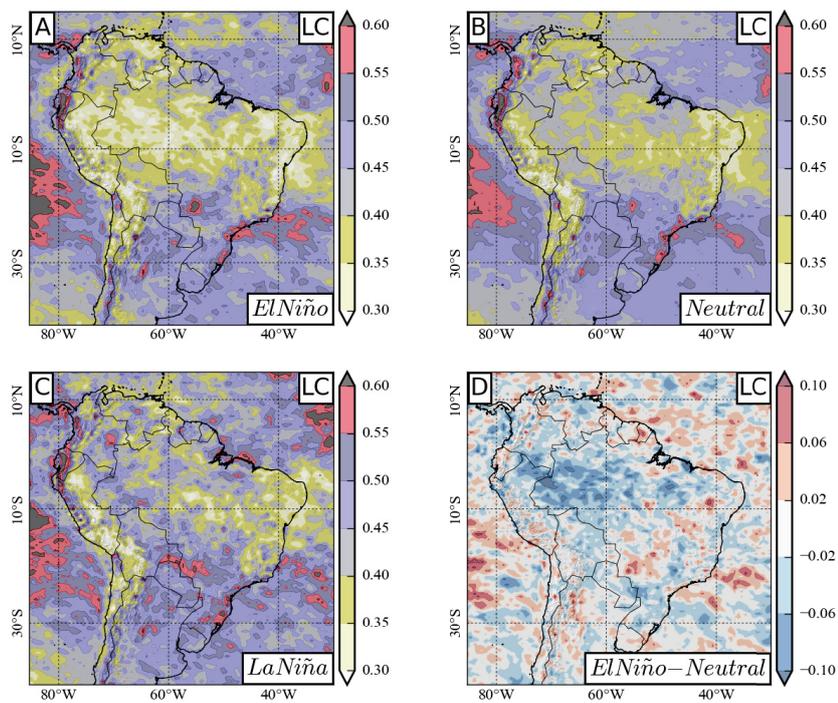
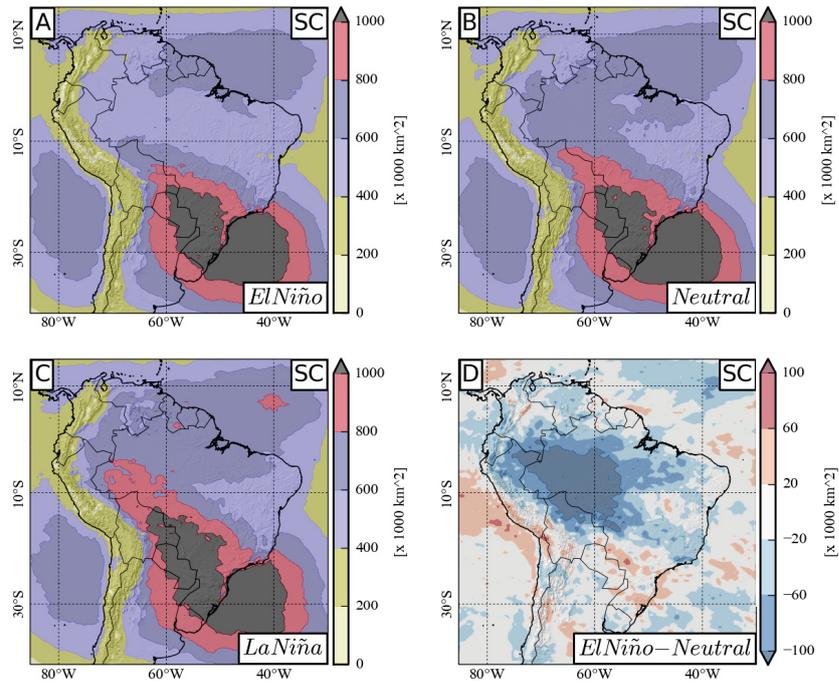


Figure 9.10.: Composites of the local clustering coefficient (LC) based on strong evapotranspiration events for warm (A), neutral (B), and cold (C) ENSO conditions, and difference of composites for warm and neutral ENSO conditions (D). Note the reduced values over the central and western Amazon Basin for El Niño conditions.



**Figure 9.11.:** As in Fig. 9.10 for the average size of connected components of simultaneous extremes (SC). Note the reduced values over the central and western Amazon Basin for El Niño conditions. Because the size of the horizontal grid cells varies depending on the latitudinal position, the values of SC have been corrected by weighting them with  $\cos(\lambda)$ , where  $\lambda$  denotes the latitudinal angle.

## 9.6. Discussion

### 9.6.1. Impacts of ENSO cycles on moisture divergence

We observe that in the eastern and northern Amazon Basin, the distribution of  $M$  is generally shifted towards higher values during El Niño conditions (Figures 9.2 and 9.3). Given the results reported by other studies (Hastenrath and Heller, 1977; Ropelewski and Halpert, 1987; Marengo et al., 2008; Bookhagen and Strecker, 2010), we attribute this to negative precipitation anomalies in the northern and eastern Amazon Basin during warm ENSO phases. In the La Plata Basin, positive as well as negative extremes of  $M$  become more pronounced (Fig. 9.4), which indicates a general increase in variance (implying both stronger precipitation and evapotranspiration extremes) during El Niño events.

Negative correlation values between MEI and  $M$  in the La Plata Basin for daily means as well as positive correlation values for the absolute values of 10th percentile thresholds are associated with positive precipitation anomalies in southeastern South America during El Niño events (Hastenrath and Heller, 1977; Bookhagen and Strecker, 2010; Cazes-Boezio et al., 2003). However, this does not apply to the 90th percentile thresholds of  $M$ , for which, surprisingly, correlations are also positive over Uruguay. Thus, the entire distribution of  $M$  becomes wider during El Niño episodes, with stronger precipitation as well as stronger evapotranspiration events. These results may be relevant in view of the importance of the La Plata Basin concerning agriculture and hydropower generation (e.g. Barros et al., 2006).

Positive correlations between MEI and daily mean and strong evapotranspiration event thresholds of  $M$  as well as negative correlations between MEI and strong precipitation event thresholds in the Atacama Desert agree with earlier studies, which have found positive precipitation anomalies during La Niña episodes (Houston, 2006b) as well as positive evapotranspiration anomalies during El Niño episodes (Houston, 2006a).

### 9.6.2. Impacts of ENSO cycles on clustering of moisture divergence

Since no clear and significant impacts of ENSO on the clustering characteristics of strong precipitation events were found in the MERRA data, we will again focus on strong evapotranspiration events in the following section.

The distance matrix  $\mathbf{L}$  (Fig. 9.6) obtained from the procedure described in Section 9.4 allows to identify times for which the spatial structures of LC resemble a specific characteristic pattern, indicated by low values of  $L^{\mu\nu}$ . Specifically, such low values of appear during El Niño conditions, which is supported by negative correlations between MEI and the row-wise mean of  $\mathbf{L}$ , i.e.  $ML$  (Fig. 9.8). This result indicates that during El Niño periods, LC assumes a characteristic spatial pattern which exhibits lower variability in terms of mutual  $L^1$ -distances than for neutral and La Niña periods. The negative correlation between MEI and GC (Fig. 9.6) implies that during El Niño phases the spatial pattern of synchronized evapotranspiration extremes is characterized by decreased GC. This is interpreted as an overall tendency

towards diminished spatial coherence of strong evapotranspiration events during El Niño conditions. Mean composites of LC for the different ENSO periods (Fig. 9.10) reveal that decreased values over the Amazon Basin are mainly responsible for this decline in GC.

Due to the interpretations of LC and SC (cf. Section 9.4), the results for SC are expected to be similar to those obtained for LC. Neglecting the influence of temporal lags, high values of LC computed from a network based on  $ES^{sym}$  should typically imply high values of SC as well, since the probability of link lengths in networks typically decays roughly exponentially with the spatial distance between two sites (Donges et al., 2009a; Radebach et al., 2013). For the case of surface air temperatures, this is reflected by the emergence of densely connected structures in networks during some time intervals, most remarkably during El Niño and La Niña periods (Tsonis and Swanson, 2008; Paluš et al., 2011; Radebach et al., 2013).

Consistent with these expectations, the mean composites of SC (Fig. 9.11) indeed reveal a reduced average size of connected components of simultaneous evapotranspiration events over the Amazon Basin during warm ENSO phases as compared to neutral and cold periods. The implications are thus qualitatively the same as for LC, although the obtained patterns of SC exhibit less spatial variability and are in this sense less informative than those of LC. However, the consistency between the results of LC and SC provides further support of the drawn conclusions.

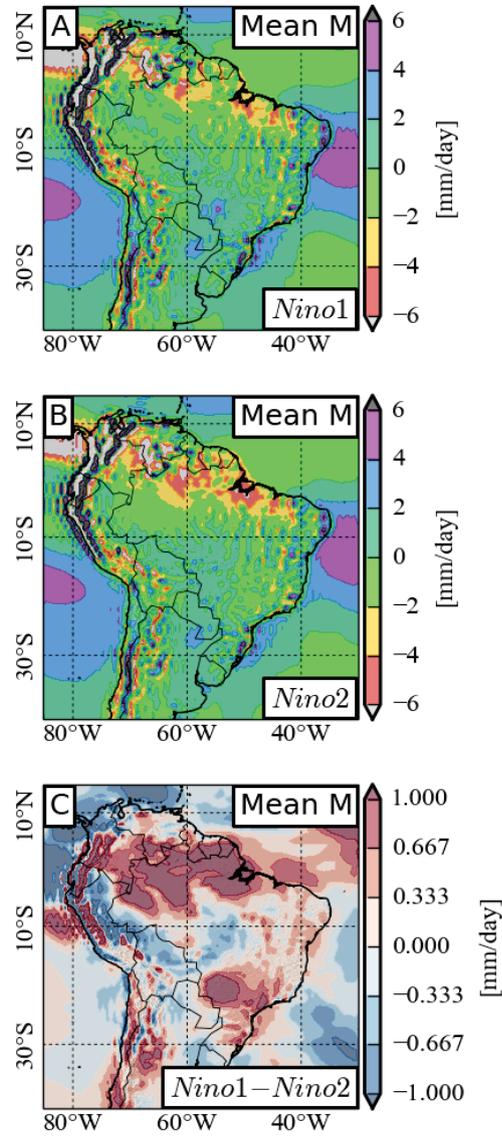
For strong events of evapotranspiration, we have found a clear block structure in the  $L^1$ -distance matrix  $\mathbf{L}$  between the LC patterns for different time intervals (Fig. 9.6). This is expressed by a relatively high degree of similarity of LC patterns within the periods 1979-1998 as well as 1992-2012 (with a mean  $L^1$ -distance of  $2.45 \cdot 10^7$ ), while the  $L^1$ -distance of LC patterns between these two periods is relatively high (on average  $2.57 \cdot 10^7$ ). The appearance of these two blocks in  $\mathbf{L}$  may indicate a possible transition from one climate “state” to another one (characterized by a somewhat different typical clustering pattern of extreme evapotranspiration events) during the mid 1990s. In fact, a recent study (Meehl et al., 2011) suggested a shift from El Niño dominated decades lasting until 1998 to a La Niña dominated period during the last 15 years, which is consistent with our finding. Specifically, the mean value of MEI for the time period from 1979 to 1998 is +0.54, while the mean value of MEI for the years from 1992 to 2011 is +0.19 (−0.12 for the years from 1998 to 2011). Recently, this shift was suggested as a possible explanation of the non-increasing global mean temperature in the last 15 years by the cooling effect of the Pacific Ocean during colder ENSO conditions (Kosaka and Xie, 2013).

### 9.6.3. Different types of El Niño events

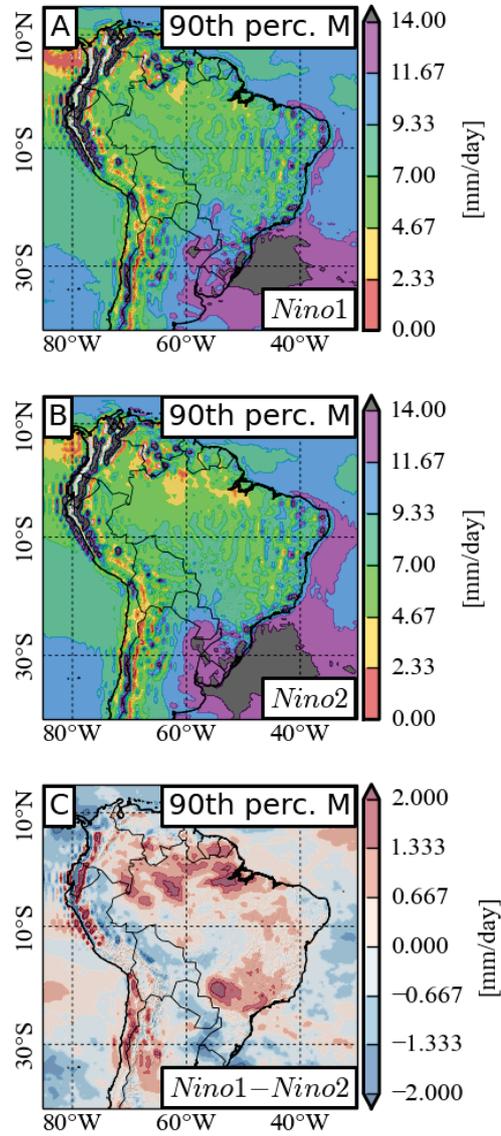
During the last years, several studies have reported evidence for a multi-phase nature of ENSO with at least two qualitatively different types of El Niño events (Ashok et al., 2007; Yeh et al., 2009; Hendon et al., 2009; Kim et al., 2011; Hu et al., 2012). However, we note that the assignment of El Niño events to these two subclasses is not fully consistent in the literature and partly depends on the chosen ENSO index. We

suggest that a more detailed discrimination of ENSO phases allows deeper insights into the spatiotemporal organization of vertical moisture flux extremes. Here, we resolve the above results with respect to the two types of El Niño events and divide all El Niño events into the following two subclasses: Nino1, consisting of the particularly strong *classical* El Niño events in 1982, 1987, and 1997, and Nino2, consisting of the *anomalous* events in 1986, 1990/1991, 1993/1994, 2002, and 2004 (Hendon et al., 2009). We remark that in (Hu et al., 2012), the El Niño event of 1987 is - in contrast to our assignment and (Yeh et al., 2009; Hendon et al., 2009; Kim et al., 2011) - taken to be an anomalous El Niño (i.e., Nino2) event. The classical El Niño (Nino1) is also referred to as *eastern Pacific El Niño*, while the anomalous El Niño (Nino2) is also called *central Pacific El Niño*, or *El Niño Modoki* (Ashok et al., 2007; Yeh et al., 2009). Nino2 is characterized by positive sea surface temperature (SST) anomalies in the central Pacific Ocean, but relatively cool SSTs to the east and west, while during (classical) Nino1 conditions, the SST maximum is located in the eastern Pacific Ocean. Furthermore, the two types of El Niño differ in their teleconnection patterns between the tropics and midlatitudes. The frequency of Nino2 episodes has increased during recent decades, possibly because of weakened tropical easterly winds (Ashok et al., 2007; Yeh et al., 2009). For a detailed analysis of the distinct impacts of Nino1 and Nino2 on precipitation over South America, we refer to (Hill et al., 2009; Hill et al., 2011; Tedeschi et al., 2013), while (Li et al., 2011) investigate the general impact on the climate of the Amazon Basin.

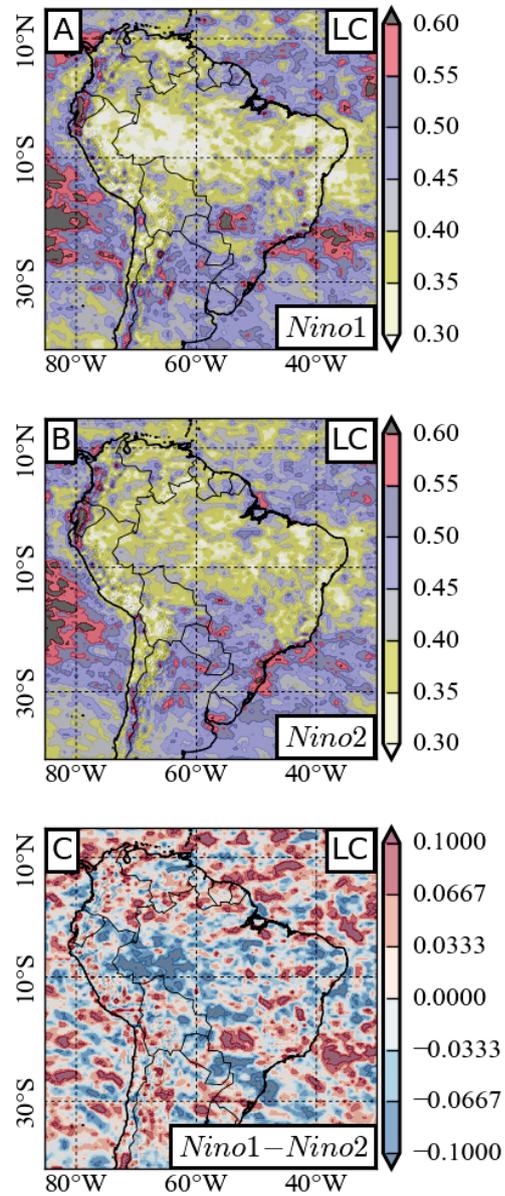
Constructing composites of means (Fig. 9.12) and 90th percentiles (Fig. 9.13) of moisture divergence as well as LC (Fig. 9.14) for these two types separately, we find that the impact of El Niño events on these three fields is mainly due to the classical type (Nino1), which shows much stronger deviations from neutral ENSO conditions than the anomalous type (Nino2). We thus infer that pronounced SST anomalies in the eastern Pacific ocean contribute much stronger to the described impacts of ENSO on moisture divergence than SST anomalies in the central Pacific ocean. This is consistent with results in (Radebach et al., 2013), where distinctively different expressions in the properties of networks constructed from global surface air temperatures were found for the two El Niño types. In particular, the impact of El Niño events on GC was also mainly assigned to classical El Niño events in that study.



**Figure 9.12.:** Composites of mean daily moisture divergence (M) for El Niño events of type Nino1 (A), Nino2 (B), and difference between Nino1 and Nino2 (C).



**Figure 9.13.:** Composites of 90th percentile thresholds of daily moisture divergence (M) for El Niño events of type Nino1 (A), Nino2 (B), and difference between Nino1 and Nino2 (C).



**Figure 9.14.:** Composites of LC for El Niño events of type Nino1 (A), Nino2 (B), and difference between Nino1 and Nino2 (C).

## 9.7. Conclusion

It was shown that magnitudes as well as spatial clustering of strong evapotranspiration events over South America are strongly affected by the phase of the El Niño Southern Oscillation (ENSO). The strongest impacts of ENSO were found over the Amazon Basin, which is particularly important since this region is known to host globally relevant and particularly vulnerable ecosystems (Davidson et al., 2012). Specifically, the Amazon rainforest is believed to be one potential tipping element of the Earth's global climate system (Lenton et al., 2008).

The key observations in this chapter are: i) The magnitudes of moisture divergence respond positively to ENSO variability (indicating negative precipitation anomalies) in the Amazon Basin with highest values during El Niño periods. ii) The network-derived local clustering of extreme positive events of moisture divergence is reduced in a characteristic way during El Niño events in the Amazon Basin. This indicates reduced spatial coherence of synchronized extreme evapotranspiration events in this region. We find that the described dependence on ENSO variability is mainly due to the three major (classical) El Niño events in 1982, 1987, and 1997, while all other events (in particular the anomalous Modoki events) play a minor role. A shift in ENSO activity during the late 1990's from warmer (El Niño) to cooler (La Niña) predominant conditions reported by others (Meehl et al., 2011) is reflected by a shift in the pattern similarity of the network's local clustering coefficients computed for extreme positive events of moisture divergence. iii) We find that our results obtained from complex network theory are consistent with, but more detailed than results obtained from a direct estimation of the spatial coherence of events.

It should be emphasized that these results were only found for the complex network analysis of extreme positive moisture divergence (i.e., evapotranspiration) events, and no corresponding significant results were found for extreme negative moisture divergence (i.e. precipitation) events.

In the previous chapters, all analysis was static in the sense that one network was constructed from the information obtained for the entire available time span. Here, by using sliding windows of networks, we showed how this can be generalized to a dynamical analysis. In future work, the impacts of La Niña episodes should be analyzed in more detail, as in this chapter the main focus was on the influence of El Niño episodes.



# Chapter 10.

## Conclusion

### 10.1. Contributions of this thesis

The aim of this dissertation was to advance the understanding of the collective dynamics of extreme rainfall. Traditionally, spatial patterns encoding the co-variability of climatic time series are derived in terms of empirical orthogonal functions (EOFs), derived from principal component analysis (PCA) of the dataset's covariance matrix. However, no technique based on the spectral decomposition of the covariance matrix is suitable for analyzing the co-variability of extreme events: If computed directly for the original time series, the covariance matrix will only capture the first two moments of the data distributions, and can thus not provide any information on the behavior of the extreme events, which are located in the tails of the frequency distribution. Alternatively, the covariance matrix could be applied to binary extreme event series derived from the original time series. However, this approach would not be suited to deal with varying time lags between events at different locations. Furthermore, a well-known caveat of PCA is that the corresponding eigenvectors (i.e., the EOFs) are not statistically independent for non-Gaussian data distributions (which is the case for extreme rainfall events), and can therefore not be assigned to independent dynamical modes of the system. A new methodology had thus to be found to overcome these problems. For this purpose, a theoretical framework for the analysis of synchronization phenomena of extreme events in significantly interrelated time series was developed on the basis of complex network theory. In principle, this framework can be applied to time series obtained from many kinds of complex systems as a general tool to explore data and, from the resulting synchronization patterns, develop hypotheses about the underlying driving mechanisms. Alongside with providing the means to analyze a given interactive system in terms of synchronization of extreme events, this mathematical tool can in certain situations be used to forecast these extreme events on a statistical footing.

The main idea of the methodology developed in this thesis is to identify the time series encoding the behavior of the single parts of a complex system with network nodes, and represent statistically significant synchronizations of extreme events between different parts of the system by network links. The rationale behind this approach is that the complex synchronization phenomena emerging from the interactions become accessible in terms of the network topology. Synchronization of all pairs of time series was estimated in terms of several modifications of the similarity measure Event

Synchronization, which allows for the delays between synchronous events to vary in time. A framework to determine the statistical significance of the outcomes of this measure in view of varying event rates and clustered event distributions was proposed, and networks were constructed with respect to this framework. In addition to established network measures, several new measures on undirected as well as on directed and weighted networks were introduced for a quantitative analysis of the network topology.

The theoretical framework was used to study the spatial characteristics of extreme rainfall synchronicity in South America on the basis of gridded, high spatiotemporal resolution satellite data. Climate networks were constructed by representing the strongest and most significant values of the similarity measure as network links. The topology of the resulting networks could then be analyzed in terms of various network measures, for which specific climatological interpretations were proposed in order to put their spatial distribution into relation with the underlying climatic mechanisms. We focussed on the following three specific fields of application:

### **Climatic Analysis**

In a first step, we analyzed the dynamical properties of extreme rainfall events of the South American monsoon by constructing undirected networks and quantifying different aspects of their topology using the standard measures degree, betweenness centrality, and clustering, as well as the measures mean geographical distance, long-ranged directedness, and directionality, which were introduced here. Upon providing climatic interpretations for these measures, we were able to show that their spatial distribution reveals several of the main features underlying the South American monsoon system.

Some of the features we revealed were already known in the literature, and in this sense some of the results can be understood as a proof of concept of the methodology. These include the main convergence zones, namely the Intertropical as well as the South Atlantic Converge Zones, as well as areas with frequent development of Mesoscale Convective Systems in southeastern South America. Events associated with these large rainfall clusters are typically above the 95th percentile, and contribute large fractions of total seasonal precipitation. Most importantly, however, we could reveal the dominant propagation route of rainfall events from the Amazon Basin along the eastern slopes of the Andes to the subtropics, driven by the trade winds and the South American Low-Level Jet (chapter 4). The specific role these features played for the dynamics of extreme events was not known due to lacking analytical tools. In this sense, these results in fact contribute to the understanding of the South American climate system, expressed as climatic interpretations of spatial pattern derived from the synchronization structure of extreme rainfall events.

The following results could not have been derived using more traditional data analysis techniques, and thus provide new insights into the functioning the South American monsoon systems: By constructing networks separately for the two phases of the the main pattern of variability associated with the monsoon system, namely the

so-called South American rainfall dipole, we could reveal the main synchronization pathways of extreme events corresponding to the two phases. This also shed light on the interplay of tropical wind systems and subtropical frontal systems for driving extreme rainfall propagation. Most notably, over the entire South American continent, the synchronization patterns of the two phases of the rainfall dipole are determined by the atmospheric waves that trigger frontal systems in the subtropics. This suggests that the well-known rainfall dipole is only the most pronounced part of an oscillation that extends beyond the equator. In the corresponding chapter 5 we showed that these results are consistent with the general understanding of the monsoon system, although they go beyond the hitherto known.

In chapter 9, we generalized our approach to study temporal changes in the network-derived spatial patterns using a sliding window approach. We could show that the network clustering of reanalysis-derived evapotranspiration events strongly responds to the El Niño Southern Oscillation, with characteristic patterns of low global clustering recurring during El Niño events.

**Associated publications:** Boers et al., *Geophysical Research Letters* (2013, P1), Boers et al., *Geophysical Research Letters* (2014, P2), Rheinwalt et al. (submitted, P3), Boers et al., *Climate Dynamics* (accepted, P8)

## Prediction of extreme events

In view of the potential predictability of extreme rainfall events, directed networks were constructed in order to analyze the synchronization structure of extreme events in a time-resolved manner. This allowed us to determine the geographical source regions of three-hourly extreme events in the main river catchments along the Andean mountain range, which we resolved with respect to the temporal and spatial scales of these events. The mountainous terrain at the eastern slopes of the Andes is exposed to high hazard risks in form of floods and landslides, calling for an assessment of the potential predictability of extreme rainfall events. We could show that for the catchments south of 20°S, extreme rainfall typically originates from frontal systems approaching from the central Argentinean plains, while the catchments north of 20°S are mainly affected by squall lines originating from the Amazon Basin (chapter 6).

By introducing the new measure network divergence in chapter 7, a general framework for the statistical prediction of extreme events in complex systems could be established. By applying this measure to three-hourly rainfall data, we were able to identify southeastern South America as a pronounced source region for extreme rainfall events, and could further show that these events typically propagate towards the eastern slopes of the Central Andes. These results were surprising at first, since the low-level moisture flow is directed in the opposite direction, leading from the tropics along the eastern Andean slopes towards southeastern South America. We identified the interplay of the low-level moisture flow from the tropics, frontal systems approaching from the South, and the topography of the Andes as the responsible

mechanism: This interplay leads to the establishment of a wind channel from the western Amazon Basin towards southeastern South America, where the warm, moist air collides with cold air in the aftermath of the frontal system. This causes the formation of large rainfall clusters which, as a result of the northward moving low-pressure anomaly and its alignment with the Andes mountain range, propagate towards the Central Andes, with sometimes devastating impacts for the local populations. On the basis of these insights, we could formulate a simple forecast rule which predicts 60% of these events in the Central Andes, and even 90% during El Niño conditions.

**Associated publications:** Boers et al. (submitted, P4), Boers et al., Nature Communications (2014, P5), Boers et al. (in preparation, P6)

## Evaluation of Climate Models

While state-of-the-art global and regional climate models perform reasonably well in reproducing mean values of climatic observables, they are known to have substantial difficulties to correctly reproduce the tails of the distributions. In particular, extreme rainfall events are notoriously hard to adequately represent. While local rainfall distributions produced by climate models can be easily analyzed and thus evaluated, the dynamical representation of extreme events and their spatial interrelations had not yet been analyzed on the basis of the produced data. In chapter 8, we compared three observational (TRMM, TRMM RT, and GPCP) and three model-derived (ERA-interim, ECHAM6, and ETA) datasets regarding their implementation of the synchronization structure of extreme rainfall events using the methodologies developed in this thesis. The presented results suggest that none of the model-derived datasets could adequately reproduce the spatial patterns found for the satellite-derived and gauge-calibrated TRMM dataset, which has been used for all investigations in the previous chapters. Surprisingly, however, spatial synchronization patterns derived from the ECHAM6 and the ETA data are still closer to the patterns found for TRMM than the patterns observed for the ERA-interim reanalysis product.

**Associated publications:** Boers et al., Journal of Climate (accepted, P7)

## 10.2. Outlook

In this thesis, statistical significance was mainly estimated on the level of the measure of similarity. By estimating and correcting for the effects of the spatial embedding, also a statement on statistical significance on the level of the network measures was provided, in the sense that the actual spatial distributions of given network measures were compared to the outcomes expected from suitable statistical null models. However, there remain several open problems concerning the statistical significance of spatial patterns, which should be addressed in future work.

It will furthermore be interesting to see how the framework introduced in this thesis performs when applied to other variables than rainfall, such as different climatic

observables, but also to time series from other complex systems, e.g. from economics or biology. For such applications, one would not need to restrict oneself to extreme events, as the methodology is general enough to be applied to any kind of events which are assumed to synchronize. A possible example for other kind of events would be spiking neurons, or certain repeating symbols, i.e. characteristic sequences or patterns in the time series. Along these lines, more general synchronization patterns than just those exhibited by extreme events could be treated.

In chapter 9 we studied temporal changes of network measures computed from reanalysis-derived moisture divergence and explained them in a climatological context. Unfortunately, the short time span for which satellite-derived and high-resolution rainfall data is available (15 years) does not yet allow to study corresponding evolving patterns of reliable estimates of extreme rainfall events in a statistically sound way. However, given the 3-hourly resolution of the TRMM 3B42 dataset, it should not take long until the time series are long enough to use a sliding window approach or at least construct the networks separately for the different stages of the El Niño Southern Oscillation and obtain statistically significant results.

In the context of model intercomparison, it would be very interesting to extend the evaluation performed in chapter 8 to the entire globe, employing the complete set of coupled ocean-atmosphere general circulation models included in the Coupled Model Intercomparison Project, which forms the baseline for estimating climate change projections of the Intergovernmental Panel on Climate Change (IPCC). Unfortunately, however, the typical horizontal resolutions on which these models are integrated are still rather coarse, with the ECHAM6 model being an exception. This imposes caveats for an analysis of the dynamics of extreme rainfall along the lines of chapter 8. If the problem of the spatial resolution could be overcome, it would furthermore be very interesting to see how the spatial patterns encoding the synchronization structure of extreme events would react to changing the parameters or equations of the climate models under consideration. This could lead to substantial improvements of the dynamical implementation of extreme rainfall in these models.



# Appendix



Appendix A.

Additional figures for chapter 4

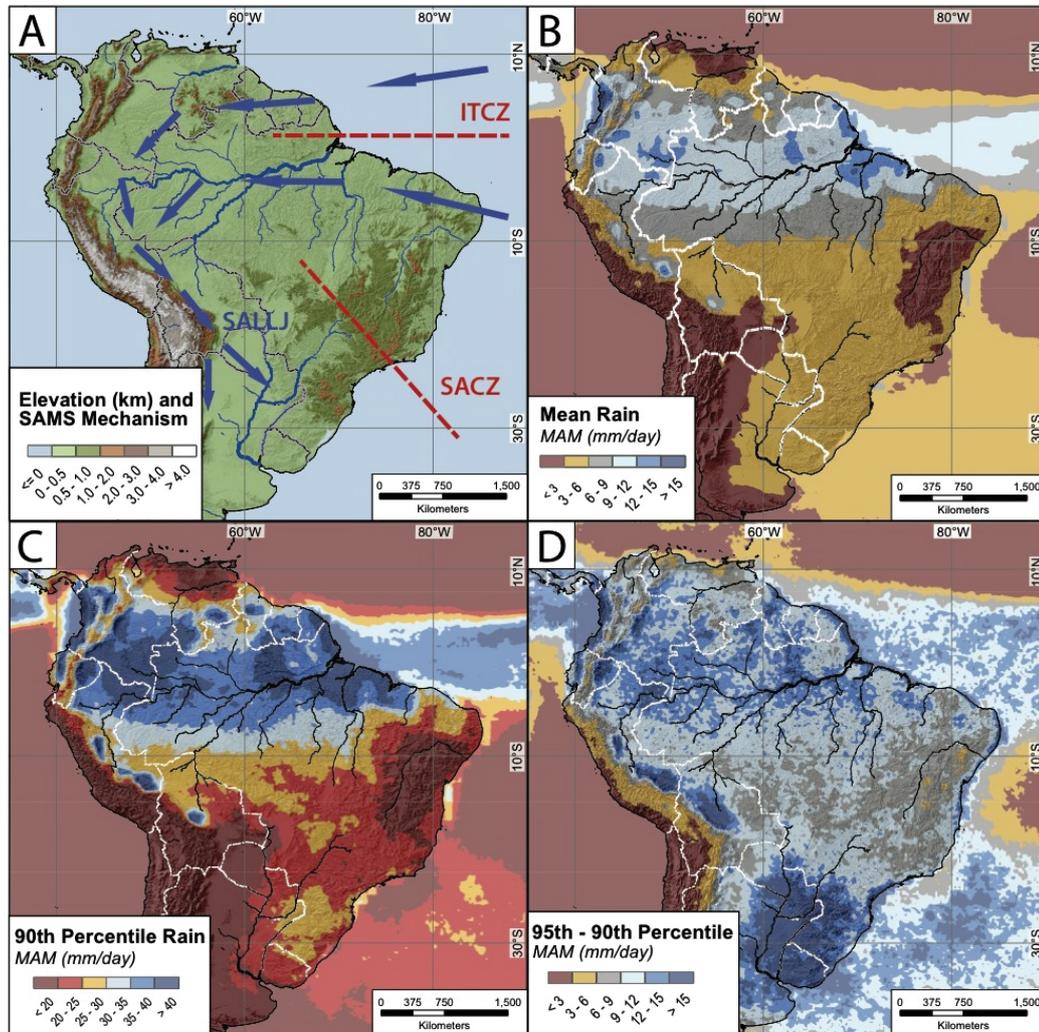


Figure A.1.: Same as Figure 4.1 but for the fall season from March to May (MAM).

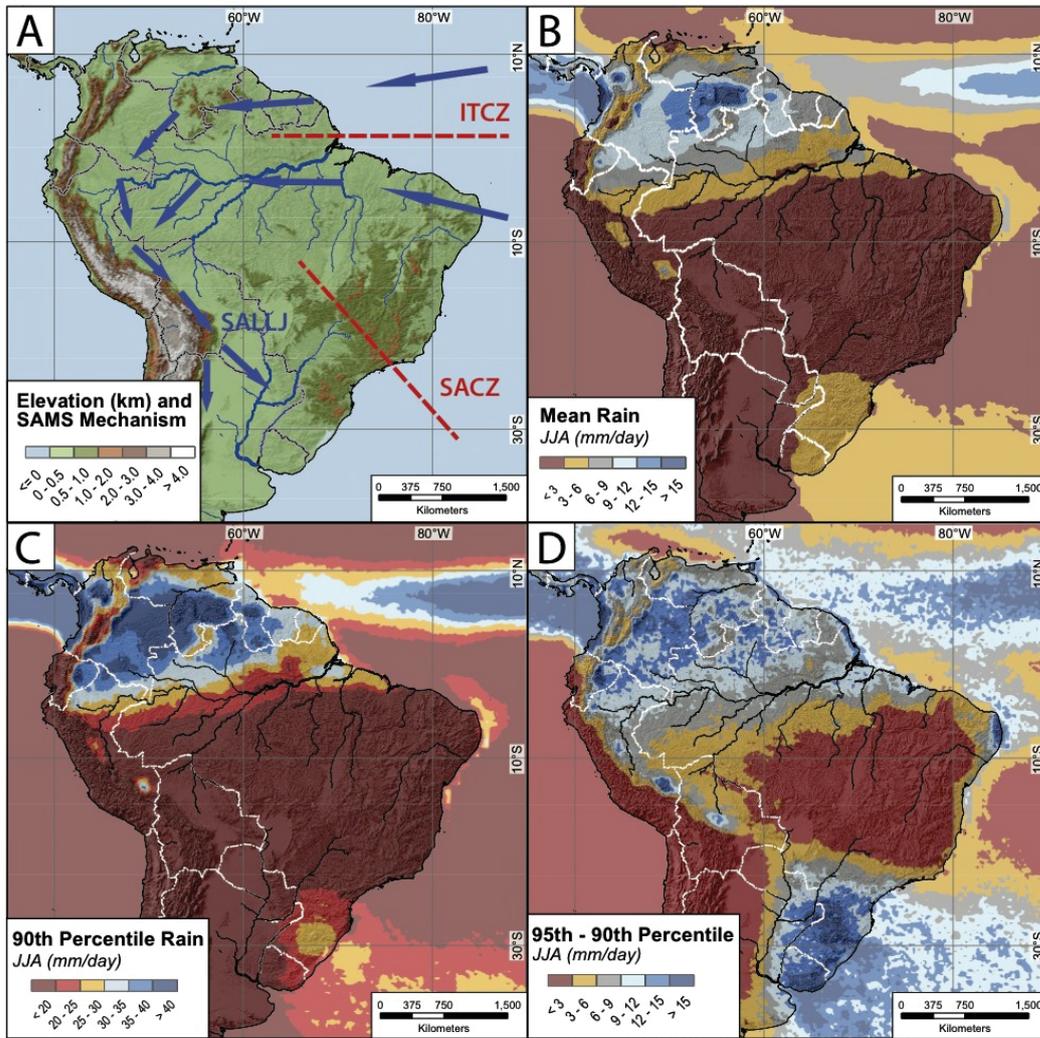


Figure A.2.: Same as Figure 4.1 but for the winter season from June to August (JJA).

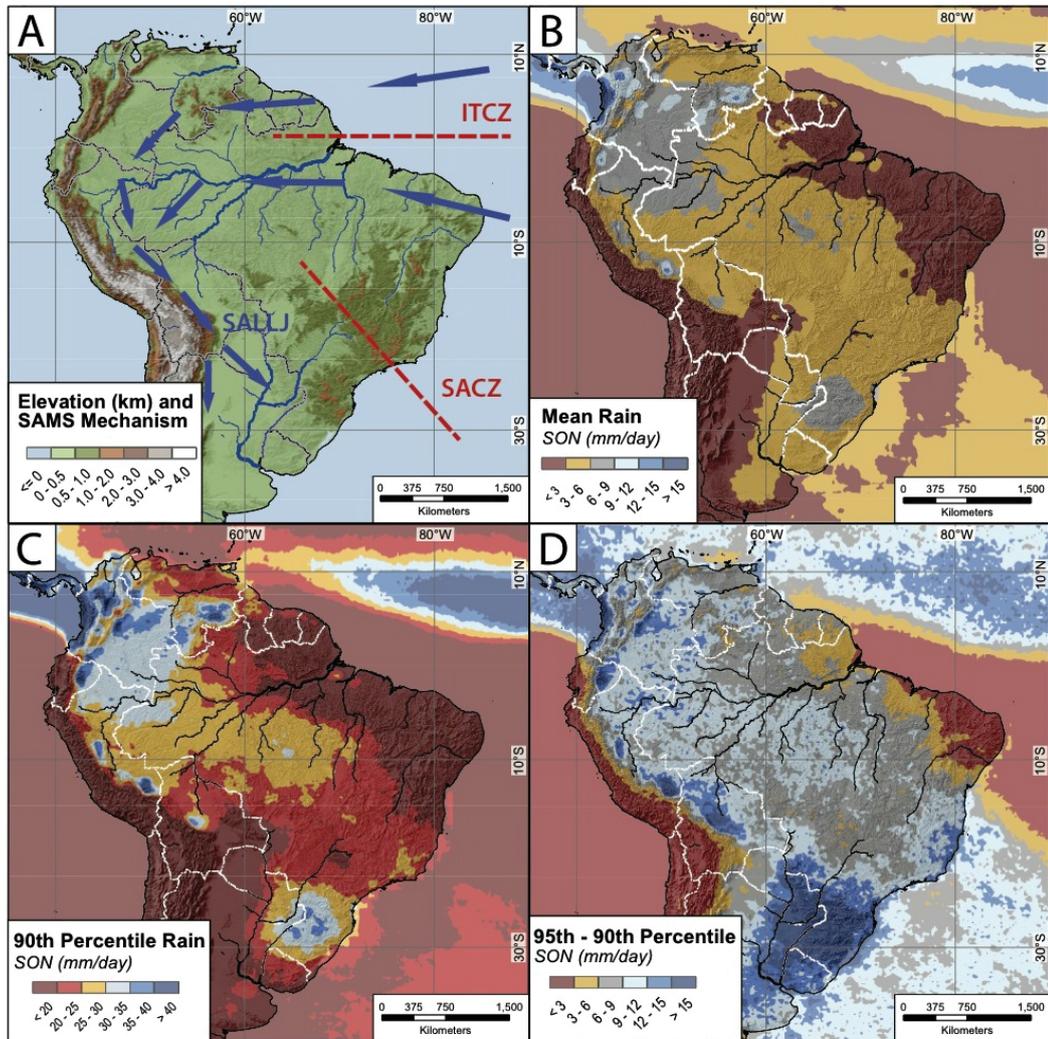


Figure A.3.: Same as Figure 4.1 but for the spring season from September to November (SON).

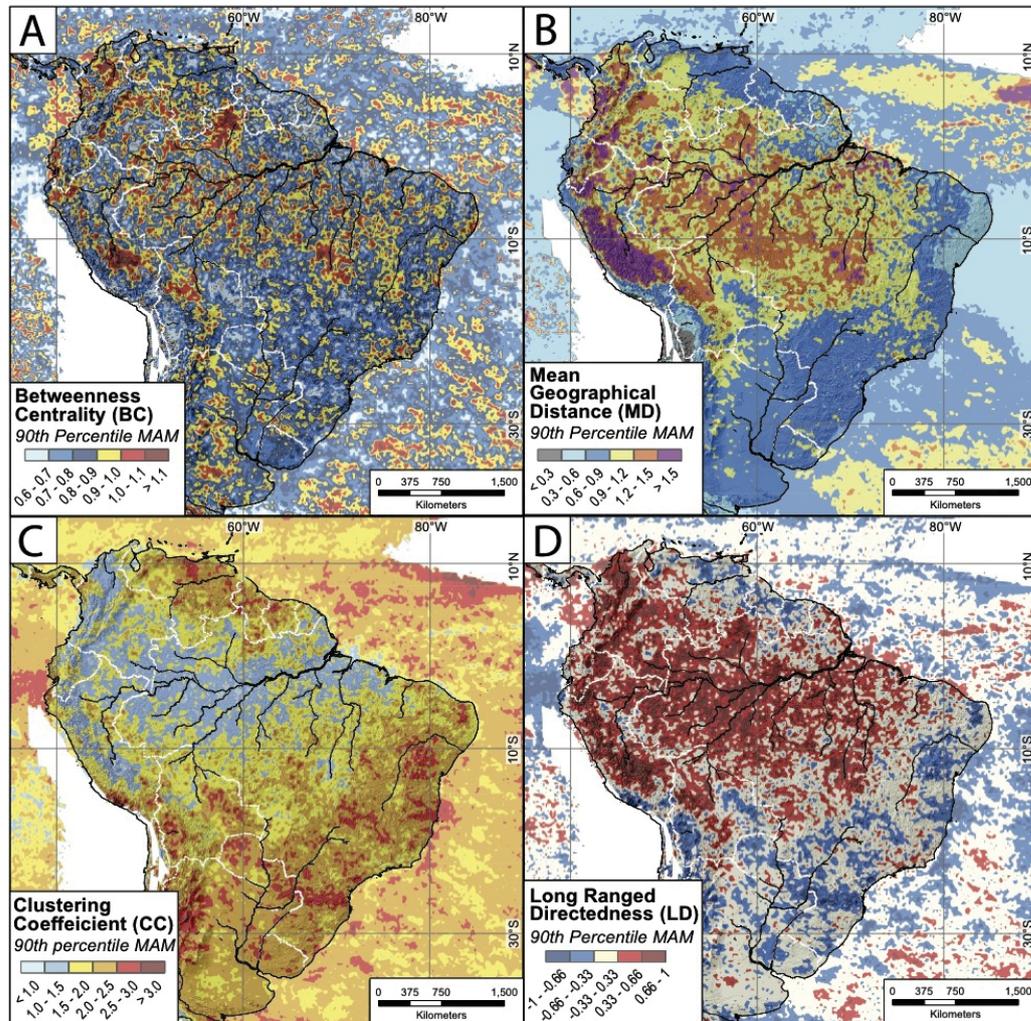


Figure A.4.: Same as Figure 4.2 but for the fall season from March to May (MAM).

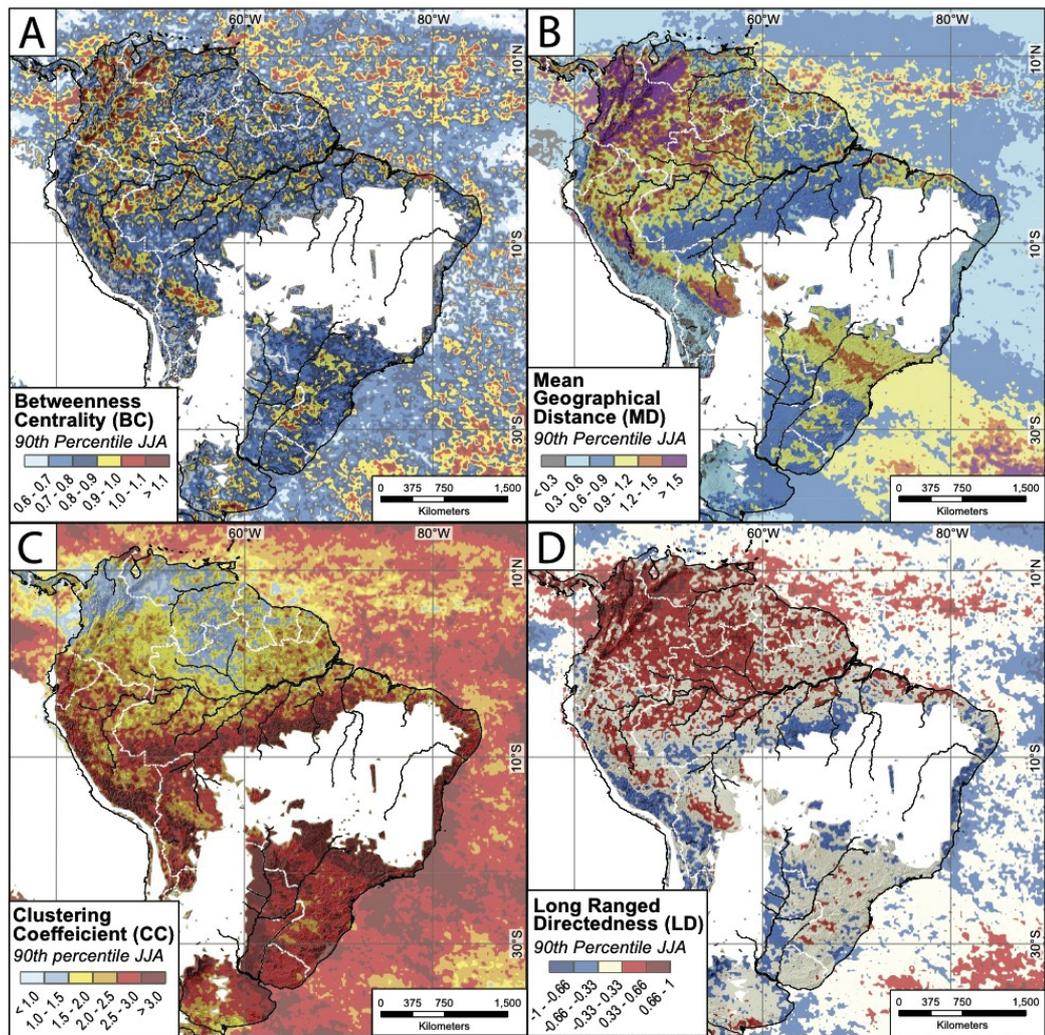


Figure A.5.: Same as Figure 4.2 but for the winter season from June to August (JJA).

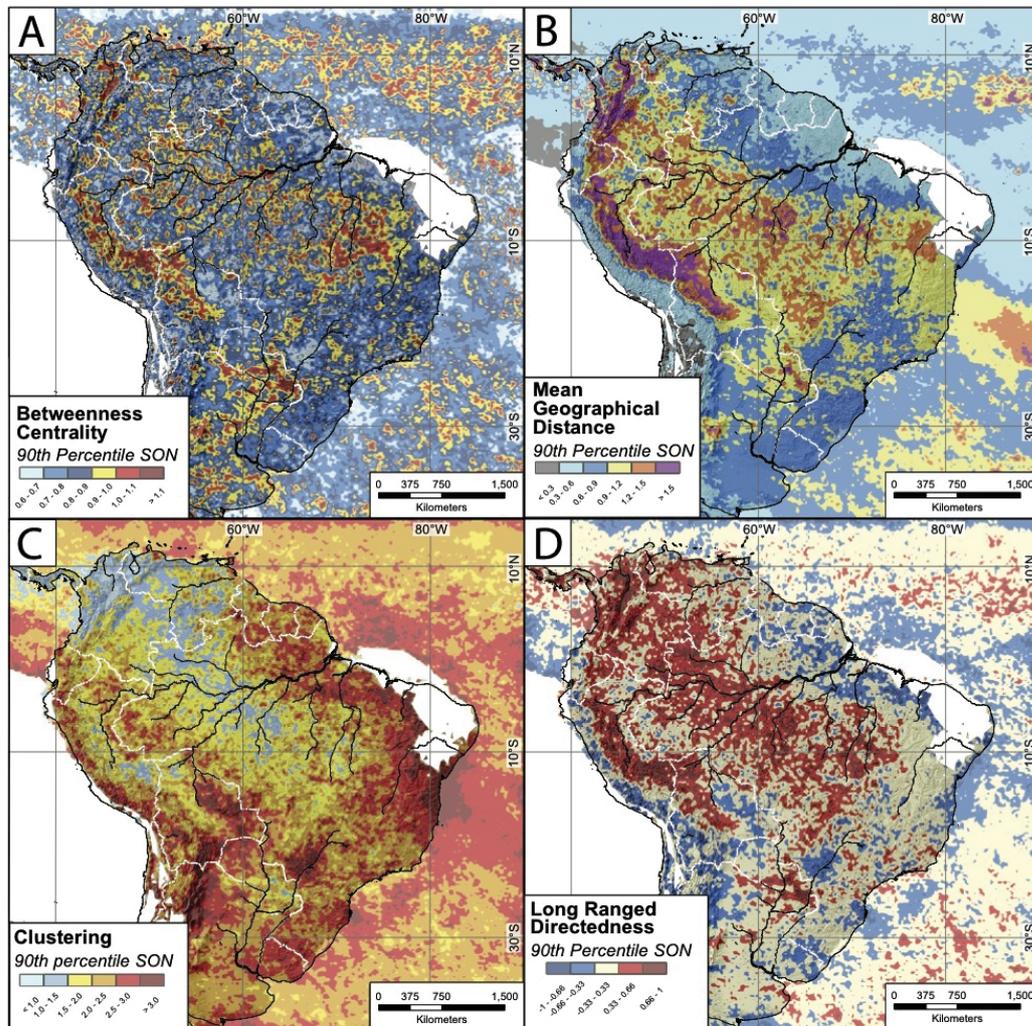


Figure A.6.: Same as Figure 4.2 but for the spring season from September to November (SON).

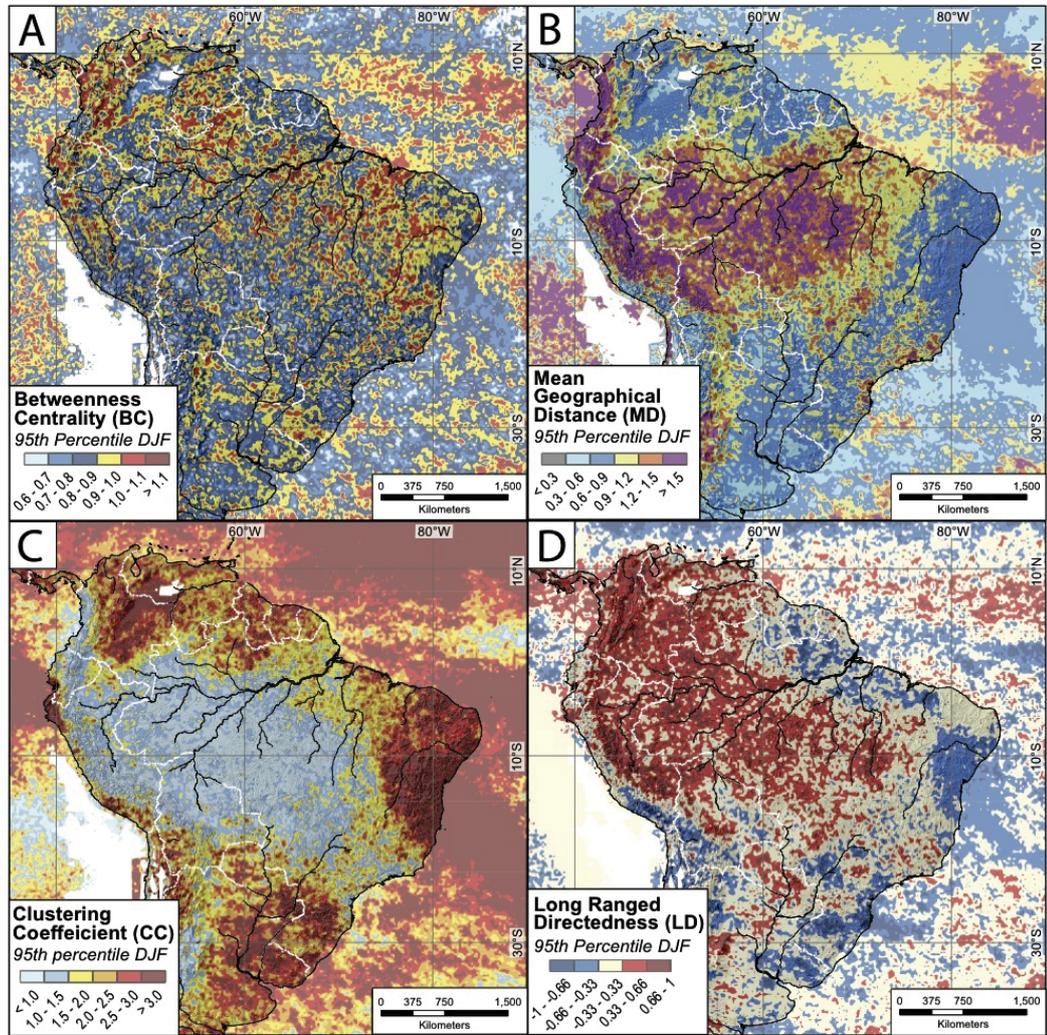


Figure A.7.: Same as Figure 4.1 but computed for events above the 95th percentile for the monsoon season from December to February (DJF).

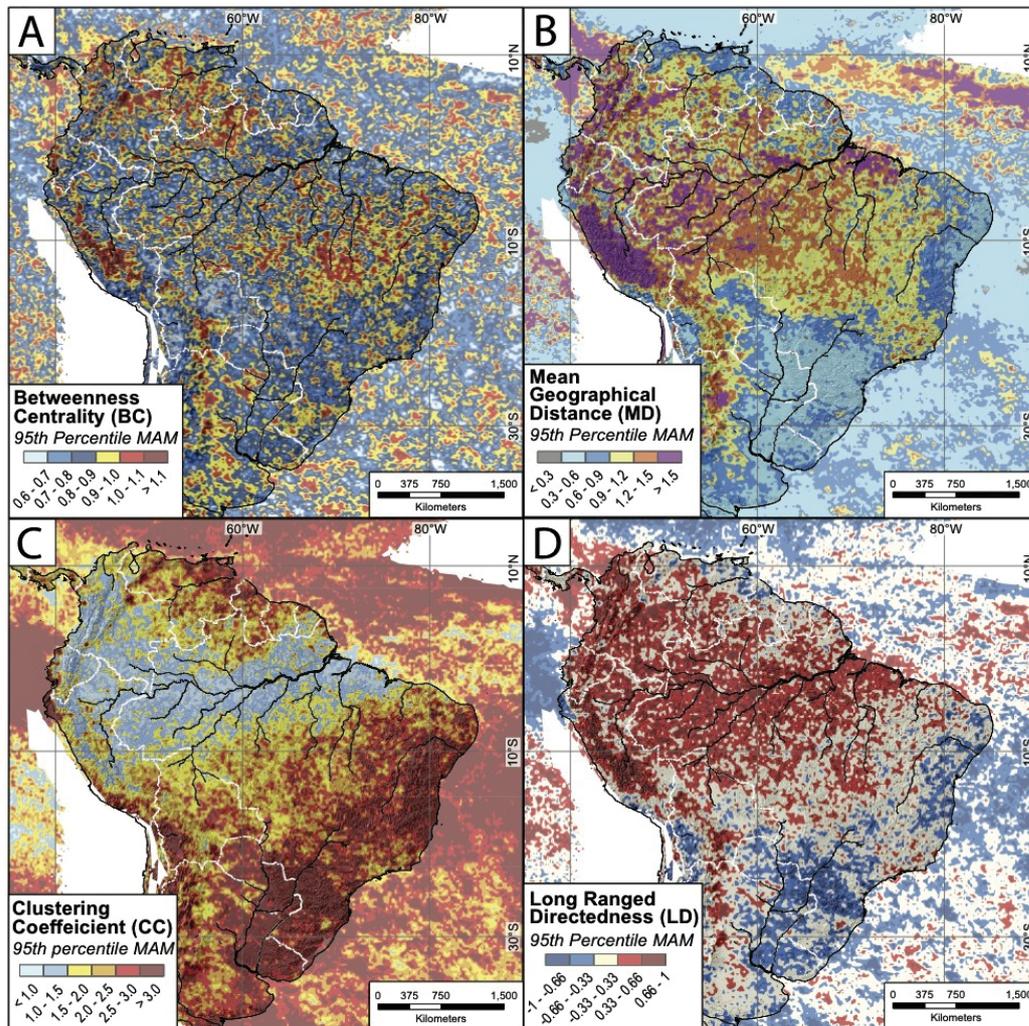


Figure A.8.: Same as Figure A.7 but for the fall season from March to May (MAM).

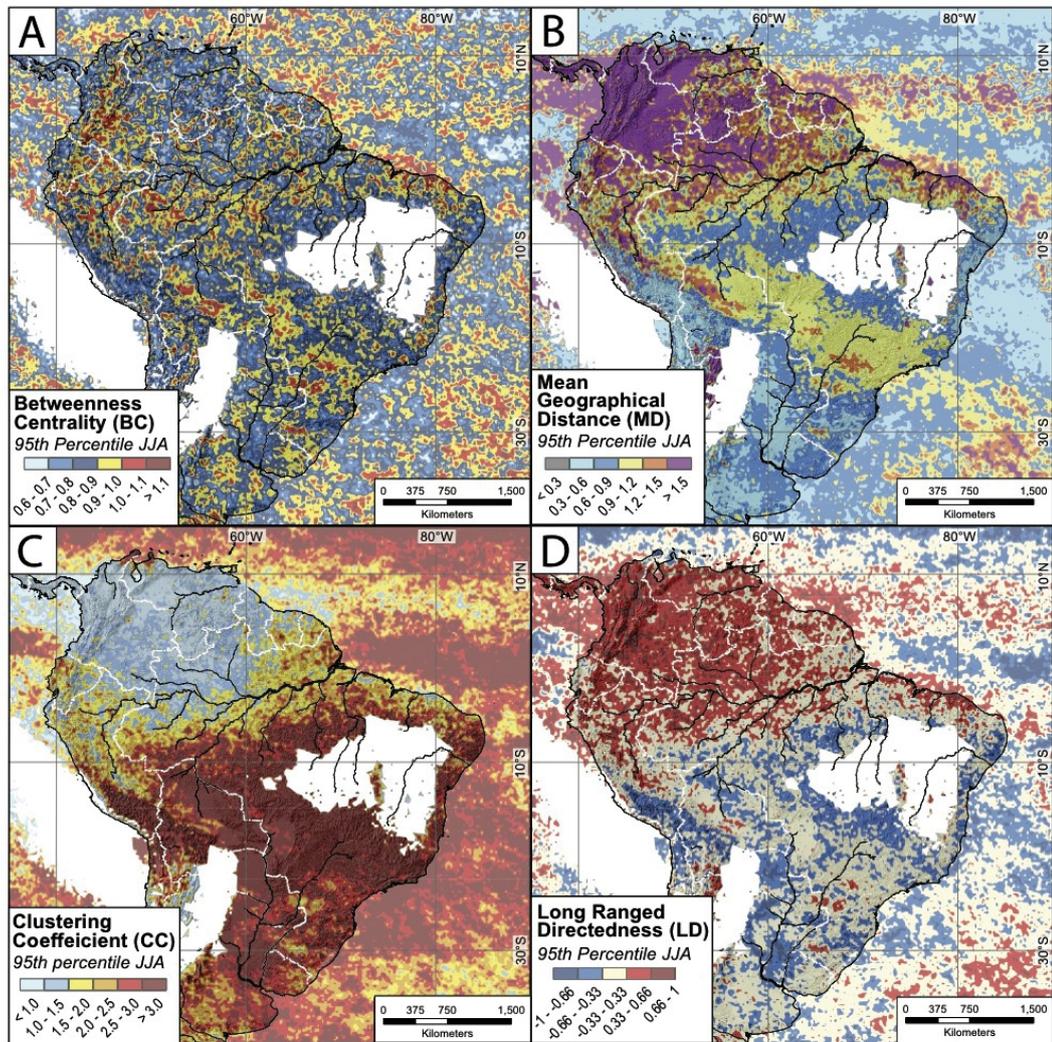
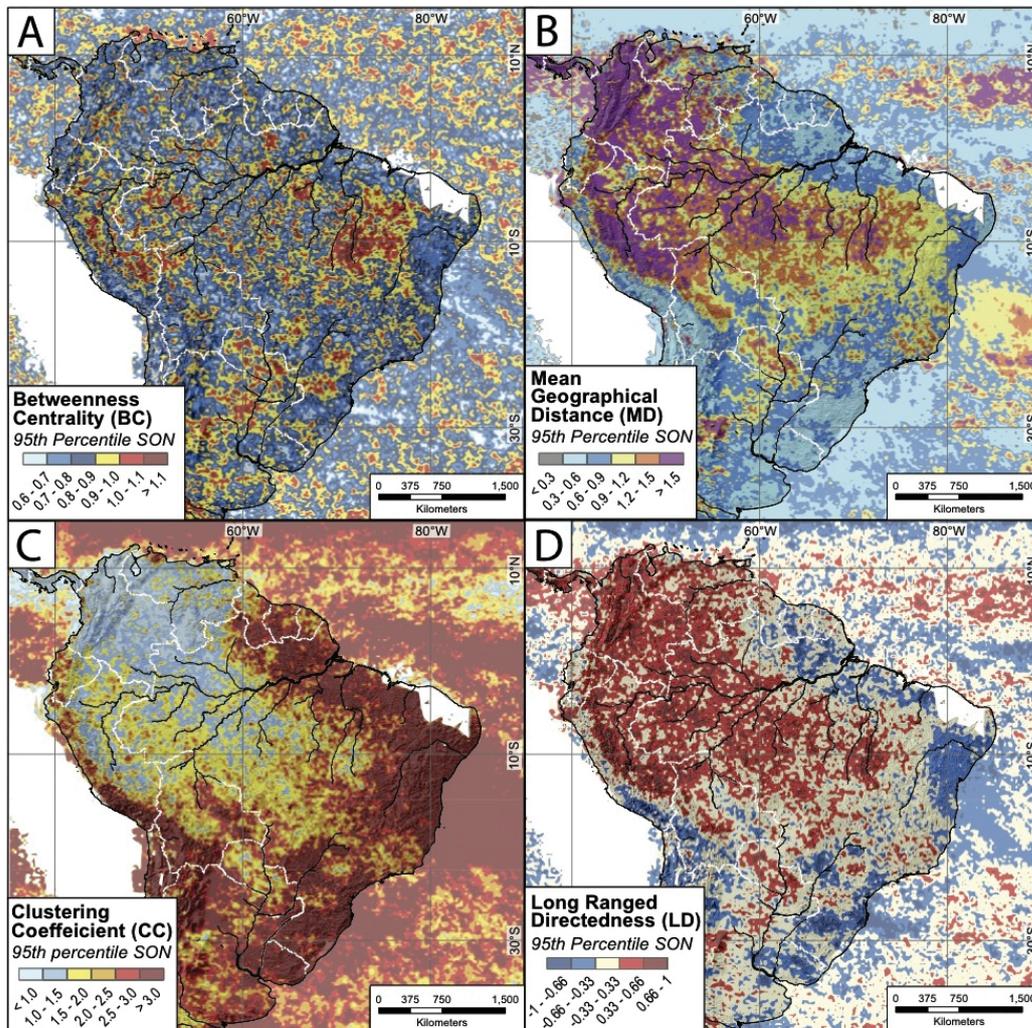


Figure A.g.: Same as Figure A.7 but for the winter season from June to August (JJA).



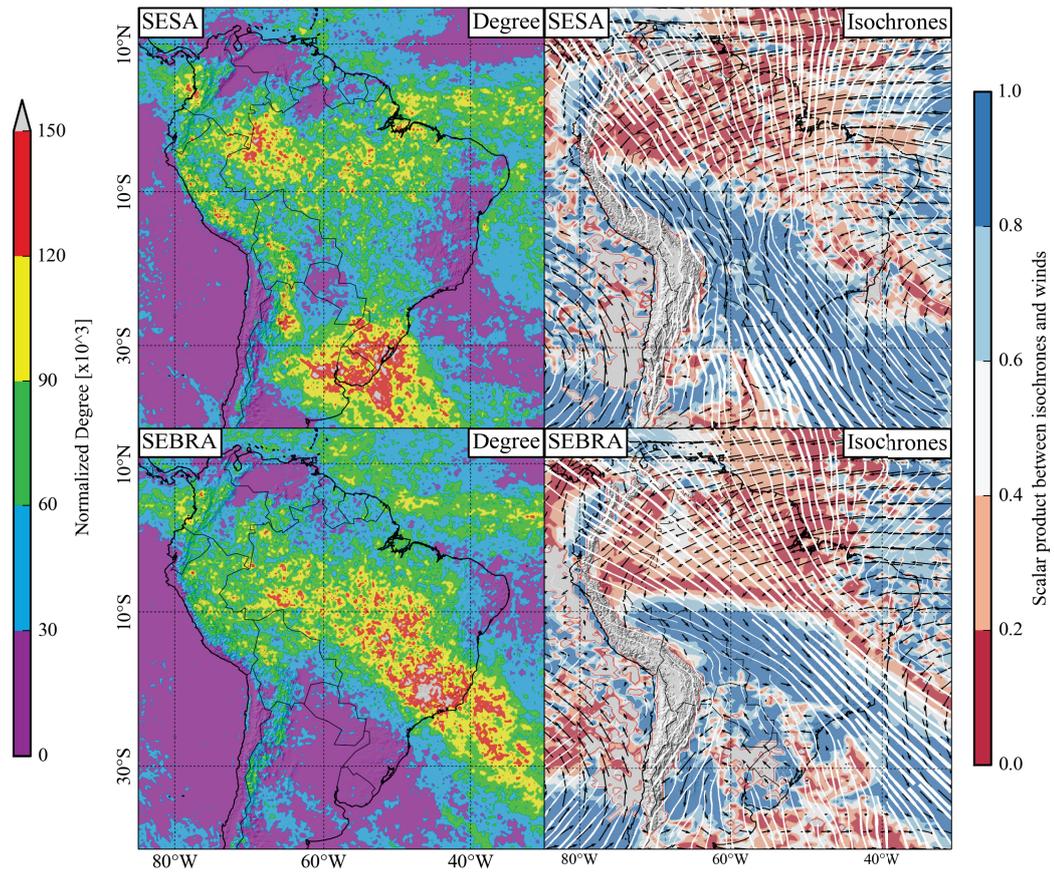
**Figure A.10.:** Same as Figure A.7 but for the spring season from September to November (SON).



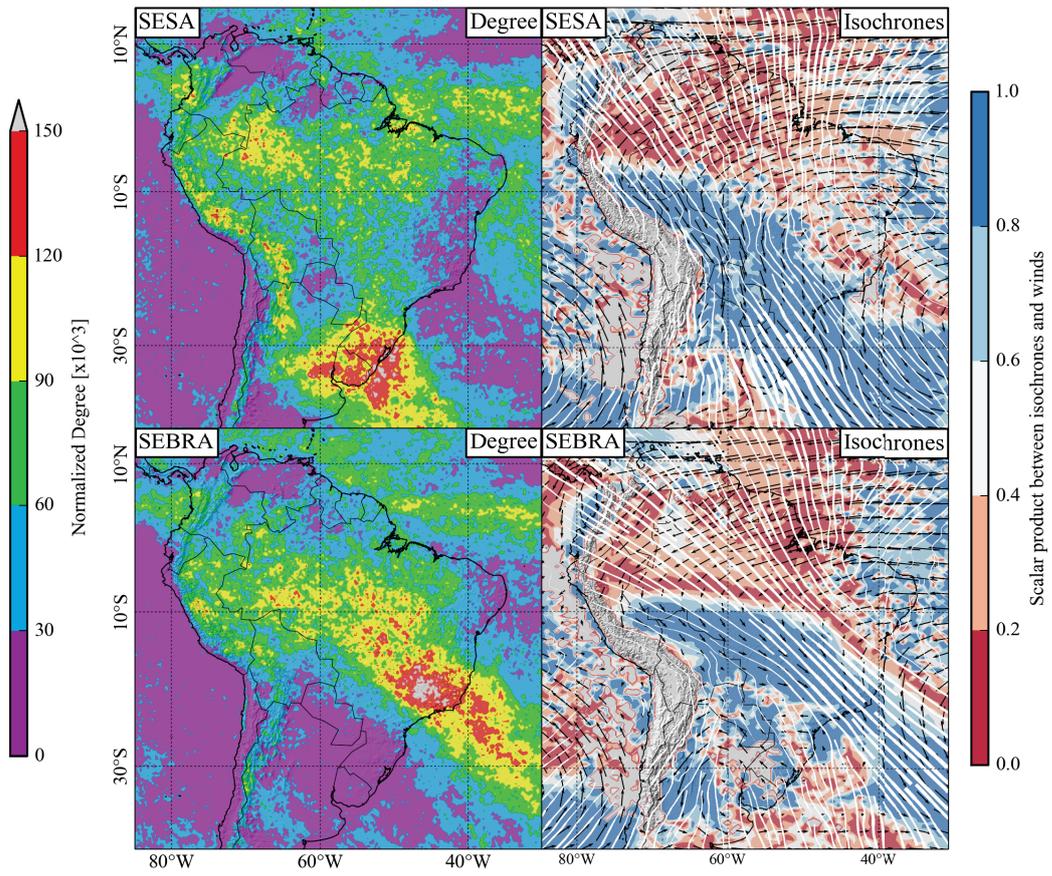
## Appendix B.

### Additional figures for chapter 5

In chapter 5, we claimed that computing the degree (DG) for the two phases of the South American rainfall dipole using  $\tau_{max} = 2$  days or  $\tau_{max} = 1$  day instead of  $\tau_{max} = 3$  days does not qualitatively change our results. Here, we show the corresponding results. Note that the directionality measure (DR, right column) is still computed for  $\tau_{max} = 0$  and thus identical with the version presented in chapter 5.



**Figure B.1.:** Same as Figure 5.2 in chapter 5, but the isochrones (right column) computed for  $\tau_{max} = 2$  days.

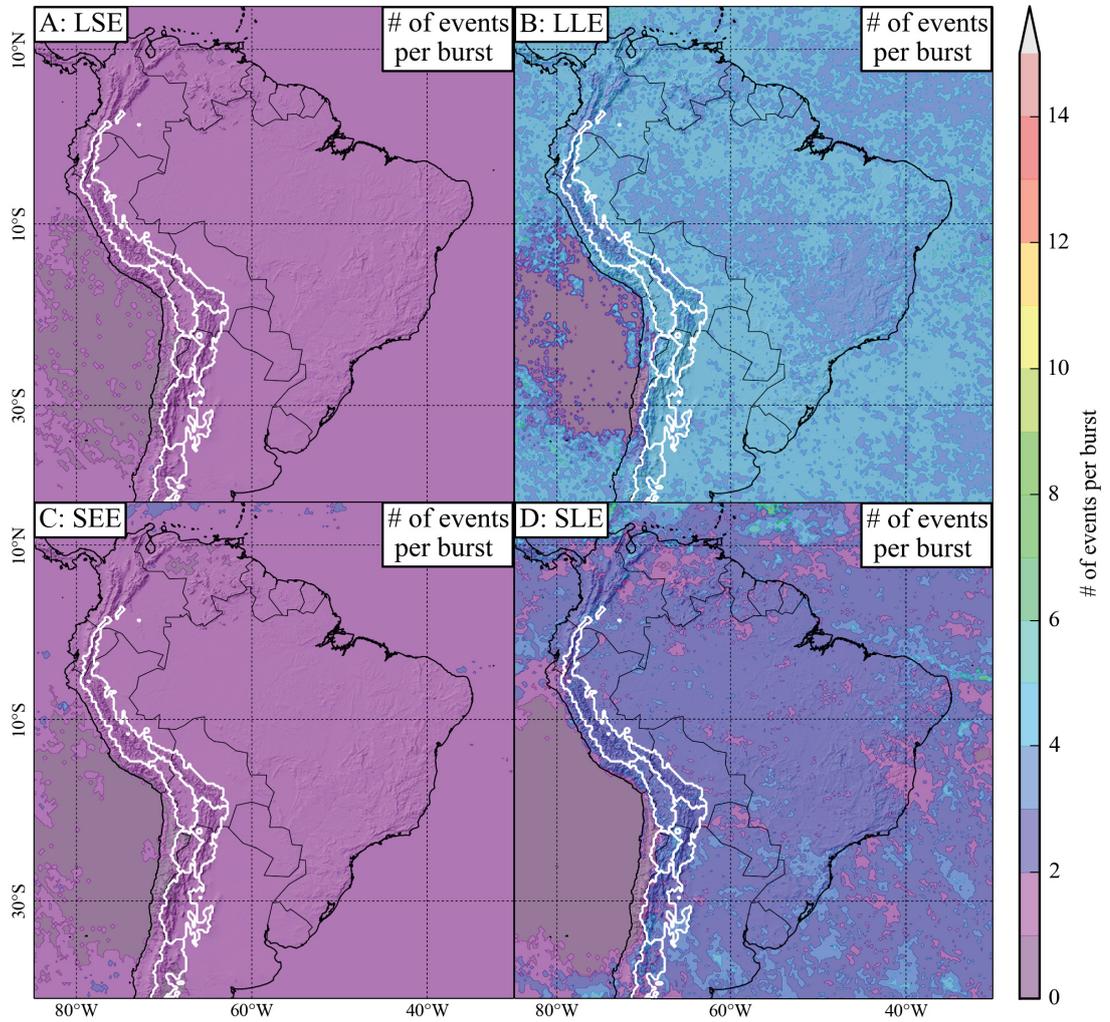


**Figure B.2.:** Same as Figure 5.2 in chapter 5, but the isochrones (right column) computed for  $\tau_{max} = 1$  day.

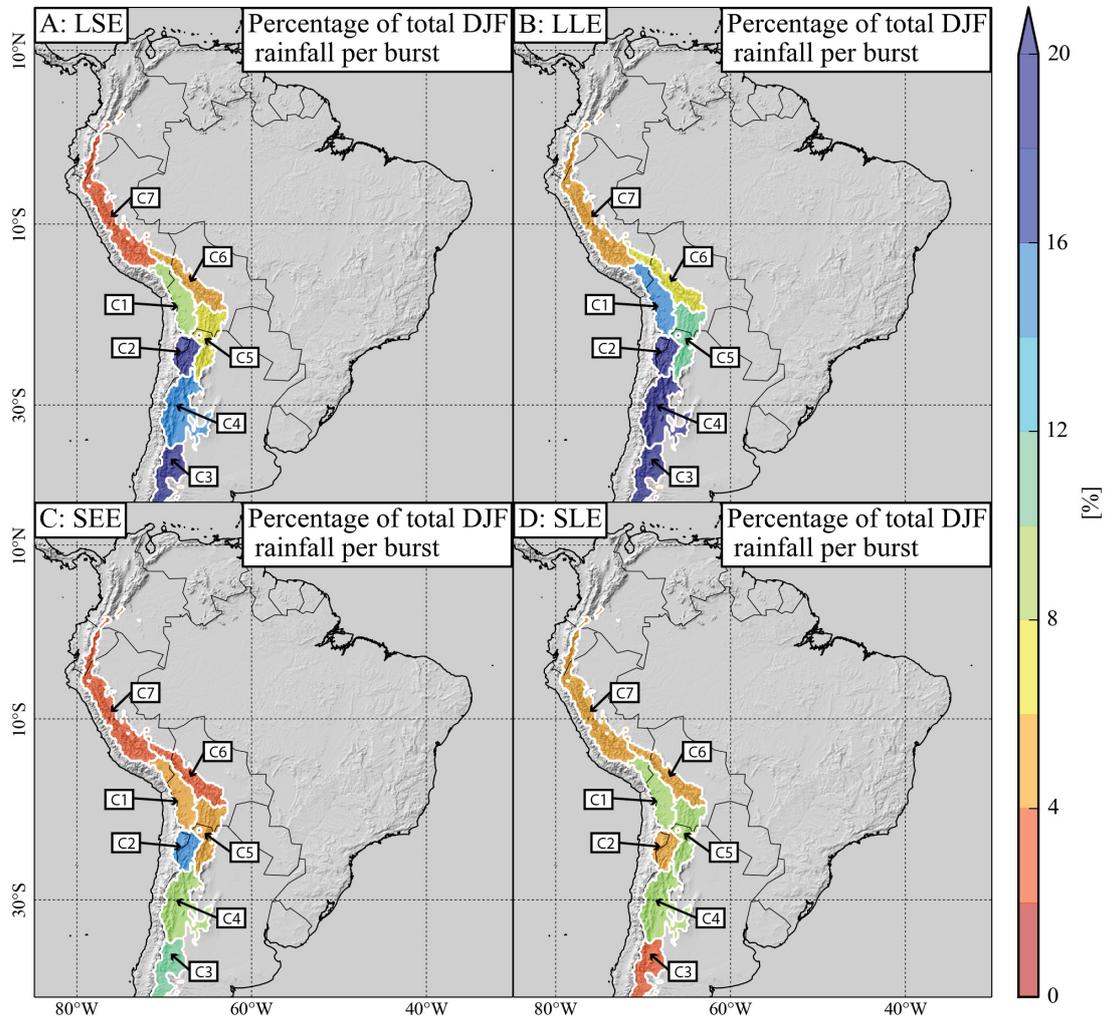


Appendix C.

Additional figures for chapter 6



**Figure C.1.:** Average number of 3-hourly events per burst of consecutive events of type local and short (A: LSE), local and long-lasting (B: LLE), spatially extensive (C: SEE), and spatially extensive long-lasting (D: SLE).

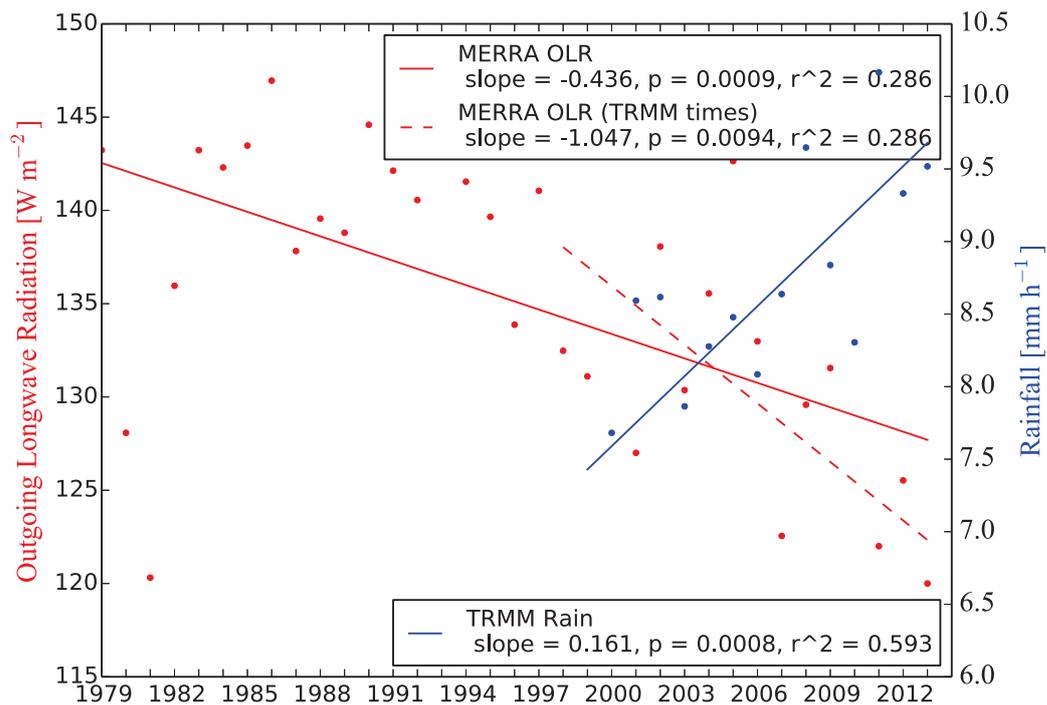


**Figure C.2.:** Average percentage of total DJF rainfall contributed to the catchments C1 to C7 by each single burst of consecutive events of type local and short (A: LSE), local and long-lasting (B: LLE), spatially extensive (C: SEE), and spatially extensive long-lasting (D: SLE).

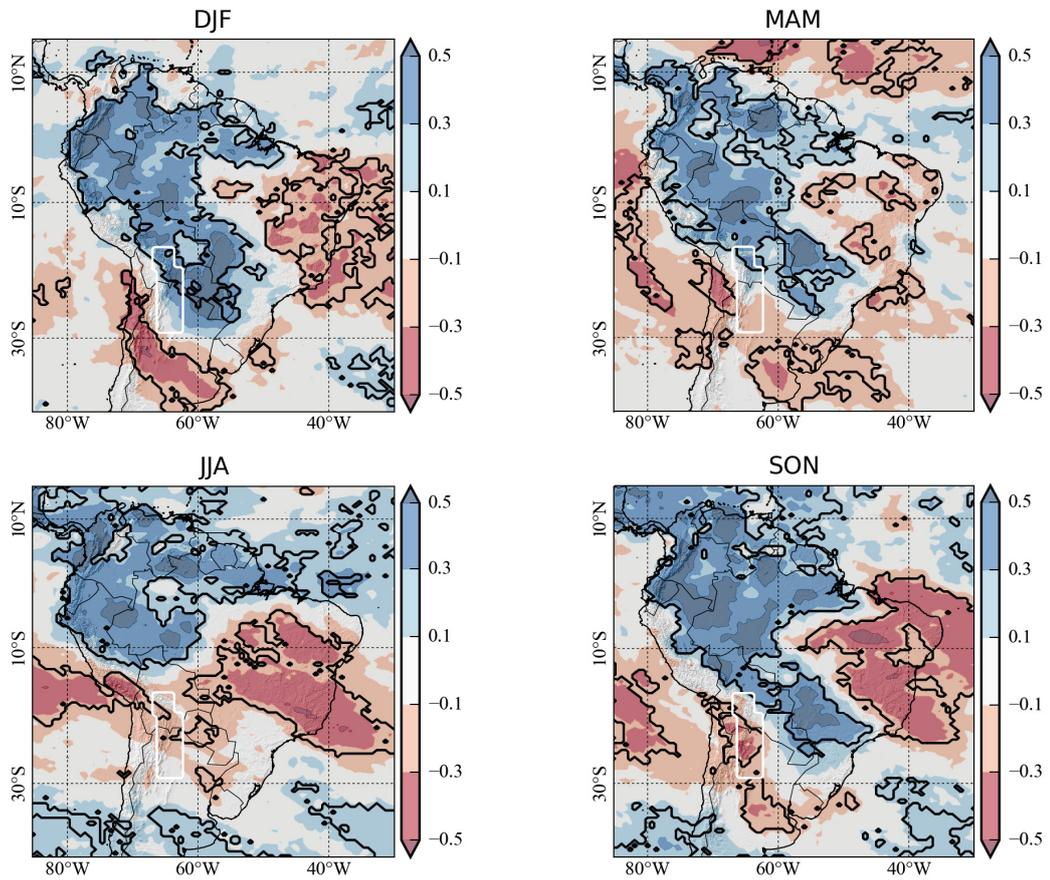


## Appendix D.

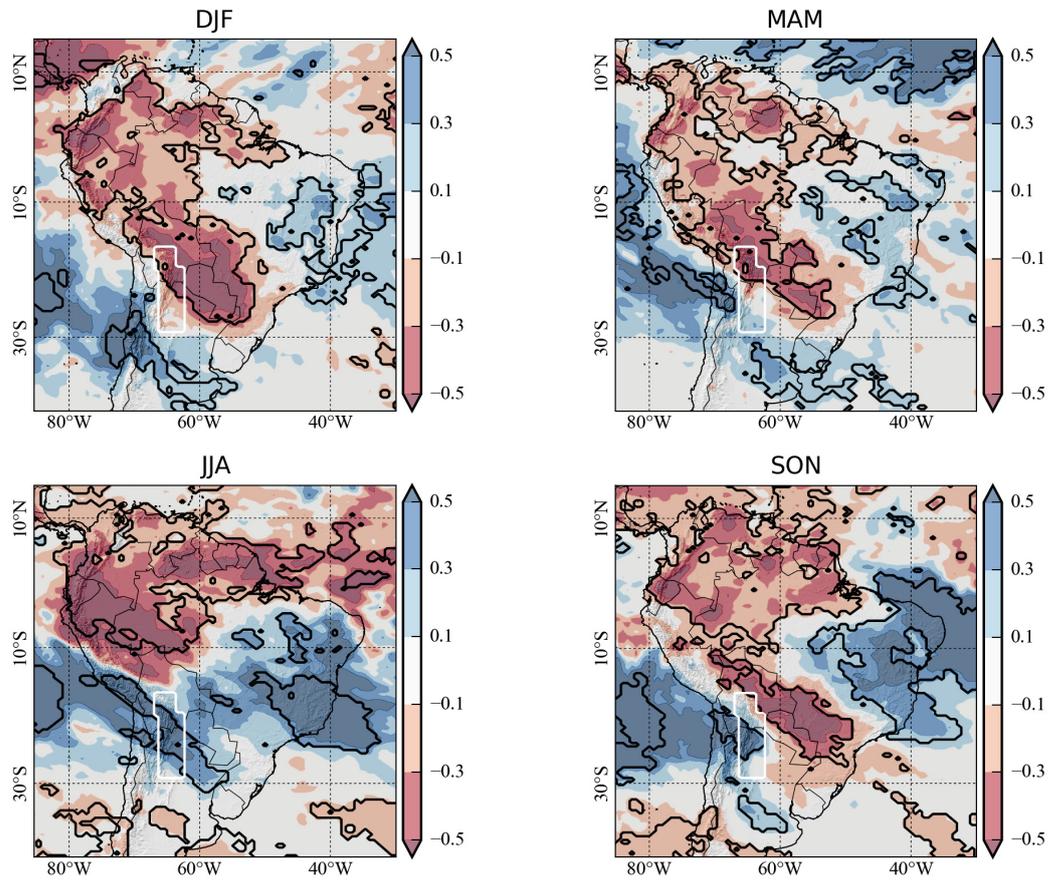
### Additional figures for chapter 7



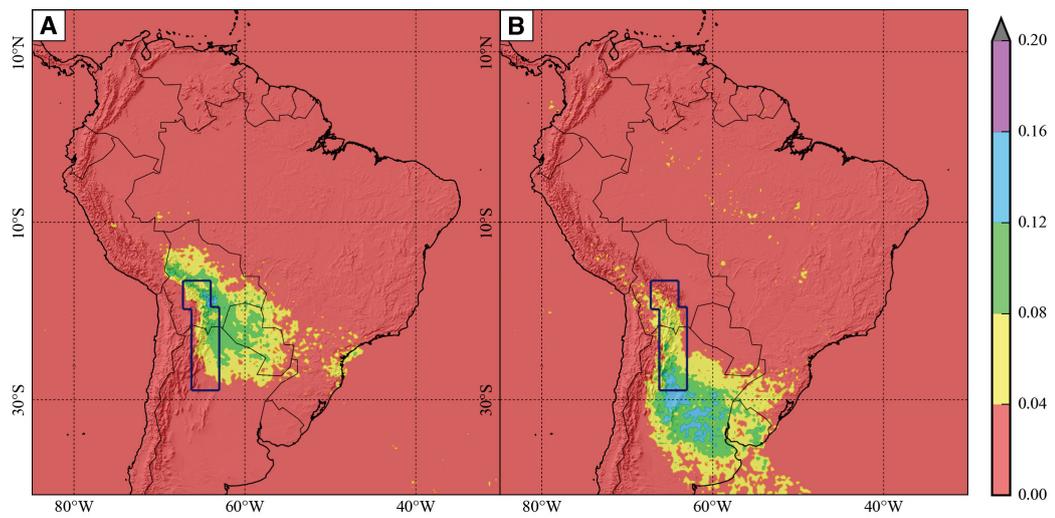
**Figure D.1.:** Linear trends for the average magnitude of the strongest 5 events averaged over boxes 6 and 7 in Figure 1A during each DJF season for rainfall derived from the gauge-calibrated TRMM 3B42 V7 (blue solid line) and outgoing longwave radiation (OLR) derived from MERRA for the period from 1979 to 2013 (red solid line) and for comparison for the period from 1998 to 2012 (red dashed line). The lower outgoing longwave radiation, the higher are corresponding rainfall amounts.



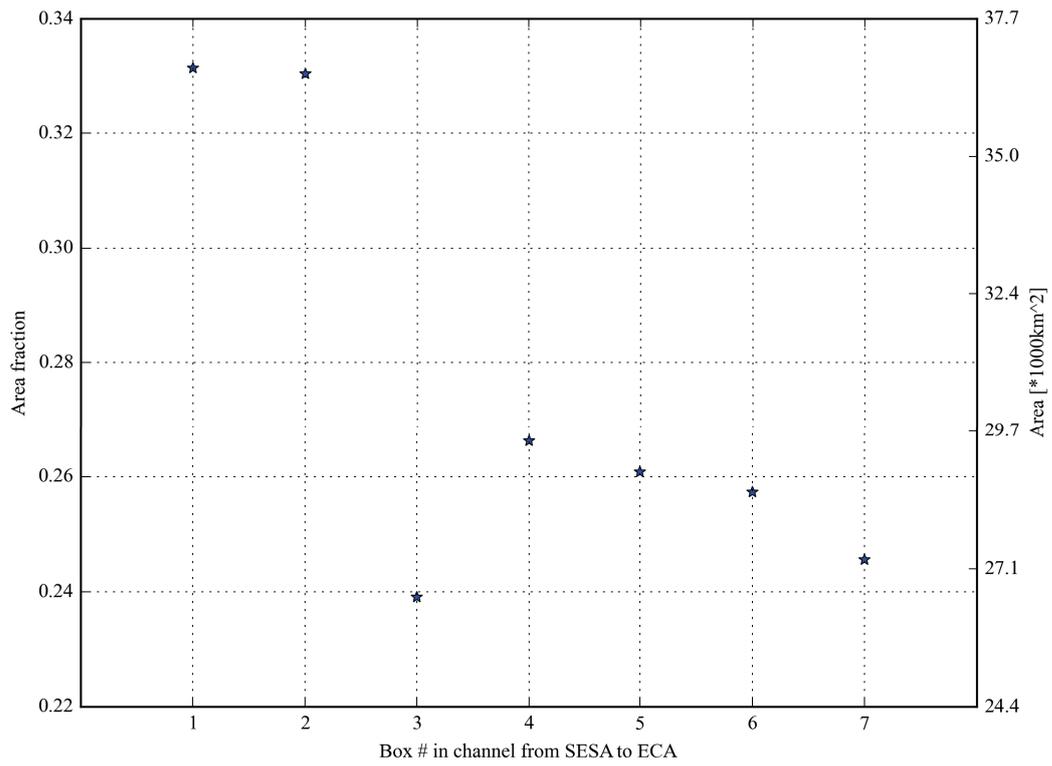
**Figure D.2.:** Linear trends for the number of extreme events per season for MERRA outgoing longwave radiation (OLR, 252 events in total) for all four three-month seasons. Areas for which results are significant at the 0.05 significance level are marked by the black contour line. The lower outgoing longwave radiation, the higher are corresponding rainfall amounts.



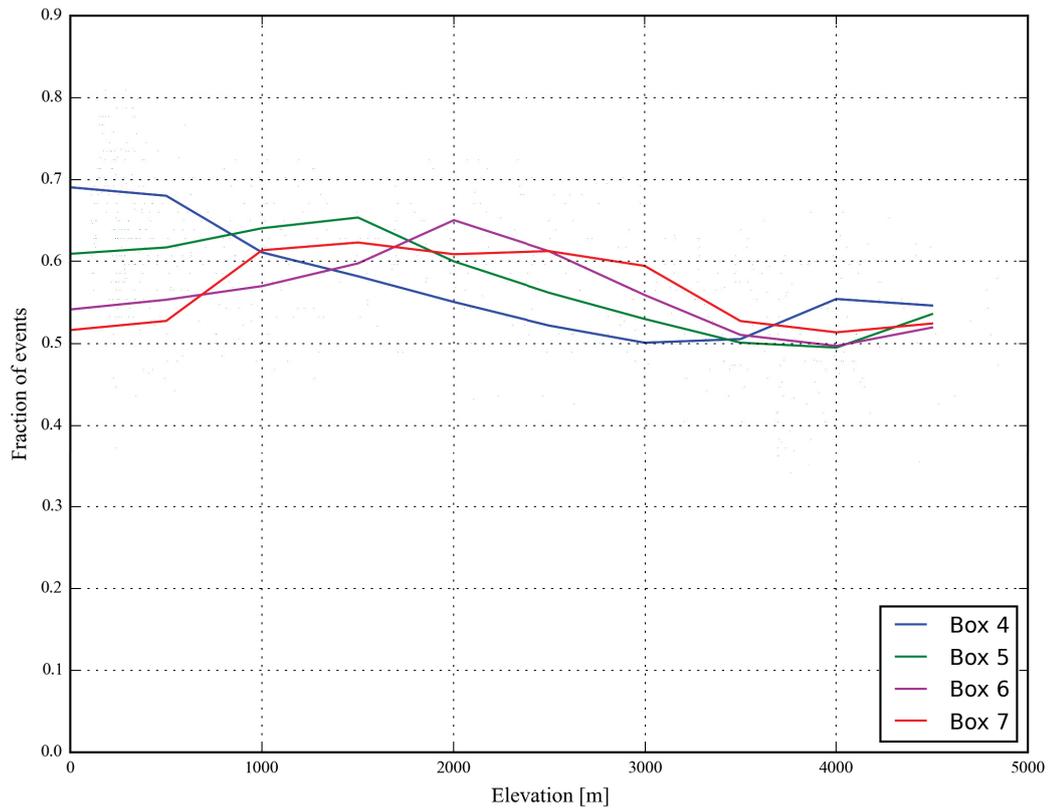
**Figure D.3.:** Trends for the average magnitude of the strongest 5 events per season for MERRA outgoing longwave radiation (OLR) for all four three-month seasons. Areas for which results are significant at the 0.05 significance level are marked by the black contour line.



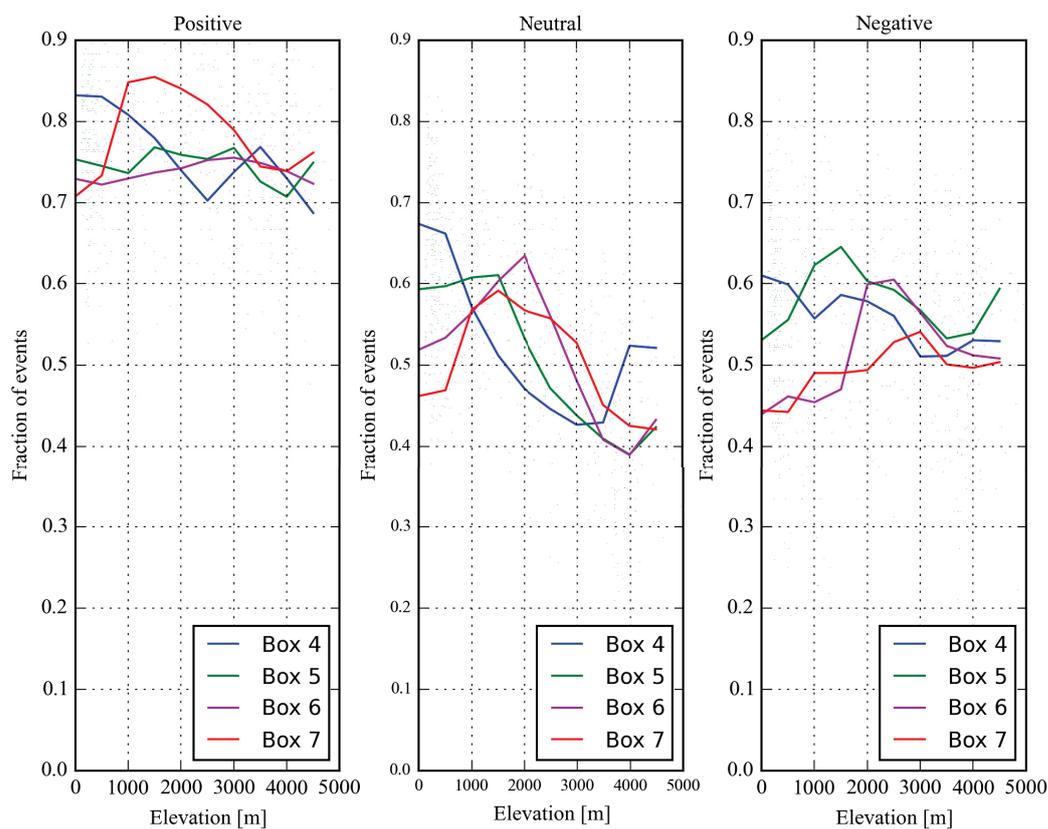
**Figure D.4.:** A. Strength out of the eastern Central Andes (ECA),  $S_i^{out}(ECA)$ . We emphasize that there are no high values over southeastern South America (SESA). B. Strength into ECA,  $S_i^{in}(ECA)$ . Note in particular the high values over SESA. Derived from the gauge-calibrated TRMM 3B42 V7.



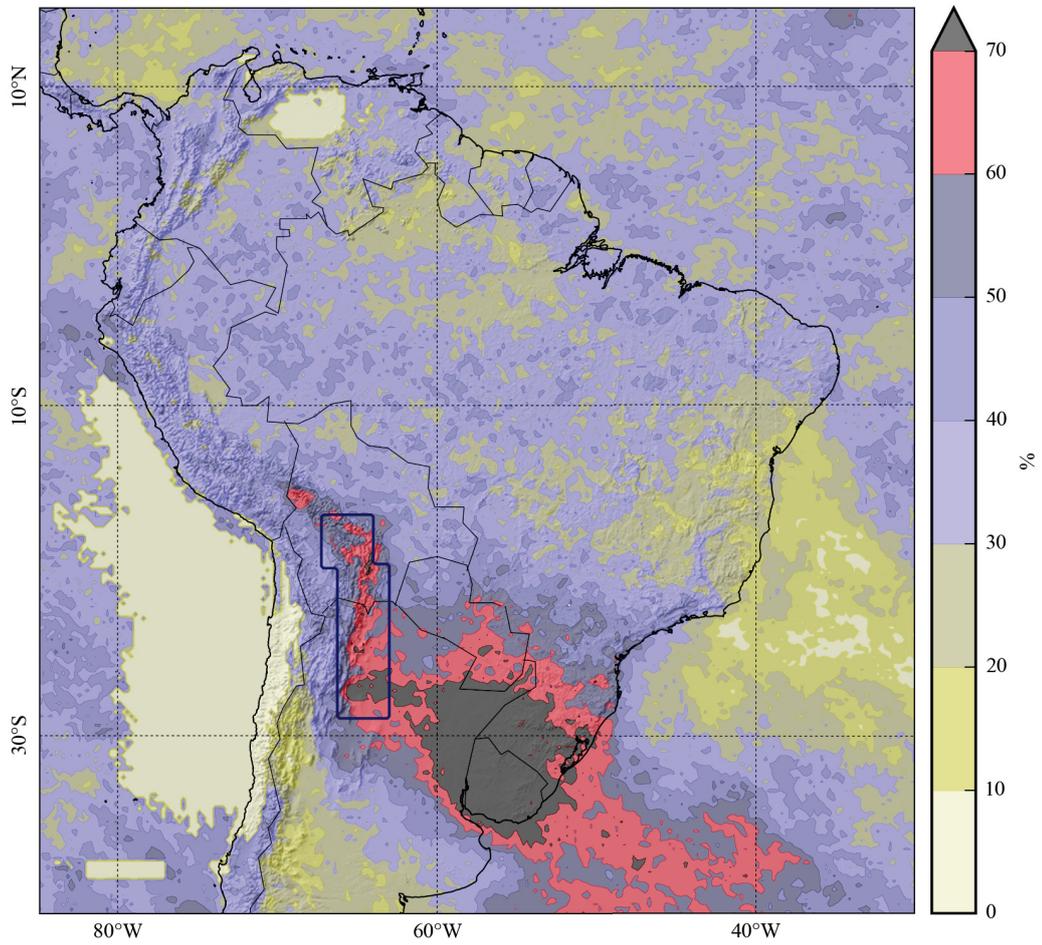
**Figure D.5.:** Average spatial extent of area receiving extreme events during prediction times and subsequent two days as fraction of box size and in km<sup>2</sup>. Derived from the (near) real-time product TRMM 3B42 V7 RT.



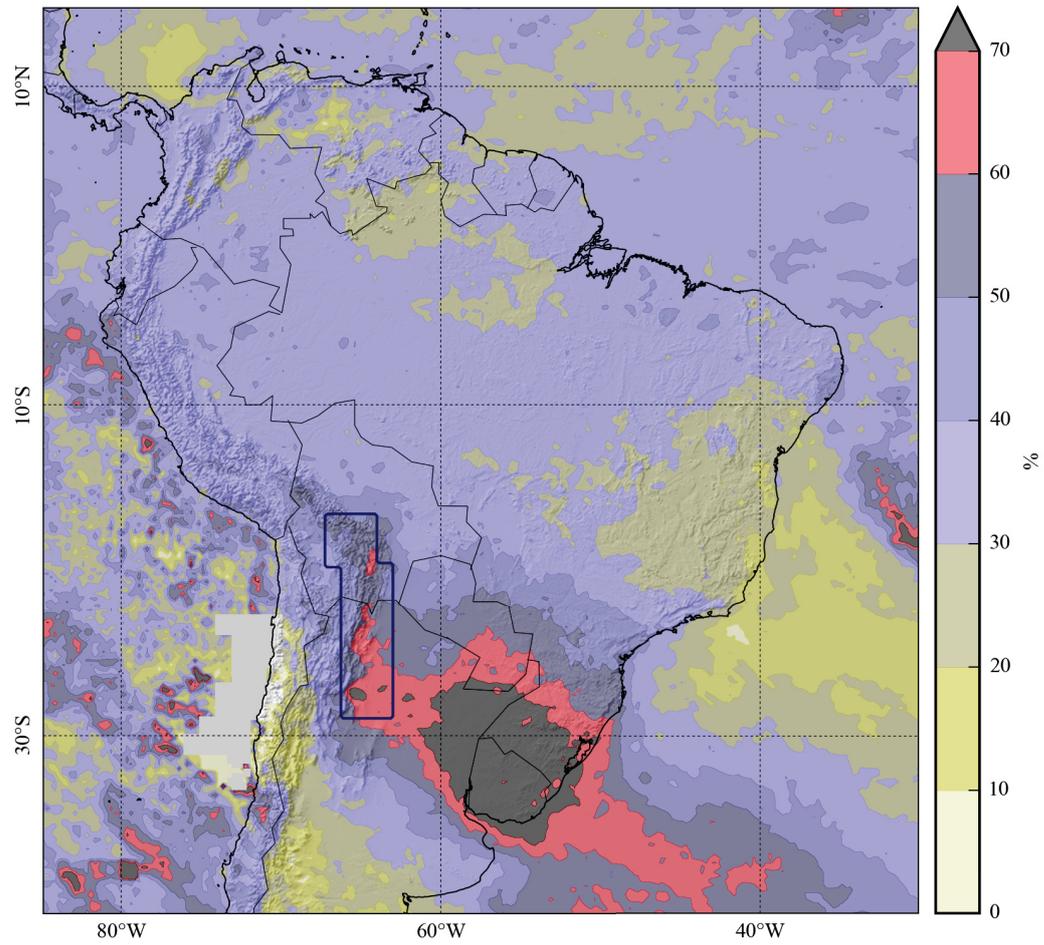
**Figure D.6.:** Relationship between elevation (meters above sea level) and relative fraction of extreme events during prediction times, spatially averaged over boxes four to seven in the ECA (see Figure 1A in the main text), for all DJF seasons. Derived from the (near) real-time product TRMM 3B42 V7 RT.



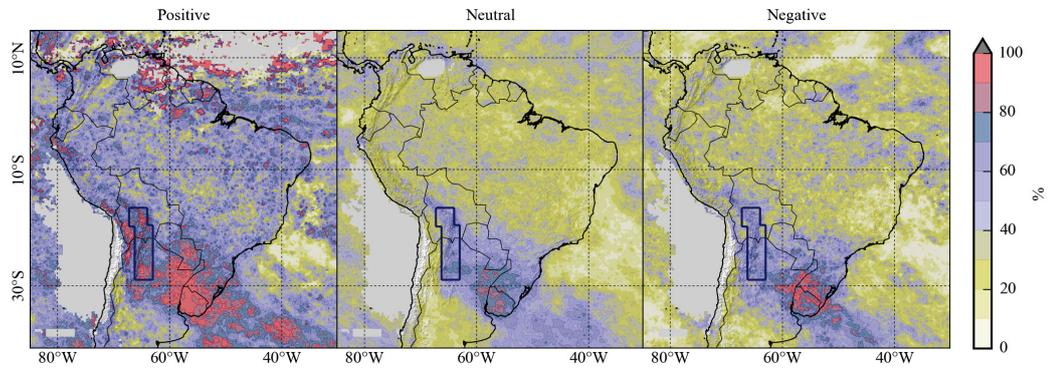
**Figure D.7.:** Same as Figure D.6 but for DJF seasons with positive (a), neutral (b) and negative (c) ENSO conditions based on the Multivariate ENSO index (MEI), retrieved from <https://climatedataguide.ucar.edu/climate-data/multivariate-enso-index>, cf. (Wolter and Timlin, 1998).



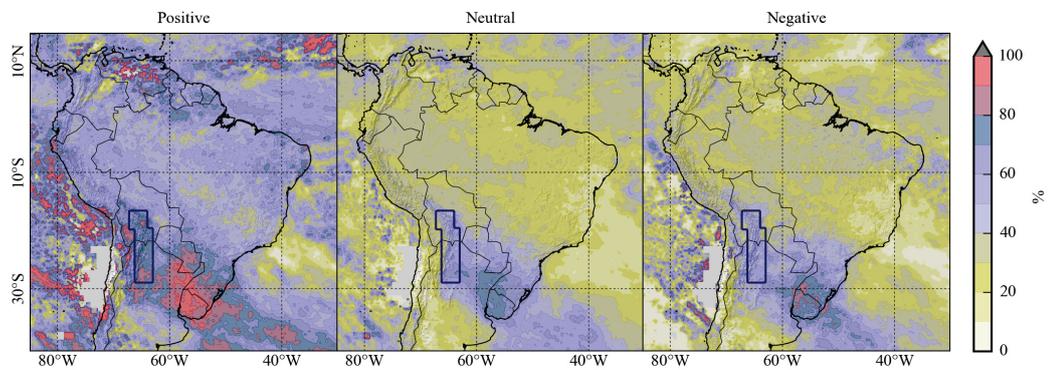
**Figure D.8.:** Relative fraction of extreme events (above 99th percentile) during prediction times and subsequent two days for all DJF seasons. Derived from the (near) real-time product TRMM 3B42 V7 RT.



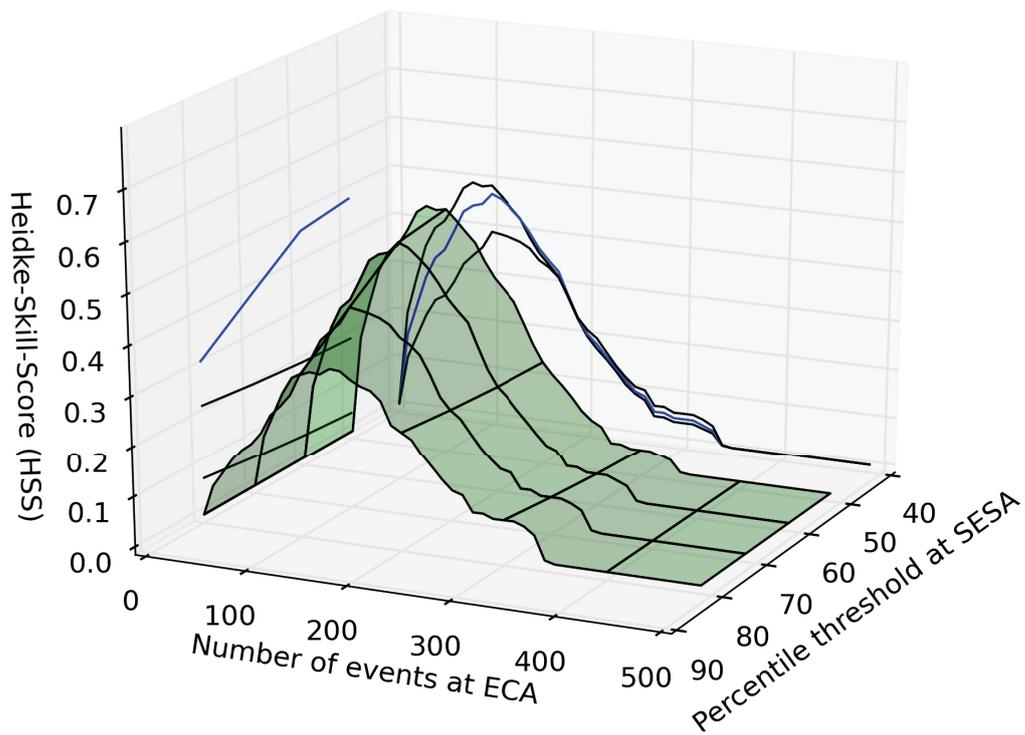
**Figure D.9.:** Relative fraction of total DJF rainfall during prediction times and subsequent two days. Derived from the (near) real-time product TRMM 3B42 V7 RT.



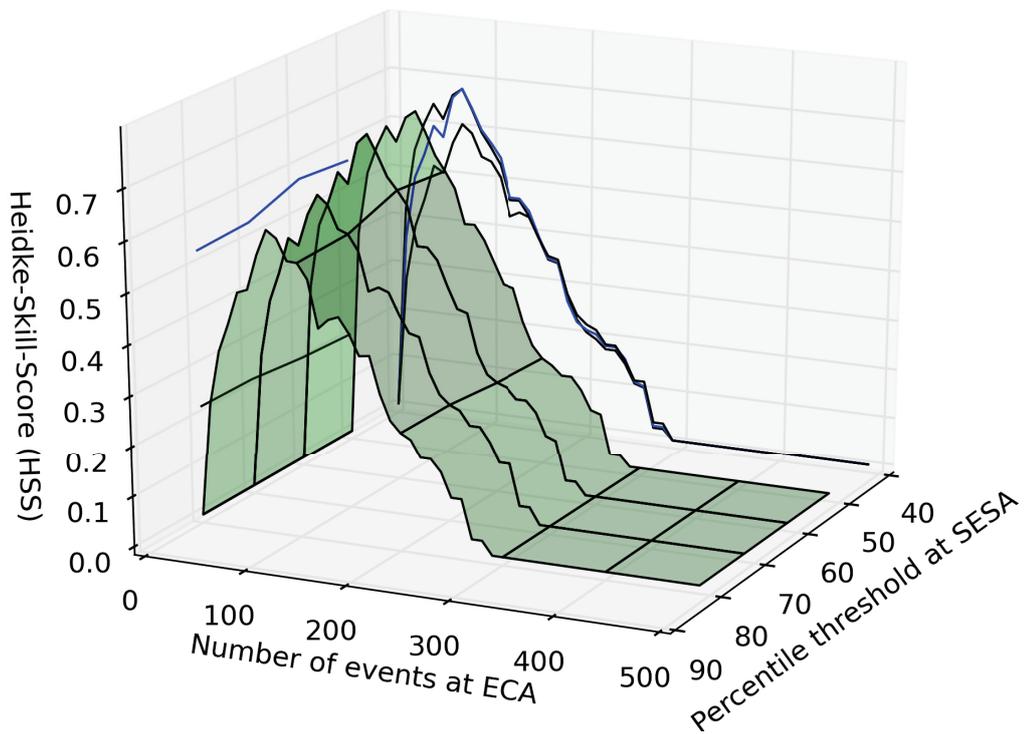
**Figure D.10.:** Same as Figure D.8 but for DJF seasons with positive, neutral, and negative ENSO conditions based on the Multivariate ENSO index (MEI).



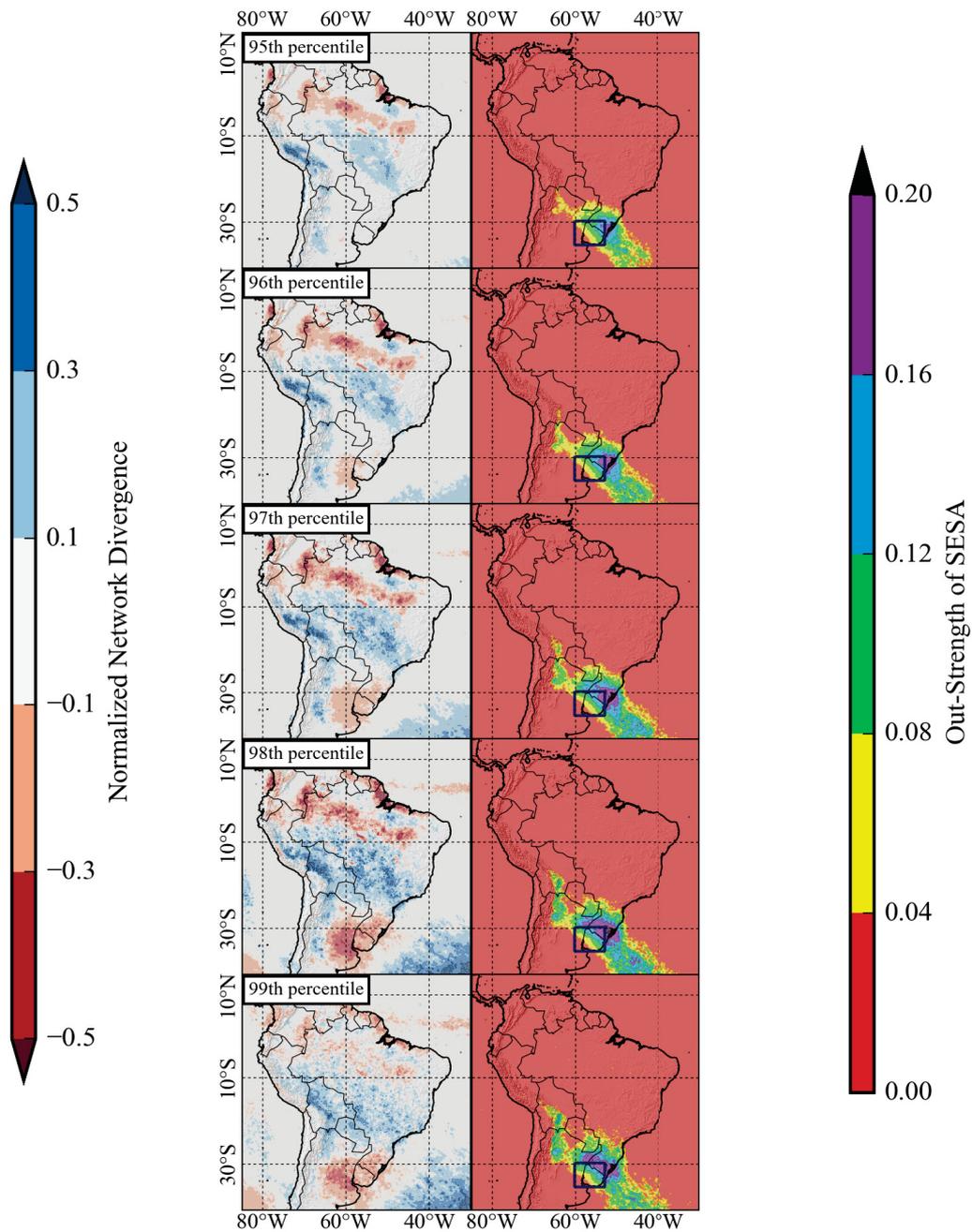
**Figure D.11.:** Same as Figure D.9 but for DJF seasons with positive, neutral, and negative ENSO conditions based on the Multivariate ENSO index (MEI).



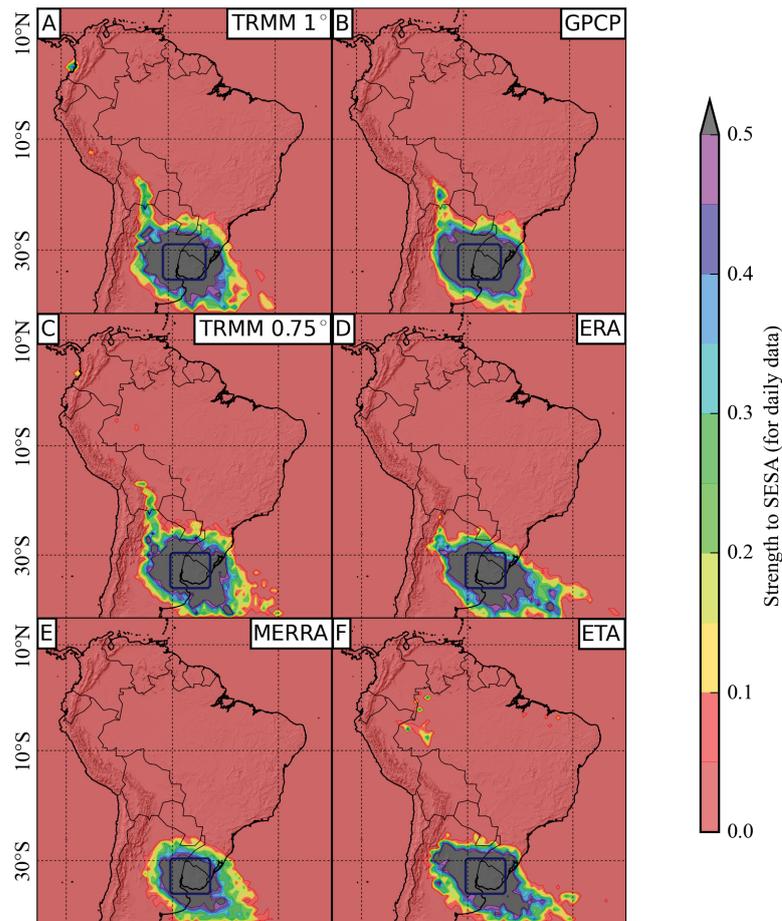
**Figure D.12.:** The Heidke-Skill-Score (HSS) for all ENSO conditions as a function of the threshold for the definition of SESA times and the number of extreme events in the ECA that are to be predicted. Derived from the (near) real-time product TRMM 3B42 V7 RT.



**Figure D.13.:** Same as Figure D.12 but for positive ENSO conditions based on the Multivariate ENSO index (MEI).

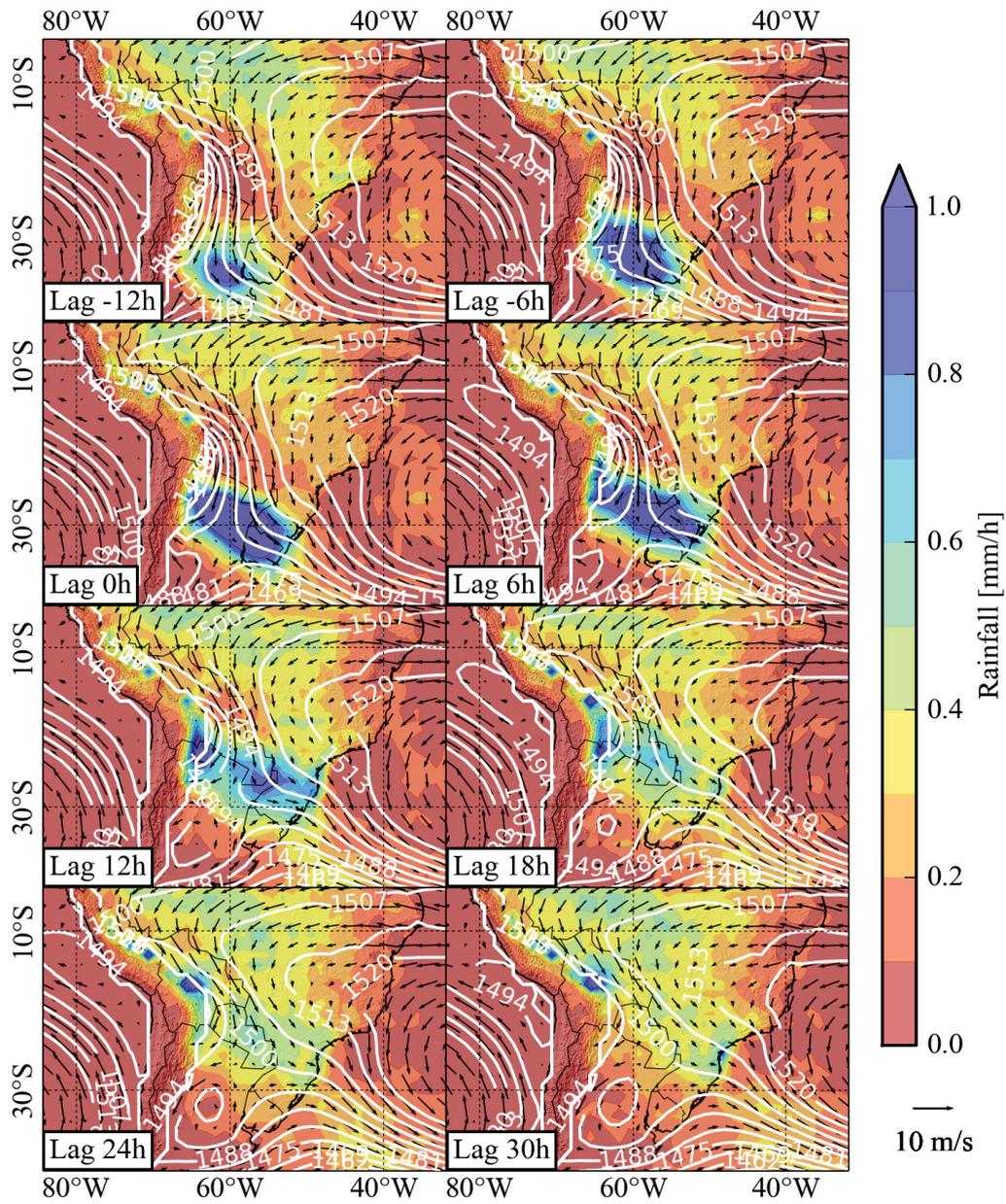


**Figure D.14.:** Network divergence  $\Delta S$  (left) and Strength out of SESA  $S^{out}(SESAs)$  (right) for the 95th to 99th percentiles as rainfall event thresholds. Note that the pattern indicating the propagation of extreme rainfall from SESA to ECA only appears for the 97th and higher percentiles.



**Figure D.15.:** Comparison of the sum of Strength into and out of SESA  $\mathcal{S}(SESA)$  for five different datasets: TRMM 3B42 V7, the satellite-gauge combination of the Global Precipitation Climatology Project (GPCP 1DD V1.2), the European Centre for Medium-Range Weather Forecasts Interim Reanalysis (ERA-interim), MERRA, as well as for output of a control run of the Brazilian ETA-CPTEC regional climate model (driven with boundary conditions from ERA-interim), which is used for operational weather forecast in South America. Since the available time period and temporal resolution varies among the different datasets, we resampled all datasets to daily values from 1998 to 2008, and summed the strength into and out of SESA for a maximal delay of  $\tau_{max} = 5$  days. TRMM is compared to GPCP on a grid with horizontal resolution of  $1^\circ$ , while the comparison to the other datasets is carried out on a horizontal resolution of  $0.75^\circ$ . For TRMM (on both horizontal grid resolutions) as well as GPCP, the propagation of extreme events from SESA to ECA is clearly recognizable. However, the reanalysis products ERA and MERRA do not show this propagation, indicating that this climatological feature could not have been found using only these datasets. Furthermore, the ETA-CPTEC model does not show the propagation neither, suggesting that weather forecast performed with this model will not be able to predict these events. If the propagation signal is not present for this sum of In- and Out-Strength using daily data, it could not be present for the Out-Strength computed for higher temporal resolutions, since in the latter case only a subset of values of Event Synchronization is summed over.

SESA-ECA: Rainfall, 850mbar GPH and Winds



**Figure D.16.:** Composites of rainfall and geopotential height and wind fields at 850mbar showing the propagation from SESA to ECA from 12 hours before to 30 hours after rainfall peaks at SESA.



# Bibliography

- Anabor, Vagner, David J. Stensrud, and Osvaldo L. L. de Moraes (Aug. 2008). “Serial Upstream-Propagating Mesoscale Convective System Events over Southeastern South America”. In: *Monthly Weather Review* 136.8, pp. 3087–3105. ISSN: 0027-0644. DOI: 10.1175/2007MWR2334.1.
- Arraut, Josefina M. and Henrique M. J. Barbosa (Oct. 2009). “Large scale features associated with strong frontogenesis in equivalent potential temperature in the South American subtropics east of the Andes”. In: *Advances in Geosciences* 22, pp. 73–78. ISSN: 1680-7359. DOI: 10.5194/adgeo-22-73-2009.
- Arraut, Josefina Moraes, Carlos Nobre, Henrique M. J. Barbosa, Guillermo Obregon, and José Marengo (Jan. 2012). “Aerial Rivers and Lakes: Looking at Large-Scale Moisture Transport and Its Relation to Amazonia and to Subtropical Rainfall in South America”. In: *Journal of Climate* 25.2, pp. 543–556. ISSN: 0894-8755. DOI: 10.1175/2011JCLI4189.1.
- Ashok, Karumuri, Swadhin K Behera, Suryachandra A Rao, Hengyi Weng, and Toshio Yamagata (2007). “El Niño Modoki and its possible teleconnection”. In: *Journal of Geophysical Research – Oceans* 112.C11.
- Barros, Vicente R., Robin Clarke, and Pedro Silva Días (2006). *Climate Change in the La Plata Basin*.
- Barros, Vicente R., Moira E. Doyle, and Inés A. Camilloni (2008). “Precipitation trends in southeastern South America: relationship with ENSO phases and with low-level circulation”. In: *Theoretical and Applied Climatology* 93.1, pp. 19–33.
- 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.
- Berbery, Ernesto H. and Vicente R. Barros (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.
- Boers, Niklas, Bodo Bookhagen, Norbert Marwan, Jürgen Kurths, and José Marengo (June 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. ISSN: 00948276. DOI: 10.1002/grl.50681.
- Bookhagen, Bodo and Manfred R. Strecker (Mar. 2008). “Orographic barriers, high-resolution TRMM rainfall, and relief variations along the eastern Andes”. In: *Geophysical Research Letters* 35.6, p. L06403. ISSN: 0094-8276. DOI: 10.1029/2007GL032011.

## Bibliography

- Bookhagen, Bodo and Manfred R. Strecker (2010). “Modern Andean rainfall variation during ENSO cycles and its impact on the Amazon drainage basin”. In: *Amazonia, Landscape and Species Evolution: A Look into the Past*.
- (2012). “Spatiotemporal trends in erosion rates across a pronounced rainfall gradient: Examples from the southern Central Andes”. In: *Earth and Planetary Science Letters* 327-328, pp. 97–110. ISSN: 0012821X. DOI: 10.1016/j.epsl.2012.02.005.
- Carvalho, L M V, C Jones, and B 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.
- (2004). “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.
- Carvalho, Leila M. V., Charles Jones, Adolfo N. D. Posadas, Roberto Quiroz, Bodo Bookhagen, and Brant Liebmann (July 2012). “Precipitation Characteristics of the South American Monsoon System Derived from Multiple Datasets”. In: *Journal of Climate* 25.13, pp. 4600–4620. ISSN: 0894-8755. DOI: 10.1175/JCLI-D-11-00335.1.
- Cazes-Boezio, Gabriel, Andrew W. Robertson, and Carlos R. Mechoso (2003). “Seasonal dependence of ENSO teleconnections over South America and relationships with precipitation in Uruguay”. In: *Journal of Climate* 16, pp. 1159–1176. ISSN: 08948755. DOI: 10.1175/1520-0442(2003)16<1159:SDOETO>2.0.CO;2.
- Chen, Sheng, Yang Hong, Jonathan J Gourley, George J Huffman, Yudong Tian, Qing Cao, Bin Yong, Pierre-Emmanuel Kirstetter, Junjun Hu, Jill Hardy, Zhe Li, Sadiq I Khan, and Xianwu Xue (2013). “Evaluation of the successive V6 and V7 TRMM multisatellite precipitation analysis over the Continental United States”. In: *Water Resources Research* 49.12, pp. 8174–8186. ISSN: 1944-7973. DOI: 10.1002/2012WR012795.
- Chou, Sin Chan, José A. Marengo, André A. Lyra, Gustavo Sueiro, José F. Pesquero, Lincoln M. Alves, Gillian Kay, Richard Betts, Diego J. Chagas, Jorge L. Gomes, Josiane F. Bustamante, and Priscila Tavares (Feb. 2011). “Downscaling of South America present climate driven by 4-member HadCM3 runs”. In: *Climate Dynamics* 38.3-4, pp. 635–653. ISSN: 0930-7575. DOI: 10.1007/s00382-011-1002-8.
- 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.
- Coppus, R and A C Imeson (2002). “Extreme events controlling erosion and sediment transport in a semi-arid sub-andean valley”. In: *Earth Surface Processes and Landforms* 27, pp. 1365–1375. ISSN: 0197-9337. DOI: 10.1002/esp.435.
- Davidson, Eric A., Alessandro C. de Araújo, Paulo Artaxo, Jennifer K. Balch, I. Foster Brown, Mercedes M. C. Bustamante, Michael T. Coe, Ruth S. DeFries, Michael Keller, Marcos Longo, J. William Munger, Wilfrid Schroeder, Britaldo S. Soares-Filho, Carlos M. Souza, and Steven C. Wofsy (2012). “The Amazon basin in transition”. In: *Nature* 481.7381, pp. 321–328. ISSN: 1476-4687. DOI: 10.1038/nature10717.

- Dee, D P, S M Uppala, A J Simmons, Paul Berrisford, P Poli, S Kobayashi, U Andrae, M A Balsameda, G Balsamo, P Bauer, P Bechtold, A C M Beljaars, L Van De Berg, J Bidlot, N Bormann, C Delsol, R Dragani, M Fuentes, A J Geer, L Haimberger, S B Healy, H Hersbach, E V Hólm, L Isaksen, P Kållberg, M Köhler, M Matricardi, A P McNally, B M Monge-Sanz, J J Morcrette, B K Park, C Peubey, P De Rosnay, C Tavolato, J N Thépaut, and F Vitart (2011). “The ERA-Interim reanalysis: configuration and performance of the data assimilation system”. In: *Quarterly Journal of the Royal Meteorological Society* 137, pp. 553–597.
- Dommenget, Dietmar and Mojib Latif (2002). “A Cautionary Note on the Interpretation of EOFs”. In: *Journal of Climate* 15.2, pp. 216–225.
- Donges, Jonathan F., Yong Zou, Norbert Marwan, and Jürgen Kurths (2009a). “Complex networks in climate dynamics”. In: *The European Physical Journal-Special Topics* 174.1, pp. 157–179.
- (2009b). “The backbone of the climate network”. In: *EPL (Europhysics Letters)* 87.4, p. 48007.
- Durkee, J D, T L Mote, and J M Shepherd (2009). “The contribution of mesoscale convective complexes to rainfall across subtropical South America”. In: *Journal of Climate* 22.17, pp. 4590–4605.
- Durkee, Joshua D. and Thomas L. Mote (2009). “A climatology of warmseason mesoscale convective complexes in subtropical South America”. In: *International Journal of Climatology* 30.3, pp. 418–431. DOI: 10.1002/joc.
- Eltahir, Elfatih A. B. and Rafael L. Bras (July 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.
- Farr, Tom G., Paul A. Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, Mimi Paller, Ernesto Rodriguez, Ladislav Roth, David Seal, Scott Shaffer, Joanne Shimada, Jeffrey Umland, Marian Werner, Michael Oskin, Douglas Burbank, and Douglas E. Alsdorf (2007). “The shuttle radar topography mission”. In: *Reviews of Geophysics* 45.2. ISSN: 87551209. DOI: 10.1029/2005RG000183.
- Garreaud, René D. (2000a). “Cold air incursions over subtropical South America: Mean structure and dynamics”. In: *Monthly Weather Review* 128.7, pp. 2544–2559.
- (Sept. 2000b). “Intraseasonal Variability of Moisture and Rainfall over the South American Altiplano”. In: *Monthly Weather Review* 128.9, pp. 3337–3346. ISSN: 0027-0644. DOI: 10.1175/1520-0493(2000)128<3337:IVOMAR>2.0.CO;2.
- (2009). “The Andes climate and weather”. In: *Advances In Geosciences* 22, pp. 1–9. ISSN: 16807340. DOI: 10.5194/adgeo-22-3-2009.
- Garreaud, René D. and Patricio Aceituno (2001). “Interannual Rainfall Variability over the South American Altiplano”. In: *Journal of Climate* 14.12, pp. 2779–2789.
- Garreaud, René D. and John M. Wallace (1997). “The diurnal march of convective cloudiness over the Americas”. In: *Monthly Weather Review* 125.12, pp. 3157–3171.
- (1998). “Summertime incursions of midlatitude air into subtropical and tropical South America”. In: *Monthly Weather Review* 126.10, pp. 2713–2733.

## Bibliography

- Garreaud, René D., Mathias Vuille, and Amy C. Clement (2003). “The climate of the Altiplano: observed current conditions and mechanisms of past changes”. In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 194.1, pp. 5–22.
- Gastner, Michael T. and M. E. J. Newman (2006a). “Optimal design of spatial distribution networks”. In: *Physical Review E* 74, p. 016117. ISSN: 15393755. DOI: 10.1103/PhysRevE.74.016117. arXiv: 0603278 [cond-mat].
- (2006b). “The spatial structure of networks”. In: *The European Physical Journal B* 49, pp. 247–252. ISSN: 1434-6028. DOI: 10.1140/epjb/e2006-00046-8. arXiv: 0407680 [cond-mat].
- Gozolchiani, a., S. Havlin, and K. Yamasaki (Sept. 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.
- Gozolchiani, Avi, Kazuko Yamasaki, Oded Gazit, and Shlomo Havlin (2008). “Pattern of climate network blinking links follows El Niño events”. In: *EPL (Europhysics Letters)* 83.2, p. 28005.
- Griffiths, Peter G., Christopher S. Magirl, Robert H. Webb, Erik Pytlak, Peter a. Troch, and Steve W. Lyon (July 2009). “Spatial distribution and frequency of precipitation during an extreme event: July 2006 mesoscale convective complexes and floods in southeastern Arizona”. In: *Water Resources Research* 45.7, W07419. ISSN: 0043-1397. DOI: 10.1029/2008WR007380.
- Grimm, Alice M. and Renata G. Tedeschi (Apr. 2009). “ENSO and Extreme Rainfall Events in South America”. In: *Journal of Climate* 22.7, pp. 1589–1609. ISSN: 0894-8755. DOI: 10.1175/2008JCLI2429.1.
- Harden, Carol P. (2006). “Human impacts on headwater fluvial systems in the northern and central Andes”. In: *Geomorphology* 79, pp. 249–263. ISSN: 0169555X. DOI: 10.1016/j.geomorph.2006.06.021.
- Hastenrath, Stefan and Leon Heller (1977). “Dynamics of climatic hazards in Northeast Brazil”. In: *Quarterly Journal of the Royal Meteorological Society* 103, pp. 77–92. ISSN: 1477870X. DOI: 10.1002/qj.49710343505.
- Hendon, Harry H., Eunpa Lim, Guomin Wang, Oscar Alves, and Debra Hudson (2009). “Prospects for predicting two flavors of El Niño”. In: *Geophysical Research Letters* 36.19, p. L19713.
- Hertwig, Eileen, Jin-song Von Storch, Irina Fast, and Thomas Krismer. “Effect of horizontal resolution on ECHAM6-AMIP performance”. In: *Climate Dynamics (under review)*.
- Hill, K. J., A. S. Taschetto, and M. H. England (2011). “Sensitivity of South American summer rainfall to tropical Pacific Ocean SST anomalies”. In: *Geophysical Research Letters* 38. ISSN: 00948276. DOI: 10.1029/2010GL045571.
- Hill, Khalia J., Andréa. S. Taschetto, and Mathew. H. England (2009). “South American rainfall impacts associated with inter-El Niño variations”. In: *Geophysical Research Letters* 36. ISSN: 00948276. DOI: 10.1029/2009GL040164.
- Hlinka, Jaroslav, David Hartman, Martin Vejmelka, Jakob Runge, Norbert Marwan, Jürgen Kurths, and Milan Paluš (2013). “Reliability of inference of directed climate networks using conditional mutual information”. In: *Entropy* 15.6, pp. 2023–2045.

- Hlinka, Jaroslav, David Hartman, Martin Vejmelka, Dagmar Novotná, and Milan Paluš (2014). “Non-linear dependence and teleconnections in climate data: sources, relevance, nonstationarity”. In: *Climate Dynamics* 42.7-8, pp. 1873–1886.
- 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.
- Houston, John (Nov. 2006a). “Evaporation in the Atacama Desert: An empirical study of spatio-temporal variations and their causes”. In: *Journal of Hydrology* 330.3-4, pp. 402–412. ISSN: 00221694. DOI: 10.1016/j.jhydro1.2006.03.036.
- (2006b). “Variability of precipitation in the Atacama Desert: its causes and hydrological impact”. In: *International Journal of Climatology* 26.15, pp. 2181–2198. ISSN: 08998418. DOI: 10.1002/joc.
- Houze Jr., Robert A. (2012). “Orographic effects on precipitating clouds”. In: *Reviews of Geophysics* 2011, pp. 1–47. DOI: 10.1029/2011RG000365.1. INTRODUCTION.
- Hu, Zeng-Zhen, Arun Kumar, Bhaskar Jha, Wanqiu Wang, Bohua Huang, and Boyin Huang (2012). “An analysis of warm pool and cold tongue El Niños: air-sea coupling processes, global influences, and recent trends”. In: *Climate Dynamics* 38.9-10, pp. 2017–2035.
- 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. Hydrometeorol* 8.1, pp. 38–55. DOI: 10.1175/JHM560.1.
- Huffman, George J., Robert F. Adler, Mark M. Morrissey, David T. Bolvin, Scott Curtis, Robert Joyce, Brad McGavock, and Joel Susskind (2001). “Global Precipitation at One-Degree Daily Resolution from Multisatellite Observations”. In: *Journal of Hydrometeorology* 2, pp. 36–50. ISSN: 1525-755X. DOI: 10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2.
- Jones, Charles, Jon Gottschalck, Leila M. V. Carvalho, and Wayne Higgins (2011). “Influence of the Madden–Julian Oscillation on Forecasts of Extreme Precipitation in the Contiguous United States”. In: *Monthly Weather Review* 139.2, pp. 332–350. ISSN: 0027-0644. DOI: 10.1175/2010MWR3512.1.
- Jorgetti, Tatiana, Pedro Leite da Silva Dias, and Edmilson Dias de Freitas (Nov. 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.
- Khan, Shiraj, Gabriel Kuhn, Auroop R. Ganguly, David J. Erickson, and George Ostrouchov (Nov. 2007). “Spatio-temporal variability of daily and weekly precipitation extremes in South America”. In: *Water Resources Research* 43.11, pp. 1–25. ISSN: 00431397. DOI: 10.1029/2006WR005384.
- Kiladis, George N. and Klaus M. Weickmann (1992). “Extratropical Forcing of Tropical Pacific Convection during Northern Winter”. In: *Monthly Weather Review* 120.9, pp. 1924–1939.

## Bibliography

- Kim, WonMoo, Sang-Wook Yeh, Joo-Hong Kim, Jong-Seong Kug, and MinHo Kwon (2011). “The unique 2009/2010 El Niño event: A fast phase transition of warm pool El Niño to La Niña”. In: *Geophys. Res. Lett.* 38.15, p. L15809.
- Kosaka, Yu and Shang-Ping Xie (Aug. 2013). “Recent global-warming hiatus tied to equatorial Pacific surface cooling”. In: *Nature* 501.7467, pp. 403–7. ISSN: 1476-4687. DOI: 10.1038/nature12534.
- Laing, Arlene G. and Michael J. Fritsch (1997). “The global population of mesoscale convective complexes”. In: *Quarterly Journal of the Royal Meteorological Society* 123.538, pp. 389–405.
- Lean, J. and D. A. Warrilow (1989). “Simulation of the regional climatic impact of Amazon deforestation”. In: *Nature* 342, pp. 411–413. ISSN: 0028-0836. DOI: 10.1038/342411a0.
- Lenters, John D. and Kerry H. Cook (1999). “Summertime precipitation variability over South America: Role of the large-scale circulation”. In: *Monthly Weather Review* 127.3, pp. 409–431.
- Lenton, Timothy M, Hermann Held, Elmar Kriegler, Jim W Hall, Wolfgang Lucht, Stefan Rahmstorf, and Hans Joachim Schellnhuber (2008). “Tipping elements in the Earth’s climate system”. In: *Proceedings of the National Academy of Sciences* 105.6, pp. 1786–1793. DOI: 10.1073/pnas.0705414105.
- Lettau, Heinz, Katharina Lettau, and Luiz Carlos B. Molion (1979). “Amazonia’s Hydrologic Cycle and the Role of Atmospheric Recycling in Assessing Deforestation Effects”. In: *Monthly Weather Review* 107.3, pp. 227–238.
- Li, Wenhong, Pengfei Zhang, Jiansheng Ye, Laifang Li, and Paul A. Baker (2011). “Impact of two different types of El Nino events on the Amazon climate and ecosystem productivity”. In: *Journal of Plant Ecology* 4, pp. 91–99. ISSN: 1752-9921. DOI: 10.1093/jpe/rtq039.
- Liebmann, B, G N Kiladis, C S Vera, A C Saulo, and L 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.
- Liebmann, Brant, George N. Kiladis, J Marengo, Tercio Ambrizzi, and John D. Glick (1999). “Submonthly convective variability over South America and the South Atlantic convergence zone”. In: *Journal of Climate* 12.7, pp. 1877–1891.
- 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.
- Maddox, Robert A. (1980). “Mesoscale convective complexes”. In: *Bulletin of the American Meteorological Society* 61.11, pp. 1374–1387.
- Malik, Nishant, Bodo Bookhagen, Norbert Marwan, and Jürgen Kurths (Aug. 2012). “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.

- 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, A. C. Saulo, J. Noguez-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.
- Marengo, J. A., S. Chou, C. Mourao, S. Solman, E. Sanchez, P. Samuelsson, R. P. da Rocha, L. Li, N. Pessacg, A. R. C. Remedio, A. F. Carril, I. F. Cavalcanti, and D. Jacob (2013a). “Simulation of rainfall anomalies leading to the 2005 drought in Amazonia using the CLARIS LPB regional climate models”. In: *Climate Dynamics* 41, pp. 2937–2955. ISSN: 09307575. DOI: 10.1007/s00382-013-1919-1.
- Marengo, José A. (2006). “On the Hydrological Cycle of the Amazon Basin: A Historical Review and Current State-of-the-art Jose Antonio Marengo”. In: *Revista Brasileira de Meteorologia* 21, pp. 1–19.
- Marengo, José A., Javier Tomasella, and Cintia R. Uvo (1998). “Trends in streamflow and rainfall in tropical South America: Amazonia, eastern Brazil, and northwestern Peru”. In: *Journal of Geophysical Research* 103.D2, pp. 1775–1783. ISSN: 0148-0227. DOI: 10.1029/97JD02551.
- 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., Carlos Nobre, Javier Tomasella, Marcos Oyama, Gilvan Sampaio de Oliveira, Rafael de Oliveira, Helio Camargo, Lincoln Alves, and Foster Brown (Feb. 2008). “The Drought of Amazonia in 2005”. In: *Journal of Climate* 21.3, pp. 495–516. ISSN: 0894-8755. DOI: 10.1175/2007JCLI1600.1.
- Marengo, José A., Richard Jones, Lincoln M. Alves, and Maria Valverde (2009). “Future change of temperature and precipitation extremes in South America as derived from the PRECIS regional climate modeling system”. In: *International Journal of Climatology* 29.15, pp. 2241–2255. DOI: 10.1002/joc.
- Marengo, José A., Maria Valverde, and G Obregon (2013b). “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.
- Matsuyama, Hiroshi, José A. Marengo, Guillermo Obregon, and Carlos A. Nobre (2002). “Spatial and temporal variabilities of rainfall in tropical South America as derived from Climate Prediction Center merged analysis of precipitation”. In: *International journal of climatology* 22, pp. 175–195. ISSN: 0899-8418. DOI: 10.1002/joc.724.
- Meehl, Gerald A, Julie M Arblaster, John T Fasullo, Aixue Hu, and Kevin E Trenberth (Oct. 2011). “Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods”. In: *Nature Clim. Change* 1.7, pp. 360–364. ISSN: 1758-678X.
- Mesinger, Fedor (1984). “A blocking technique for representation of mountains in atmospheric models”. In: *Riv. Meteor. Aeronautica* 44, pp. 195–202.

## Bibliography

- Mesinger, Fedor and Thomas L. Black (1992). "On the Impact on Forecast Accuracy of the Step-Mountain ( Eta ) vs . Sigma Coordinate". In: *Meteorology and Atmospheric Physics* 60, pp. 47–60.
- Mesinger, Fedor, Sin Chan Chou, Jorge L. Gomes, Dusan Jovic, Paulo Bastos, Josiane F. Bustamante, Lazar Lazic, André a. Lyra, Sandra Morelli, Ivan Ristic, and Katarina Veljovic (Feb. 2012). "An upgraded version of the Eta model". In: *Meteorology and Atmospheric Physics* 116.3-4, pp. 63–79. ISSN: 0177-7971. DOI: 10.1007/s00703-012-0182-z.
- Messerli, Bruno, Martin Grosjean, and Mathias Vuille (1997). "Water availability, protected areas, and natural resources in the Andean desert altiplano". In: *Mountain Research and Development* 17.3, pp. 229–238.
- Monahan, Adam H., John C. Fyfe, Maarten H. P. 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. ISSN: 0894-8755. DOI: 10.1175/2009JCLI3062.1.
- Moreiras, Stella M. (2005a). "Climatic effect of ENSO associated with landslide occurrence in the Central Andes, Mendoza Province, Argentina". In: *Landslides*. Vol. 2, pp. 53–59. ISBN: 1612-510X. DOI: 10.1007/s10346-005-0046-4.
- (2005b). *Climatic effect of ENSO associated with landslide occurrence in the Central Andes, Mendoza Province, Argentina*. DOI: 10.1007/s10346-005-0046-4.
- Negrón Juárez, Robinson I., Wenhong Li, Rong Fu, Katia Fernandes, and Andrea de Oliveira Cardoso (2009). "Comparison of Precipitation Datasets over the Tropical South American and African Continents". In: *Journal of Hydrometeorology* 10, pp. 289–299. ISSN: 1525-755X. DOI: 10.1175/2008JHM1023.1.
- Newman, Mark (2010). *Networks: An Introduction*. 1st ed. Oxford University Press, USA. ISBN: 0199206651.
- 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.
- Nieto-Ferreira, Rosana, Thomas M. Rickenbach, and Emily a. Wright (Apr. 2011). "The role of cold fronts in the onset of the monsoon season in the South Atlantic convergence zone". In: *Quarterly Journal of the Royal Meteorological Society* 137.657, pp. 908–922. ISSN: 00359009. DOI: 10.1002/qj.810.
- Nogués-Paegle, J and K C Mo (1997). "Alternating wet and dry conditions over South America during summer". In: *Monthly Weather Review* 125.2, pp. 279–291.
- O’Hare, Greg and Sara Rivas (2005). "The landslide hazard and human vulnerability in La Paz City, Bolivia". In: *Geographical Journal* 171, pp. 239–258. ISSN: 0016-7398. DOI: 10.1111/j.1475-4959.2005.00163.x.
- Paluš, Milan, David Hartman, Jaroslav Hlinka, and Martin Vejmelka (2011). "Discerning connectivity from dynamics in climate networks". In: *Nonlinear Processes in Geophysics* 18, pp. 751–763. ISSN: 16077946. DOI: 10.5194/npg-18-751-2011.

- Papalexiou, Simon Michael and Demetris Koutsoyiannis (Jan. 2013). “Battle of extreme value distributions: A global survey on extreme daily rainfall”. In: *Water Resources Research* 49.1, pp. 187–201. ISSN: 00431397. DOI: 10.1029/2012WR012557.
- Pesquero, José Fernando, Sin Chan Chou, Carlos Afonso Nobre, and José Antonio Marengo (Mar. 2009). “Climate downscaling over South America for 1961–1970 using the Eta Model”. In: *Theoretical and Applied Climatology* 99.1-2, pp. 75–93. ISSN: 0177-798X. DOI: 10.1007/s00704-009-0123-z.
- Poveda, Germán, Liliana Jaramillo, and Luisa F. Vallejo (Jan. 2014). “Seasonal precipitation patterns along pathways of South American low-level jets and aerial rivers”. In: *Water Resources Research* 50.1, pp. 98–118. ISSN: 00431397. DOI: 10.1002/2013WR014087.
- Programa de las Naciones Unidas para el Desarrollo (PNUD) (2011). *Tras las huellas del cambio climático en Bolivia*, p. 46.
- Quian Quiroga, R, T Kreuz, and P Grassberger (2002). “Event synchronization: A simple and fast method to measure synchronicity and time delay patterns”. In: *Phys. Rev. E* 66.4, p. 41904.
- Radebach, Alexander, Reik V Donner, Jakob Runge, Jonathan F Donges, and Jürgen Kurths (2013). “Disentangling different types of El Niño episodes by evolving climate network analysis”. In: *Physical Review E* 88, p. 052807. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.88.052807.
- Rasmusson, Eugene M. and Thomas H. Carpenter (1982). “Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño”. In: *Monthly Weather Review* 110, pp. 354–384. DOI: [http://dx.doi.org/10.1175/1520-0493\(1982\)110<0354:VITSST>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1982)110<0354:VITSST>2.0.CO;2).
- Renard, B., K. Kochanek, M. Lang, F. Garavaglia, E. Paquet, L. Neppel, K. Najib, J. Carreau, P. Arnaud, Y. Aubert, F. Borchi, J.-M. Soubeyroux, S. Jourdain, J.-M. Veysseire, E. Sauquet, T. Cipriani, and A. Auffray (Feb. 2013). “Data-based comparison of frequency analysis methods: A general framework”. In: *Water Resources Research* 49.2, pp. 825–843. ISSN: 00431397. DOI: 10.1002/wrcr.20087.
- 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, 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.

## Bibliography

- Romatschke, Ulrike and Robert A. Houze (Feb. 2013). "Characteristics of Precipitating Convective Systems Accounting for the Summer Rainfall of Tropical and Subtropical South America". In: *Journal of Hydrometeorology* 14.1, pp. 25–46. ISSN: 1525-755X. DOI: 10.1175/JHM-D-12-060.1.
- Ropelewski, C.F. and M.S. Halpert (1987). "Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation". In: *Monthly Weather Review* 115.8, pp. 1606–1626.
- Salio, P, M Nicolini, and A C Saulo (2002). "Chaco low-level jet events characterization during the austral summer season". In: *Journal of Geophysical Research* 107.D24, p. 4816.
- Salio, Paola, Matilde Nicolini, and Edward J. Zipser (Apr. 2007). "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.
- Saulo, A. Celeste, Marcelo E. Seluchi, and Matilde Nicolini (2004). "A case study of a Chaco low-level jet event". In: *Monthly Weather Review* 132.11, pp. 2669–2683.
- Schuster, Robert L, Daniel A Salcedo, and Luis Valenzuela (2002). "Overview of catastrophic landslides of South America in the twentieth century". In: *Catastrophic landslides: effects, occurrence, and mechanisms, reviews in engineering geology. Geological Society of America, Boulder, CO*, pp. 1–33.
- 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.
- Seluchi, Marcelo E. and A. Celeste Saulo (2003). "The northwestern Argentinean low: A study of two typical events". In: *Monthly Weather Review* 131.10, pp. 2361–2378.
- Serinaldi, Francesco and Chris G Kilsby (Jan. 2014). "Rainfall extremes: Toward reconciliation after the battle of distributions". In: *Water Resources Research* 50.1, pp. 336–352. ISSN: 0043-1397. DOI: 10.1002/2013WR014211.
- Shukla, Jagadish, Carlos A. Nobre, and Piers J. Sellers (1990). "Amazon Deforestation and Climate Change". In: *Science* 247.4948, pp. 1322–1325.
- Silva, Gyrlene A. M., Tercio Ambrizzi, and José A. Marengo (Feb. 2009). "Observational evidences on the modulation of the South American Low Level Jet east of the Andes according the ENSO variability". In: *Annales Geophysicae* 27.2, pp. 645–657. ISSN: 1432-0576. DOI: 10.5194/angeo-27-645-2009.
- Silva, Viviane B. S., Vernon E. Kousky, and R. Wayne Higgins (2011). "Daily Precipitation Statistics for South America: An Intercomparison between NCEP Reanalyses and Observations". In: *Journal of Hydrometeorology* 12, pp. 101–117. ISSN: 1525-755X. DOI: 10.1175/2010JHM1303.1.
- 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.
- Smith, James A., Mary Lynn Baeck, Alexandros A. Ntelekos, Gabriele Villarini, and Matthias Steiner (Apr. 2011). "Extreme rainfall and flooding from orographic

- thunderstorms in the central Appalachians". In: *Water Resources Research* 47.4, pp. 1–24. ISSN: 00431397. DOI: 10.1029/2010WR010190.
- Solman, Silvina A., Mario N. Nuñez, and María Fernanda Cabré (Sept. 2007). "Regional climate change experiments over southern South America. I: present climate". In: *Climate Dynamics* 30.5, pp. 533–552. ISSN: 0930-7575. DOI: 10.1007/s00382-007-0304-3.
- Solman, Silvina A., E. Sanchez, P. Samuelsson, R. P. da Rocha, L. Li, J. Marengo, N. L. Pessacg, A. R C Remedio, S. C. Chou, H. Berbery, H. Le Treut, M. de Castro, and D. Jacob (2013). "Evaluation of an ensemble of regional climate model simulations over South America driven by the ERA-Interim reanalysis: Model performance and uncertainties". In: *Climate Dynamics* 41, pp. 1139–1157. ISSN: 09307575. DOI: 10.1007/s00382-013-1667-2.
- Steinhaeuser, Karsten, Auroop R. Ganguly, and Nitesh V. Chawla (June 2012). "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.
- Stevens, Bjorn, Marco Giorgetta, Monika Esch, Thorsten Mauritsen, Traute Crueger, Sebastian Rast, Marc Salzmann, Hauke Schmidt, Jürgen Bader, Karoline Block, Renate Brokopf, Irina Fast, Stefan Kinne, Luis Kornblueh, Ulrike Lohmann, Robert Pincus, Thomas Reichler, and Erich Roeckner (June 2013). "Atmospheric component of the MPI-M Earth System Model: ECHAM6". In: *Journal of Advances in Modeling Earth Systems* 5.2, pp. 146–172. ISSN: 19422466. DOI: 10.1002/jame.20015.
- Tedeschi, Renata G., Iracema F A Cavalcanti, and Alice M. Grimm (2013). "Influences of two types of ENSO on South American precipitation". In: *International Journal of Climatology* 33, pp. 1382–1400. ISSN: 08998418. DOI: 10.1002/joc.3519.
- Trenberth, KE (1997). "The definition of el nino". In: *Bull. Amer. Meteor. Soc.* 78, pp. 2771–2777. DOI: [http://dx.doi.org/10.1175/1520-0477\(1997\)078<2771:TDOENO>2.0.CO;2](http://dx.doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2).
- Trenberth, KE and DP Stepaniak (2001). "Indices of El Niño evolution". In: *Journal of Climate* 14, pp. 1697–1701. DOI: [http://dx.doi.org/10.1175/1520-0442\(2001\)014<1697:LIOENO>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(2001)014<1697:LIOENO>2.0.CO;2).
- Tsonis, A A and P J Roebber (2004). "The architecture of the climate network". In: *Physica A* 333, pp. 497–504.
- Tsonis, Anastasios and Kyle Swanson (June 2008). "Topology and Predictability of El Niño and La Niña Networks". In: *Physical Review Letters* 100.22, p. 228502. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.100.228502.
- Tsonis, Anastasios, Kyle Swanson, and Paul Roebber (2006). "What Do Networks Have to Do with Climate?" In: *Bulletin of the American Meteorological Society* 87.5, pp. 585–595. DOI: 10.1175/BAMS-87-5-585.
- Tsonis, Anastasios, Kyle Swanson, and Sergey Kravtsov (2007). "A new dynamical mechanism for major climate shifts". In: *Geophysical Research Letters* 34, p. L13705. DOI: 10.1029/2007GL030288.
- Urrutia, Rocío and Mathias Vuille (Jan. 2009). "Climate change projections for the tropical Andes using a regional climate model: Temperature and precipitation

## Bibliography

- simulations for the end of the 21st century". In: *Journal of Geophysical Research* 114.D2, p. D02108. ISSN: 0148-0227. DOI: 10.1029/2008JD011021.
- 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.
- Vera, Carolina S. (2002). "Cold season synoptic-scale waves over subtropical South America". In: *Monthly Weather Review* 130.3, pp. 684–699.
- Vuille, Mathias (1999). "Atmospheric circulation over the Bolivian Altiplano during dry and wet periods and extreme phases of the Southern Oscillation". In: *International Journal of Climatology* 19.14, pp. 1579–1600.
- Wilks, Daniel S. (2006). "Statistical Methods in the Atmospheric Sciences". In: *International Geophysics Series* 102, pp. 380–380. ISSN: 19326203. DOI: 10.1198/jasa.2007.s163.
- Wolter, Klaus and Michael S. Timlin (1993). "Monitoring ENSO in COADS with a seasonally adjusted principal component index". In: *Proc. of the 17th Climate Diagnostics Workshop, Norman, OK, NOAA/NMC/CAC, NSSL, Oklahoma Clim. Survey, CIMMS and the School of Meteor., Univ. of Oklahoma*, pp. 52–57.
- (1998). "Measuring the strength of ENSO events - how does 1997/98 rank?" In: *Weather* 53, pp. 315–324.
- Xue, Xianwu, Yang Hong, Ashutosh S Limaye, Jonathan J Gourley, George J Huffman, Sadiq Ibrahim Khan, Chhimi Dorji, and Sheng Chen (2013). "Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins?" In: *Journal of Hydrology* 499.0, pp. 91–99. ISSN: 0022-1694. DOI: <http://dx.doi.org/10.1016/j.jhydro1.2013.06.042>.
- Yamasaki, K, A Gozolchiani, and S Havlin (2008). "Climate networks around the globe are significantly affected by El Niño". In: *Physical Review Letters* 100.22, p. 228501.
- Yeh, Sang-Wook, Jong-Seong Kug, Boris Dewitte, Min-Ho Kwon, Ben P Kirtman, and Fei-Fei Jin (Sept. 2009). "El Niño in a changing climate". In: *Nature* 461.7263, pp. 511–4. ISSN: 1476-4687. DOI: 10.1038/nature08316.
- Yin, Lei, Rong Fu, Elena Shevliakova, and Robert E. Dickinson (2013). "How well can CMIP5 simulate precipitation and its controlling processes over tropical South America?" In: *Climate Dynamics* 41, pp. 3127–3143. ISSN: 09307575. DOI: 10.1007/s00382-012-1582-y.
- 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 (Aug. 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.
- Zulkaffi, Zed, Wouter Buytaert, Christian Onof, Bastian Manz, Elena Tarnavsky, Waldo Lavado, and Jean-Loup Guyot (Dec. 2014). "A comparative performance

## *Bibliography*

analysis of TRMM 3B42 (TMPA) versions 6 and 7 for hydrological applications over Andean-Amazon river basins". In: *Journal of Hydrometeorology* 15.2, pp. 581–592. ISSN: 1525-755X. DOI: 10.1175/JHM-D-13-094.1.



# Selbständigkeitserklärung

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

Berlin, den 11. November 2014

Niklas Felix Boers