



Institute for Earth and Environmental Science

# Unraveling spatio-temporal climatic patterns via multi-scale complex networks

by
**Ing. Ankit Agarwal**

**2018**

A cumulative dissertation for the degree
Doctor of Engineering "doktoringenieur "(Dr. Ing.) in
Geoecology

Submitted to
The Faculty of Mathematics and Natural Sciences at
The University of Potsdam

Supervisors

Prof. Dr. rer. nat. Jürgen Kurths
*University of Potsdam, Institute of Earth and Environmental Science.*
*Research Domain Transdisciplinary Concepts and Methods, Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Telegrafenberg, Potsdam, Germany.*
*Institute of Physics, Humboldt Universität zu Berlin, Germany.*

Prof. Dr. Ing. Bruno Merz
*University of Potsdam, Institute of Earth and Environmental Science.*
*Section 5.4, Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences.*

Dr. Norbert Marwan
*Research Domain Transdisciplinary Concepts and Methods, Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Telegrafenberg, Potsdam, Germany.*

Submission date: 28th September 2018
Date of PhD defence:

Referees:

Prof. Dr. rer. nat. Jürgen Kurths
*University of Potsdam, Institute of Earth and Environmental Science.*
*Research Domain Transdisciplinary Concepts and Methods, Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Telegrafenberg, Potsdam, Germany.*
*Institute of Physics, Humboldt Universität zu Berlin, Germany.*

Prof. Dr. Ing. Bruno Merz
*University of Potsdam, Institute of Earth and Environmental Science.*
*Section 5.4, Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences.*

Prof. Dr. Ing. Jürgen Jensen
*Forschungsinstitut Wasser und Umwelt (fwu), Lehrstuhl Wasserbau und Hydromechanik, University of Siegen.*

Published online at the
Institutional Repository of the University of Potsdam:
https://doi.org/10.25932/publishup-42395
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-423956

Sarvadharmān parityajya mām ekaṃ śaraṇaṃ vraja;
Ahaṃ tvāṃ sarvapāpebhyo mokṣayiṣyāmi mā śucaḥ.

(–**Bhagavad Gita 2.47**)

You have a right to perform your prescribed duty, but you are not entitled to the fruits of action. Never consider yourself the cause of the results of your activities, and never be attached to not doing your duty.

# Copyright

*"Dedicated to the memory of my brother,Rahul Goyal (*14 *November* 1989-14 *September* 2017*), who always believed in my ability to be successful in the academic arena. You are gone but your belief in me has made this journey possible."*

## Abstract

The climate is a complex dynamical system involving interactions and feedbacks among different processes at multiple temporal and spatial scales. Although numerous studies have attempted to understand the climate system, nonetheless, the studies investigating the multiscale characteristics of the climate are scarce. Further, the present set of techniques are limited in their ability to unravel the multi-scale variability of the climate system. It is completely plausible that extreme events and abrupt transitions, which are of great interest to climate community, are resultant of interactions among processes operating at multi-scale. For instance, storms, weather patterns, seasonal irregularities such as El Niño, floods and droughts, and decades-long climate variations can be better understood and even predicted by quantifying their multi-scale dynamics. This makes a strong argument to unravel the interaction and patterns of climatic processes at different scales. With this background, the thesis aims at developing measures to understand and quantify multi-scale interactions within the climate system.

In the first part of the thesis, I proposed two new methods, viz, multi-scale event synchronization (MSES) and wavelet multi-scale correlation (WMC) to capture the scale-specific features present in the climatic processes. The proposed methods were tested on various synthetic and real-world time series in order to check their applicability and replicability. The results indicate that both methods (WMC and MSES) are able to capture scale-specific associations that exist between processes at different time scales in a more detailed manner as compared to the traditional single scale counterparts.

In the second part of the thesis, the proposed multi-scale similarity measures were used in constructing climate networks to investigate the evolution of spatial connections within climatic processes at multiple timescales. The proposed methods WMC and MSES, together with complex network were applied to two different datasets.

In the first application, climate networks based on WMC were constructed for the univariate global sea surface temperature (SST) data to identify and visualize the SSTs patterns that develop very similarly over time and distinguish them from those that have long-range teleconnections to other ocean regions. Further investigations of climate networks on different timescales revealed (i) various high variability and co-variability regions, and (ii) short and long-range teleconnection regions with varying spatial distance. The outcomes of the study not only re-confirmed the existing knowledge on the link between SST patterns like El Niño/ Southern Oscillation and the Pacific Decadal Oscillation, but also suggested new insights into the characteristics and origins of long-range teleconnections.

In the second application, I used the developed non-linear MSES similarity measure to quantify the multivariate teleconnections between extreme Indian precipitation and climatic patterns with the highest relevance for Indian sub-continent. The results confirmed significant non-linear influences that were not well captured by the traditional methods. Further, there was a substantial variation in the strength and nature of teleconnection across India, and across time scales.

Overall, the results from investigations conducted in the thesis strongly highlight the need for considering the multi-scale aspects in climatic processes, and the proposed methods provide robust framework for quantifying the multi-scale characteristics.

## Zusammenfassung

Das Klima ist ein komplexes Zusammenspiel verschiedener Mechanismen und Rückkopplungen auf mehreren zeitlichen und räumlichen Skalen. Viele Studien beschäftigten sich mit dem diesem System, nur wenige jedoch konzentrierten sich auf das Multiskalenverhalten des Klimas. Vor allem die bis dato verfügbaren Techniken schränkten eine vertiefte Analyse der Klimavariabilität auf unterschiedlichen Skalen ein. Von großen Interesse in der aktuellen Klimaforschung sind Extremereignisse und plötzliche Veränderungen, welche höchstwahrscheinlich aus dem Zusammenwirken von Prozessen auf unterschiedlichen Skalen hervorgehen. Um Stürme, wiederkehrende Wetterlagen, jahreszeitliche Phänomene wie El Niño, Fluten, Dürren oder Klimaschwankungen über Jahrzehnte besser zu verstehen oder sogar vorhersagen zu können, müssen wir deren Dynamik auf unterschiedlichen Skalen quantifizieren. In der vorliegenden Arbeit werden Mittel und Wege präsentiert um das Zusammenwirken auf verschiedenen Skalen im Klimasystem besser zu verstehen und zu quantifizieren.

Im ersten Teil dieser Arbeit stelle ich zwei Methoden, multi-scale event synchronization (MSES) und wavelet multi-scale correlation (WMC) vor, welche skalenspezifischen Eigenschaften in klimatischen Prozessen abbilden. Die vorgestellte Methode wurde mit mehreren synthetischen und realen Zeitreihen getestet um ihre Anwendbarkeit und Reproduzierbarkeit zu überprüfen. Die Ergebnisse zeigen, dass beide Methoden Beziehungen auf unterschiedlichen zeitlichen Skalen detaillierter als traditionelle Ansätze abbilden können.

Im zweiten Teil dieser Arbeit bilde ich klimatische Netzwerke mithilfe eines Maßes zur Ähnlichkeit auf Multiskalen. Dabei untersuche ich die Entwicklung von räumlichen Beziehungen um klimatische Prozesse auf mehreren Zeitskalen zu verstehen. Die Methoden WMC und MSES werden zusammen mit komplexen Netzwerken auf zwei Datensätze angewendet.

In der ersten Anwendung wurden Klimanetzwerke auf Basis von WMC aufgebaut für das univariate globale Meer Oberflächentemperatur (SST) Daten zur Identifizierung und Visualisierung der SST-Muster, die sich sehr ähnlich entwickeln mit Zeit und unterscheiden sie von denen, die Fernverbindungen zu anderen Meeresregionen haben. Weiter Untersuchungen von Klimanetzwerken auf verschiedenen Zeitskalen ergaben (i) verschiedene hohe Variabilität und Co-Variabilität und (ii) Kurz- und Langstrecken-Televerbindungsbereiche mit unterschiedlichem räumlichen Abstand. Die Ergebnisse der Studie bestätigten nicht nur das vorhandene Wissen über den Zusammenhang zwischen SST-Mustern wie El Niño/ Southern Oscillation und Pacific Decadal Oscillation, sondern schlug auch neue Erkenntnisse über die Merkmale und Ursprünge von Telekommunikations-Verbindungen mit großer Reichweite.

In der zweiten Anwendung verwendeten wir das von mir entwickelte, nicht-lineare MSES Ähnlichkeitsmaß um multivariate Fernbeziehungen zwischen Starkniederschlägen und klimatischen Mustern über Indien zu quantifizieren. Unsere Ergebnisse bestätigen signifikante, nicht-lineare Einflüsse, welche von traditionellen Methoden bisher unzureichend abgebildet wurden. Des Weiteren fanden wir deutliche Schwankungen in der Stärke und in der Ausprägung von Fernbeziehungen über Indien und über Zeitskalen.

Zusammenfassend zeigen die Ergebnisse dieser Fallstudien, dass Multiskalen in Klimaprozessen entschieden berücksichtigt werden müssen und dass der entwickelte methodische Rahmen adäquat die charakteristischen Prozesse quantifizieren kann.

# List of publications

This dissertation is partly based on the following publications.

### Papers

(P1) Agarwal, A., Marwan, N., Rathinasamy, M., Merz, B. and Kurths, J.: Multi-scale event synchronization analysis for unravelling climate processes: a wavelet-based approach, Nonlinear Process. Geophys., 24(4), 599–611, doi:10.5194/npg-24-599-2017, 2017.

(P2) Agarwal, A., Rathinasamy, M., Marwan, N., Caesar, L., and Kurths, J.: A wavelet-based multi-scale similarity measure for complex networks, European Physics Journal-B, https://doi.org/10.1140/epjb/e2018-90460-6.

(P3) Agarwal, A., Caesar, L., Marwan, N., Rathinasamy, M., Merz, B. and Kurths, J.: Detection of short- and long-range teleconnections in SST patterns on different timescales. (Climate dynamics-in review).

(P4) Agarwal, A., Marwan, N., Maheswaran, R., Krishnan,R., Kurths, J., and Merz, B.: Unravelling the spatial diversity of Indian rainfall teleconnections. (Scientific Reports-In review).

(P5) Agarwal, A., Marwan, N., Maheswaran, R., Merz, B. and Kurths, J.: Quantifying the roles of single stations within homogeneous regions using complex network analysis, Journal of Hydrology, doi:10.1016/j.jhydrol.2018.06.050, 2018.

(P6) Agarwal, A., Marwan, N., Rathinasamy, M., Ozturk, U., Merz, B. and Kurths, J.: Optimal Design of Hydrometric Station Networks Based on Complex Network Analysis, Hydrol. Earth Syst. Sci. Discuss., 1–21, doi:10.5194/hess-2018-113, 2018.

#### Other papers of the author in collaboration with topical similarities

(P7) Ozturk, U., Marwan, N., Korup, O., Saito, H., Agarwal, A., Grossman, M. J., Zaiki, M. and Kurths, J.: Complex networks for tracking extreme rainfall during typhoons, Chaos: An Interdisciplinary Journal of Nonlinear Science, 28(7), 075301, doi:10.1063/1.5004480, 2018a.

(P8) Ekhtiari, N., Agarwal, A., Donner, R., Marwan, N.,: Disentangling the multiscale interdependence between precipitation and sea-surface temperature: A network-of-network approach, Chaos: An Interdisciplinary Journal of Nonlinear Science, (In progress).

### Conference proceedings

(C1) Ankit Agarwal, Norbert Marwan, Rathinasamy Maheswaran, Bruno Merz, Ravi Krishnan Juergen Kurths: Unravelling Regionswise Teleconnections of Indian Rainfall Using Event Synchronization-Based Multiscale Nonlinear Method. Asian Oceanic Geosciences meeting-2018; 06/2018.

(C2) Ankit Agarwal Norbert Marwan, Rathinasamy Maheswaran, Bruno Merz, Juergen Kurths, R: Complex network-based approach for identification of influential and expandable station across rainfall network. European Geophysical Union 2017; 04/2018.

(C3) Ankit Agarwal Norbert Marwan, Rathinasamy Maheswaran, Bruno Merz, Juergen Kurths, R: Multiscale complex network analysis: An approach to study spatiotemporal rainfall pattern in south Germany. European Geophysical Union 2017; 04/2017.

(C4) Ankit Agarwal Norbert Marwan, Rathinasamy Maheswaran, Ugur Ozturk, Bruno Merz, Juergen Kurths: Multiscale event synchronization analysis for unraveling climate processes: A wavelet-based approach. American Geophysical Union 2016, San Francisco; 12/2016.

(C5) Ankit Agarwal Norbert Marwan, Rathinasamy Maheswaran, Ugur Ozturk, Bruno Merz, Juergen Kurths: Multiscale event synchronization measure: A wavelet-based approach. Perspectives in nonlinear dynamincs (PNLD) 2016, Humboldt-University Berlin, Germany; 07/2016, DOI:10.13140/RG.2.2.16589.64485.

(C6) Ankit Agarwal Rathinasamy Maheswaran, Juergen Kurths, Rakesh Khosa: Wavelet Spectrum and Self-Organizing Maps-Based Approach for Hydrologic Regionalization -a Case Study in the Western United States. Water Resources Management, European Geophysical Union-2016; 04/2016.

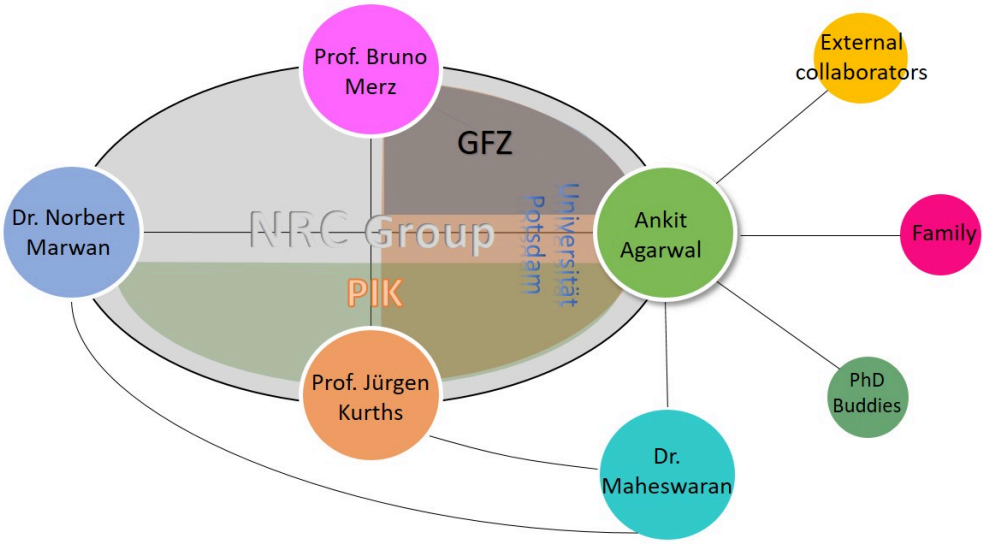
(C7) Ankit Agarwal Maheswaran Rathinasamy, Rakesh Khosa: Hydrologic Regionalization Using Wavelet Based Multi-Scale Entropy Method. AGU-2015, San Fransisco, USA; 12/2015, DOI:10.13140/RG.2.1.2155.5289.

### Book chapters

(B1) Agarwal, A., Marwan, N., Ozturk, U., & Maheswaran, R. (2019). Unfolding Community Structure in Rainfall Network of Germany Using Complex Network-Based Approach. In Water Resources and Environmental Engineering II (pp. 179-193). Springer, Singapore.

# Acknowledgements

The best part about writing a thesis is recalling the efforts of people who made it possible. I am deeply indebted to the Deutsche Forschungsgemeinschaft (German Research Foundation) via the NatRiskChange Graduate school (GREKO 2043/1) for generously supporting me through my research years. I would also like to thank the University of Potsdam for providing an amicable environment, copious infrastructure and an equally competent administration during my entire stay at Golm. I would also like to gratefully acknowledge the University of Potsdam and Potsdam Institute for Climate Impact Research (PIK) for supporting this work by providing resources on the high performance computer systems. The University of Potsdam and PIK have been the best institutions any scholar could desire.




I owe a debt of gratitude to people without whom I could not have completed this journey. While my work is based on complex networks in climatology, these networks are analogous to various important system of human connections that made this thesis possible. Foremost, I take this opportunity to express my heartfelt indebtedness to the three intellectual pillars, my PhD supervisors. I am grateful to Dr. Norbert Marwan, who in the challenging times, kept a sense of humor when I had lost mine. His indulgence created a very congenial working atmosphere. Prof. Bruno Merz was ever generous and his incredible attention to details drove me to ultimately learning the right way of punctuating proses. Despite his busy schedule, Prof. Jürgen Kurths sent detailed comments which were invaluable. I am glad that I met such a selfless and caring person who combined intellectual inspiration

with warmth. I am deeply thankful to each one of them for steering deeply stimulating discussions which led to further explorations.

I also extend my sincere thanks to my research-coach and mentor Dr. Maheswaran Rathinasamy. I am indebted to him for all his kindness and for being a constant moral and academic support.

The magnitude of my life at University of Potsdam and PIK would have only been a fraction of itself, had it not been in the presence of the NatRiskChange Graduate School. I am very grateful to Prof. Axel Bronstert for being ever-accessible and encouraging. A special thanks goes to my fellow doctoral students and buddies Dadiyorto Wendi, Ugur Ozturk, Georg Veh, Jennifer von Keyserlingk, Jonas Laudan, Tobias Sieg, Sebastian Specht, Bernhard Fiedler, Thomas Moran, Madlen Fischer, Berry Bossenkool, Erwin Rottler and Irene Crisologo for being the best graduate companions one could ask for. Their constant feedbacks, cooperation and friendship made me feel like home, miles away from home. Also, I extend my gratitude to NatRiskChange staff members, Prof. Annegret Thieken, Dr. Theresia Petrow, Dr. Kristin Vogel and Dr. Stephanie Natho for making it a smooth ride for me altogether.

In addition, I would like to acknowledge the support of my extended research family, Nikoo Ekhtiari, Dr. Bedartha Goswami, Dr. Chiranjit Mitra, Dr. Paul Schultz, Dr. Aljoscha Rheinwalt, Dr. Sabine Auer, Dr. Deniz Eroglu, Dr. Veronika Stolbova, Dr. Jonathan Donges, Dr. Jobst Heitzig, Dr. Frank Hellmann, Dr.Reik Donnar, Dr. Kai Schröter, Levke Caesar and the entire research group of Prof. Kurths and Prof. Bruno Merz.

I thank the staff of the University of Potsdam, Potsdam Institute for Climate Impact Research and GFZ, section 5.4 Hydrology for their constant backing and assistance throughout the PhD.

To my friends and family, who held me up and encouraged me throughout this journey, you value more than I can express on paper. It is difficult to name them all, but some need special mention. I owe special thanks to my PhD confidants, Niharika Tyagi, (TERI University, India), Roopam Shukla (TERI University, India), Pooja Arya (University of Potsdam) and Vinit Sehgal (Texas, A&M) who were always there with a listening ear and a word of inspiration. Several people read the thesis and suggested ways of improving it. I am thankful to Dr. Stephanie Natho, Nishitha Bajaj, and Sumedha Premi for proofreading my thesis. Their insights encouraged me to write better.

Finally, I would like to thank my entire family for supporting me financially and emotionally throughout these years. It is never easy to work away from home but my parents, Rajesh Agarwal and Rama Agarwal, and my uncle Naresh Agarwal have been most understanding and compassionate all through my stay in Berlin. They believed in me, even in times when I didn't believe in myself.

# Contents

# List of Figures

# List of Tables

# List of frequently used abbreviations

## Abbreviations

| | |
|---|---|
| Z | With-in module degree |
| P | Participation coefficient |
| Q | Strength of event synchronization |
| M | Modularity |
| WT | Wavelet transform |
| ES | Event synchronization |
| NH | Northern Hemisphere |
| TH | Tropical Hemisphere |
| SH | Southern Hemisphere |
| KE | Kriging error |
| WMC | Wavelet multiscale correlation |
| SST | Sea surface temperature |
| PCC | Pearson's correlation coefficient |
| IOD | Indian Ocean Dipole |
| NAO | North Atlantic Oscillation |
| PDO | Pacific Decadal Oscillation |
| AMO | Atlantic Multidecadal Oscillation |
| SAM | Southern Annular mode |
| DWT | Discrete wavelet transform |
| MODWT | Maximum overlap discrete wavelet transform |
| MSES | Multiscale event synchronizaion |
| ENSO | El Niño/Southern Oscillation |

# Chapter 1

# Introduction

*"Different physics at different scale"*

## 1.1 Background and motivation

Climatology, a branch of atmospheric science and a subfield of Geography, is a scientific study of climate, which is defined as a long-term (>30 years) average of a place's weather (Werndl, 2016). Climate varies naturally on a whole range of timescales, and these variations can have profound impacts on weather conditions around the world, such as heavy rainfall and heat waves (Schiermeier, 2018). It is a complex system involving many variables including air temperature, sea surface temperature, relative humidity, type and amount of precipitation, air pressure and winds. A remarkable progress had been made in the last century in terms of understanding the climate and its variables. It all started with an observation of a Chinese scientist Shen Kuo (1031–1095) who inferred that climate naturally shifted over an enormous span of time (Bowden et al., 2005). However, the understanding of climate system expanded exponentially after Helmut Landsberg (1906–1985) fostered the use of statistical analysis in climatology, which led to its evolution into physical science.

Before the $16th$ century, when the first thermoscope and barometer were invented, researchers believed that all climatic variables such as precipitation, temperature, and pressure, etc. are independent of each other and have their individual dynamics. However, soon scientists realized that climate is a coupled system, and that interactions among climatic variables account for global climate and are responsible for its variability. Since then, a multitude of studies were dedicated around the world to study each climatic variable such as precipitation (Berndtsson, 1988; Eltahir and Bras, 1996; Hu and Feng, 2002; Malik et al., 2016), pressure (Berberan-Santos et al., 1997), temperature (Miralles et al., 2014; Vecchi and Harrison, 2002), relative humidity (Putthividhya and Tanaka, 2012), sea surface temperature (Kucharski et al., 2006; Trenberth and Shea, 2005), etc. separately using various kinds of datasets (observed, satellite, reanalysis or modeled) around the world.

Even though there has been a significant progress in our understanding of the climatic processes, there remains a large scope for a holistic understanding owing to the immense complexity involved in such processes. Moreover, the recent challenges in terms of change in global climate, frequent occurrence of extreme events (rainfall, floods, and droughts) call for more advanced research in improving our existing knowledge of the climate systems. Extreme events have become the utmost concern across many regions in the world where risks due to such events have to be managed and mitigated. Hence, to investigate the cause

of extreme events-induced natural disasters and provide suitable remedial measures, research-training group (RTG) "Natural hazards and risks in a changing world (NatRiskChange)" has been formed with an aim to foster the scientific knowledge basis for these natural disasters. Key scientific aims of the RTG are identification, quantification and prediction of transient natural hazards and the associated risk. This thesis is dedicated to the quantification of global climatic processes and interactions, which could be further used for developing better prediction mechanisms.

In the pursuit of understanding the global climate and extreme events phenomena, several researchers have used different approaches to investigate and model climate systems (Steinbach et al., 2003; Steinhaeuser et al., 2012; Tsonis et al., 2007; Tsonis and Roebber, 2004). All these studies showed that climatic systems are a complex system, and are often fueled by multiple feedbacks and interactions. In such systems, information is often held, stored, and perceived differently at a different scale, resulting from mutual interactions of intertwined sub-components interacting over a wide range of scales (Miralles et al., 2014; Peters et al., 2004; Rathinasamy et al., 2014; Paluš, 2014a; Molini et al., 2010). Numerous definitions of the term "scale" exist which can have multiple connotations depending on the field of the study and hence, in the subsequent section, we briefly define "scale" with reference to climatology.

In climatology, the term "scale" has three different connotations (Fig. 1.1), namely (i) *spatial scale*, which refers to the region at which its parameters are being measured and defined. It can range from local climate processes (Jarrell et al., 2012; Shimono and Beggs, 2015; Schroeter et al., 2015; Lee et al., 2016) to regional and global climate phenomena (Bullmore and Bassett, 2011), (ii) *temporal scale*, refers to the characteristic time of the process under investigation and can range from sub-daily to days (Khambhati et al., 2015; Burns et al., 2014) to that of decades and multi-decades (Zuo et al., 2010; Betzel et al., 2014; Gu et al., 2015b) and finally, (iii) *topological scale* which is defined as an arrangement of a network (Havlin et al., 2012) and can range from individual nodes to densely interconnected subsystem (a.k.a. community/cluster) to the network as a whole (Stam and Reijneveld, 2007; Rubinov and Sporns, 2011; Bullmore and Sporns, 2009).

Collectively, these scales define the axes of a three-dimensional space in which any analysis of climate network data exists (Fig. 1.1). Most climate network analyses exist as points in this space i.e. they focus on networks defined singularly at one spatial, temporal, and topological scale. We argue that, while such multi-scale studies have proven illuminating, but in order to better understand the climate's true multi-scale, it is essential that our network analyses begin to form bridges that link different scales (temporal, spatial and topological) to one another.

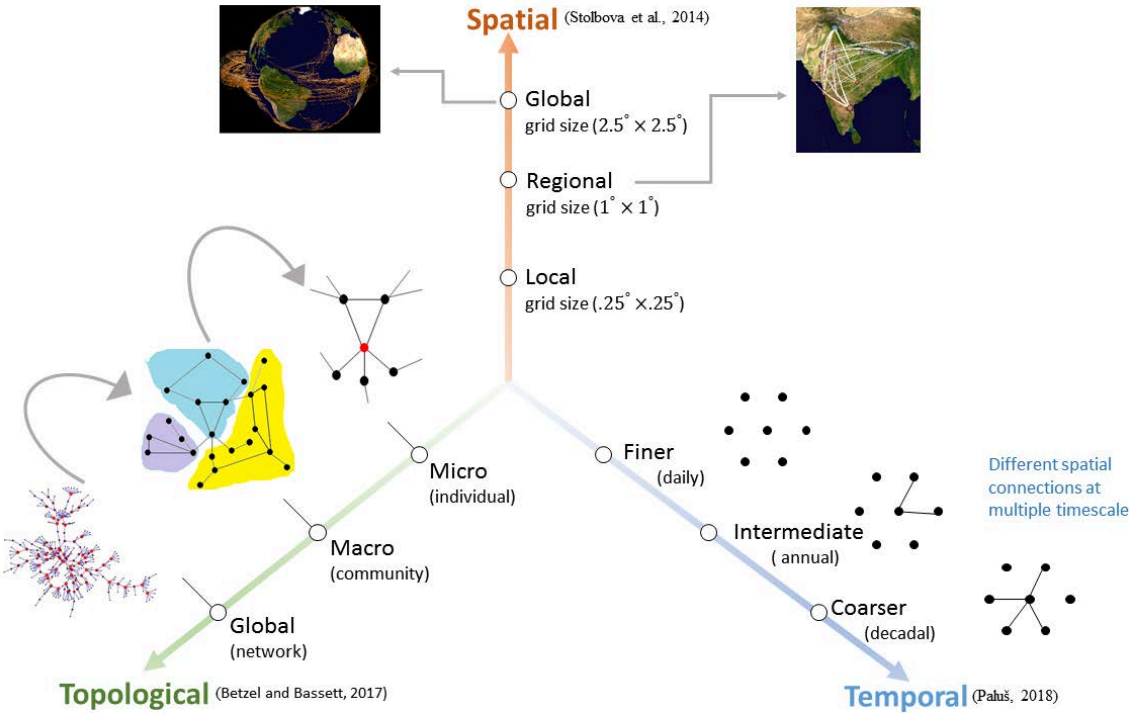



**Figure 1.1.:** Three-dimensional schematic of multi-scale in climate network. A climate network is defined as a collection of nodes (geographical location of climatological data) interconnected with links based on statistical similarity measure in a non-trivial manner. Although the term "Global" is same for spatial and topological scale, however the meaning in both the contexts is different.*(adapted from Betzel and Bassett (2017) and modified for climate network).*

In the past, several studies such as Casagrande et al. (2015), Steinhaeuser et al. (2012), Paluš (2014a), and Girard et al. (2015) evidently showed that the climatic processes occur at different scales and these processes are driven by factors operating at other scales. For instance, in case of a multiple timescale process, finer scale processes might be the result of a convective or localized phenomena whereas processes at a coarser scale might be driven by large-scale teleconnections. Also, it is reasonable to state that the process observed at finer scale indeed carries the influence of process that operates at a coarse scale. For example (Fig. 1.2), the observed daily rainfall series is a part of the seasonal cycle within any given year and carries the influence of the dynamics of a seasonal pattern. The seasonal rainfall series, in turn, is a component in the long-term cycle and may carry an influence of the underlying long-term patterns, such as a decadal features.

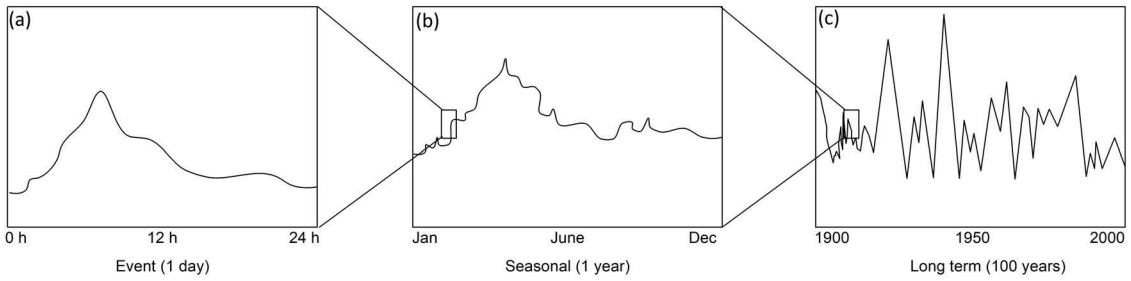



**Figure 1.2.:** Schematic of multi-timescale phenomena in climatology (adapted from Blöschl and Sivapalan (1995)) and modified for climatic processes). (a) Daily rainfall series, (b) Seasonal rainfall series, (c) Long-term rainfall series. The observed daily rainfall series is a part of the seasonal cycle within any given year and carries the influence of the dynamics of a seasonal pattern. The seasonal rainfall series, in turn, is a component in the long-term cycle and may carry the influence of an underlying long-term pattern, such as decadal features.

Hence, it is logical to expect that the coupling between climate variables takes place at different time and spatial scales. Untangling this multi-scale variability and interactions of a climatic process are vital as it would improve the understanding about global climate and its variability.

Over the years, several studies have developed methods to visualize, analyse and predict the dynamics of climatic variables, their interactions, and patterns. A significant progress has indeed been made. However, there are only a few studies dedicated to quantification of multi-scale variability in climatic processes. It is vital to note that the conventional statistical methods such as Pearson's correlation coefficient, principal component analysis and others, lack the ability to discern and identify the dominant features that are present at different scales. Individual, scale-specific, characteristic details get camouflaged at an integrated observation timescale and therefore satisfactory/holistic understanding is missed. Further, in conventional approaches to study geophysical processes, we tend to focus on one particular scale. The effects of finer scales are understood through the constitutive relation, and effects of coarser scales are neglected by assuming that the system is homogeneous at larger scales (E, 2011). Sometimes changes in the climatic system occur rapidly and their effects may appear in a short time, or sometime take a long (Lovejoy, 2015). As an illustration, for example, record-breaking high temperature year-after-year triggered rapid disappearance of Arctic sea ice and gradually increased sea level.

The philosophy of a multi-scale approach is based on the premise that any process (climate/hydrology) of interest can always be represented at various timescales with different complexities. Such a representation allows an in-depth analysis of the process when the aggregate behavior of the process at an observational scale is no longer adequate to study. Further, it also gives us a basis for understanding the scale of dominance for the process. In many situations, the process of interest can be described adequately at a coarser scale, except in some small regions where more detailed information is required. In such cases, we may get more detailed insights into the process by studying it at multiple scales.

In conclusion, it may be averred that multi-scale analysis of climatic processes holds the promise of better understanding the system dynamics that may be missed when processes are analysed at one timescale only (Peters et al., 2007; Rathinasamy and Khosa,

2012a; Casagrande et al., 2015; Steinhaeuser et al., 2012; Agarwal, 2015). Notwithstanding the progress made in understanding the climate dynamics, there is an urgent quest to investigate and quantify the climatic process at a multi-scale.

## 1.2 Scope

The scope of the current research is to investigate the multi-scale variability in the climatic process and to foster new advanced methods that have an innate capability to enable the following:

a Detection of various scale specific features that are normally associated with typical geophysical systems in general and climatic system in particular.

b Capturing linear association and patterns in the large spatial dataset at different scales.

c Capturing and quantifying non-linearity, noise contamination, scale emergence and scale dominance in the signal which are vital to comprehend climate processes at different scales.

d Spatial data visualization or data representation at multiple scales.

It is affirmed that this aforementioned potential will lead to develop methods to capture the multi-scale variability and the interaction of the process at different scales. The implied aspiration leads to a multitude of research questions that this study will attempt to investigate. The guiding research questions are as follows:

1. Do geophysical signals such as sea surface temperature, precipitation, air pressure, turbulence, hydrologic fluxes, or the El Niño–Southern Oscillation show different dynamics/variability at different timescale? The answer is yes, but then how to quantify the different timescales of dynamics?
2. Can the existing similarity measures such as Pearson's Correlation Coefficient (PCC), Event Synchronization (ES) effectively investigate the multi-scale process? If not, how can we modify the existing measure to capture the multi-scale process?
3. How to study the spatial connections of univariate multi-scale processes such as sea surface temperature (SST) variability? How to visualize and represent the spatial patterns?
4. How to quantify the linkages in the multi-variate process such as linkages between dominant climatic patterns and extreme precipitation? How to visualize and represent the spatial patterns?

A review of literature around these research questions has provided a way to proceed forward.

Recent developments in wavelet theory have presented newer approaches to quantify multi-scale geophysical phenomena (Addison, 2005; Maheswaran and Khosa, 2012; Torrence and Compo, 1998). With the capability to enable multi-scale resolution

and frequency localization in time, wavelets offer the advantage of facilitating a decomposition of the given climate variable time series into its various, but scale specific, dynamic components as proxies of the corresponding physical processes at those scales (Rathinasamy and Khosa, 2012). Numerous studies based on wavelets application to investigate geophysical time series have been reported in the recent past. For example, Venugopal et al. (1999) used wavelet packets to study the energy distribution of rainfall over time & frequency and showed the evidence of dynamic scaling in space-time rainfall. In karstic region, Labat et al. (2001) applied wavelet analysis to investigate the multi-scale phenomena in daily observed rainfall and runoff, and showed that classical correlation and spectral analyses are limited to quantify the multi-scale phenomenon. He further observed that nonlinearity is one of the main problems in Karstic hydrogeology, which can be taken care via wavelet transform. Gaucherel (2002), Lafreniere and Sharp (2003), and Anctil and P. Coulibaly (2004) have used wavelet analysis to interpret temporal patterns of different basin responses that include either rapid processes or slow recharges. Hu and Si (2016) used multiple wavelet coherence techniques on evaporation and other meteorological factors (i.e. relative humidity, mean temperature, sun hours, and wind speed) and showed that the scale-specific and localized relationship exists in geophysical data, which remains hidden in the investigation at observed scale. Even though wavelets are useful in identifying the multi-scale patterns but its application and interpretation involving large spatial data sets will be cumbersome and tedious.

Recently, complex networks have been one of the most active research topics in climatology and other closely related disciplines. Though analyzing networks as nodes connected by links dates back as far as Euler's famous solution of the Königsberg bridge problem (Euler, 1741) but with the improvement of computing technology and with substantial inputs from mathematics and statistics, the theory of complex networks has proven to be a powerful tool for the analysis of climate systems (Zanin, 2014). Complex network theory has received popularity owing to its simplicity, in particular, within the fields of brain (Pfurtscheller and Lopes da Silva, 1999; Betzel and Bassett, 2017; Achard, 2006), cardiovascular research (O'Connor et al., 2013), power grids (Schultz et al., 2016; Nitzbon et al., 2017) non-linear chaotic systems (Arnhold et al., 1999; Mitra et al., 2017b) and climate sciences (Runge et al., 2015; Stolbova et al., 2014; Donges et al., 2009a; Paluš, 2018). Spatial data representation and data visualization which reveal structure and patterns are salient features of a complex network that also contribute to its exponential use. Also, it allows reducing large datasets into simple structures of interactions, which can easily be studied by means of statistical measures such as degree, clustering coefficient, bridgeness, etc. while removing all unnecessary details (Zanin, 2014).

Under the "NatRiskChange" framework, I studied multi-scale interactions and patterns in univariate and multivariate processes by combining the data-driven complex network approach with wavelet transform approach and further attempted to develop an advanced analytical method in climatology. The main challenge, thus, being to understand how state-of-the-art wavelet transform along with integrating the most recent advancements from physics to better understand the multi-scale climate variability.

## 1.3 Objectives

Specifically, the research seeks to develop measures and methods that enable the study of climatic system at different scales such as temporal, spatial, and topological (Fig.1.1) using complex network-based approach. The overarching goals of this study can be subdivided into three objectives as follows:

**Objective I:** To investigate the geophysical process in general and climatic variables in particular, at different scales for their intrinsic but characteristics features.

**Objective II:** To propose new advanced methods by modifying the single scale similarity measures to quantify multi-scale interactions and patterns.

**Objective III:** To study the efficacy of the proposed methods for a range of climatic applications such as (a) capturing multi-scale associations and patterns into univariate sea surface temperature, (b) unravelling multi-scale interactions and spatial diversity in between climatic patterns and regional extreme precipitation.

Here, to show the application of the proposed methods, first, we investigate the patterns and interactions of univariate sea surface temperature (SST) at a different spatial and temporal scale for the entire globe. In the second application, we quantify the multi-scale (temporal, spatial and topological) variability and spatial diversity of Indian rainfall teleconnection. We select India as study region, since, Indian Monsoon is a large-scale climatic phenomenon affecting more than 1.7 billion people. The variability, strength, onset and withdrawal of monsoonal rainfall have an enormous effect on Indian agriculture, economy, life and prosperity of the inhabitants of the Indian sub-continent. Consequently, understanding the mechanisms of the Indian monsoon and its successful forecasting is not only a question of great interest but also a significant societal challenge (Stolbova et al., 2014).

## 1.4 Organization of the thesis

The thesis is organized into six chapters and the methodology used is described in the corresponding chapters.

Chapter 2 investigates the geophysical signal in general and climate signal in particular for their intrinsic but characteristic features. The investigation on a range of prototypical examples confirms that climate-related processes show variability at a range of temporal scales (Fig.1.1) and concludes that the study of these processes at only one time scale might not reveal complete mechanism and interactions. The study further fosters a new method to quantify multi-scale relationships.

Chapter 3 evaluates the ability of state-of-the-art method Pearson's correlation coefficient (PCC) to capture multi-scale (temporal scale, Fig.1.1) interactions and patterns among large spatial data and conclude that PCC alone is not enough to capture the underlying different scale features. We modify PCC by combining it with a wavelet transform to capture multi-scale interactions and patterns. We use complex network to visualize and represent the spatial interactions and patterns over different scales. The applicability of the method is shown on the global sea surface temperature (SST) dataset with physical interpretation.

The application of proposed methods on real climate dataset is presented in chapter

4 and chapter 5. We present two applications on uni-variate and multi-variate climate data to mark the urgent need to consider multi-scale dynamics in climate studies. The two application also make a bridge among different type of scales (temporal, spatial and topological).

In chapter 4, I continue working with the same approach and use global SST dataset with a basic difference that this study focus on both spatial and time scales (Fig.1.1). The study offer a visualization of short and long-range teleconnections of different spatial distance at different temporal scale and physically interpret them.

In chapter 5, we examine the spatial diversity of Indian rainfall teleconnection at different timescales. First we identify homogenous communities following which we compute nonlinear linkages between the identified communities (spatial regions) and dominant climatic patterns, represented by climatic indices such as El Niño/Southern Oscillation, Indian Ocean Dipole, North Atlantic Oscillation, Pacific Decadal Oscillation and Atlantic Multi-decadal Oscillation. This chapter involves topological scale, temporal and spatial scales (Fig.1.1).

In chapter 6, we present a summary and discuss several open issues that surfaced while conducting this research. Appendix A quantifies the topological scale and provides the foundation for chapter 5. In addition, the insights gained from Appendix A are extended to research carried out in Appendix B. Though, Appendix B does not form a direct link with the main structure of the thesis, it discusses in detail the various network measures such as degree and betweeness centrality that have been recurrently used through the text in the thesis.

## 1.5 Author contributions

**Agarwal, A.**, Marwan, N., Maheswaran, R., Merz, B. and Kurths, J.: Multi-scale event synchronization analysis for unraveling climate processes: a wavelet-based approach, Nonlinear Processes in Geophysics, doi.org/10.5194/npg-24-599-2017.

AA developed the theoretical formalism, performed the analytic calculations, and verified the methodology for multi-scale event synchronization. AA collected, preprocessed the dataset used in the study and prepared all the figures. AA took the lead in writing the manuscript. All authors discussed the results, drew conclusions, and provided critical feedback to the manuscript text. All authors were involved in deciding the context and narrative within which the main results have been communicated.

**Agarwal, A.**, Rathinasamy, M., Marwan, N., Caesar, L., and Kurths, J.: A wavelet-based multiscale similarity measure for complex networks. (European physics journal B).

AA & RM jointly conceived the presented idea. AA developed a theoretical framework, performed computations and validated methodology on synthetic case studies. AA downloaded and preprocessed SST dataset, prepared all figures and took lead to write the main text of the manuscript. LC helped to draw a meaningful conclusion from the real-world case study. All authors discussed the results, drew conclusions, and contributed to the manuscript text. RM, NM & JK supervised the project.

**Agarwal, A.**, Caesar, L., Marwan, N., Rathinasamy, M., Merz, B. and Kurths, J.: Detection of short- and long-range teleconnections in SST patterns on different timescales. (Climate dynamics-in draft).

AA substantially contributed to conception, design, acquisition of data, analysis and interpretation of data. AA plotted all the figures and took the lead in writing the manuscript. LC participated in drafting the article and revising it critically for important intellectual content. BM & NM were involved in the extensive discussion and designing the content. RM explicitly contributed in formulating method and provided expert feedback on the work. BM & JK supervised the work and gave final approval of the version to be submitted.

**Agarwal, A.**, Marwan, N., Maheswaran, R., Krishnan,R., Kurths, J., and Merz, B.: Unravelling the spatial diversity of Indian rainfall teleconnections (Scientific reports).

AA devised the project, the main conceptual ideas, and proof outline. AA & RM jointly developed the theoretical framework. AA took lead and implemented the method on real dataset and performs the analysis and tests. AA arranged, preprocessed the dataset, prepared all the figures and wrote the main test. BM closely supervised the work, helped extensively in rewriting the text for journal submission and helped to improve the figures. Intense discussion with NM & JK helped to reach to appropriate conclusion relevant for the climatological interpretation. RK provided the additional help to draw a meaningful conclusion in term of climatological interpretations.

**Agarwal, A.**, Marwan, N., Maheswaran, R., Merz, B. and Kurths, J.: Quantifying the roles of single stations within homogeneous regions using complex network analysis, Journal of Hydrology, doi:10.1016/j.jhydrol.2018.06.050, 2018.

AA designed the study. AA, BM & NM were involved in the planning of theoretical framework. RM helped to arrange the raw Indian precipitation dataset for experiment. AA took lead, processed the precipitation data, performed analysis, drafted the manuscript and designed the figures. AA performed the Z-P space calculations for entire Indian raingauges. RM & BM aided in interpreting the results and worked on the manuscript. All authors discussed the results and commented on the manuscript.

**Agarwal, A.**, Marwan, N., Rathinasamy, M., Ozturk, U., Merz, B. and Kurths, J.: Optimal Design of Hydrometric Station Networks Based on Complex Network Analysis, Hydrol. Earth Syst. Sci. Discuss., 1–21, doi:10.5194/hess-2018-113, 2018.

AA designed and implemented the research model. AA developed the node ranking algorithm and performed several test cases. UO tested the node ranking algorithm on various other networks. NM & BM closely supervised the work and encouraged AA to investigate the node ranking algorithm for hydrological dataset. AA implemented the method and performed the analyses and tests. RM & BM helped to interpret the findings. All authors discussed the results and contributed to the final manuscript.

# Part I.

# Investigation and Quantification

# Chapter 2

# Multi-scale event synchronization analysis for unravelling climate processes: a wavelet-based approach

*Agarwal, A., Marwan, N., Rathinasamy, M., Merz, B. and Kurths, J.: Multi-scale event synchronization analysis for unravelling climate processes: a wavelet-based approach, Nonlinear Process. Geophys., 24(4), 599–611, doi:10.5194/npg-24-599-2017, 2017.*

The temporal dynamics of climate processes are spread across different timescales and, as such, the study of these processes at only one selected timescale might not reveal the complete mechanisms and interactions within and between the (sub-)processes. To capture the non-linear interactions between climatic events, the method of event synchronization has found increasing attention recently. The main drawback with the present estimation of event synchronization is its restriction to analysing the time series at one reference timescale only. The study of event synchronization at multiple scales would be of great interest to comprehend the dynamics of the investigated climate processes. In this paper, the wavelet-based multi-scale event synchronization (MSES) method is proposed by combining the wavelet transform and event synchronization. Wavelets are used extensively to comprehend multi-scale processes and the dynamics of processes across various timescales. The proposed method allows the study of spatio-temporal patterns across different timescales. The method is tested on synthetic and real-world time series in order to check its replicability and applicability. The results indicate that MSES is able to capture relationships that exist between processes at different timescales.

## 2.1 Introduction

Synchronization is a widespread phenomenon that can be observed in numerous climate-related processes, such as synchronized climate changes in the northern and southern polar regions (Rial, 2012), see-saw relationships between monsoon systems (Eroglu et al., 2016), or coherent fluctuations in flood activity across regions (Schmocker-Fackel and Naef, 2010) and among El Niño and the Indian summer monsoon (Maraun, 2005; Mokhov et al., 2011). Synchronous occurrences of climate-related events can be of great societal relevance. The occurrence of strong precipitation or extreme runoff, for instance, at many locations within a short time period may overtax the disaster management capabilities.

Various methods for studying synchronization are available, based on recurrences (Marwan et al., 2007; Donner et al., 2009; Arnhold et al., 1999; Quyen et al., 1999; Quiroga et al., 2000; Quiroga et al., 2002a; Schiff et al., 1996), phase differences (Schiff et al., 1996; Rosenblum et al., 1997), or the quasi-simultaneous appearance of events (Tass et al., 1998; Stolbova et al., 2014; Malik et al., 2012; Rheinwalt et al., 2015). For the latter, the method of event synchronization (ES) has received popularity owing to its simplicity, in particular within the fields of brain (Pfurtscheller and Lopes da Silva, 1999) and cardiovascular research (O'Connor et al., 2013), non-linear chaotic systems (Callahan et al., 1990), and climate sciences (Tass et al., 1998; Stolbova et al., 2014; Malik et al., 2012). ES has also been used to understand driver–response relationships, i.e. which process leads and possibly triggers another based on its asymmetric property. It has been shown that, for event-like data, ES delivers more robust results compared to classical measures such as correlation or coherence functions which are limited by the assumption of linearity (Liang et al., 2016).

Particularly in climate sciences, ES has been successfully applied to capture driver–response relationships, time delays between spatially distributed processes, strength of synchronization, and moisture source and rainfall propagation trajectories, and to determine typical spatio-temporal patterns in monsoon systems (Stolbova et al., 2014; Malik et al., 2012). Furthermore, extensions of the ES approach have been suggested to increase its robustness with respect to boundary effects (Stolbova et al., 2014; Malik et al., 2012) and number of events (Rheinwalt et al., 2016).

Even though ES has been successfully used, it is still limited by measuring the strength of the non-linear relationship at only one given temporal scale, i.e. it does not consider relationships at and between different temporal scales. However, climate-related processes typically show variability at a range of scales. Synchronization and interaction can occur at different temporal scales, as localized features, and can even change with time (Rathinasamy et al., 2014; Herlau et al., 2012; Steinhaeuser et al., 2012). Features at a certain timescale might be hidden while examining the process at a different scale. Also, some of the natural processes are complex due to the presence of scale-emergent phenomena triggered by non-linear dynamical generating processes and long-range spatial and long-memory temporal relationships (Barrat et al., 2008). In addition, single-scale measures, such as correlation and ES, are valid and meaningful only for stationary systems. For non-stationary systems, they may underestimate or overestimate the strength of the relationship (Rathinasamy et al., 2014).

The wavelet transform can potentially convert a nonstationary time series into stationary components (Rathinasamy et al., 2014), and this can help in analysing nonstationary time series using the proposed method.

Therefore, the multi-scale analysis of climatic processes holds the promise of better understanding the system dynamics that may be missed when analysing processes at one timescale only (Perra et al., 2012; Miritello et al., 2013). According to this background, we propose a novel method, multi-scale event synchronization (MSES), which integrates ES and the wavelet approach in order to analyse synchronization between event time series at multiple temporal scales. To test the effectiveness of the proposed methodology, we apply it to several synthetic and real-world test cases.

The paper is organized as follows: Section 2.2 describes the proposed methodology and sec-

tion 2.3 introduces selected case studies. The results are discussed in section 2.4 Conclusions are summarized in Section 2.6.

## 2.2 Methods

Here we describe the methodology for the proposed MSES approach. In this we combine two already well-established approaches (DWT and ES) to analyse synchronization at multiple temporal scales. The following sub-sections briefly introduce wavelets and ES and subsequently provide the mathematical framework for estimating MSES.

### 2.2.1. Discrete wavelet transform

Wavelet analysis has become an important method in spectral analysis due to its multi-resolution and localization capability in both time and frequency domains. A wavelet transform converts a function (or signal) into another form which makes certain features of the signal more amenable to study (Addison, 2005). A wavelet $\psi(t)$ is a localized function which satisfies certain admissibility conditions. The wavelet transform $T_{a,b}(\mathrm{x})$ of a continuous function $x(t)$ can be defined as a simple convolution between $x(t)$ and dilated and translated versions of the mother wavelet $\psi(t)$:

$$T_{a,b}(x) = \int_{-\infty}^{\infty} x(t)\psi_{a,b}(t)\mathrm{d}t \tag{2.1}$$

where $a$ and $b$ refer to the scale and location variables (real numbers) and $\psi_{a,b}$ is defined as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}) \tag{2.2}$$

Depending on the way we sample parameters $a$ and $b$, we get either a continuous wavelet transform (CWT) or a discrete wavelet transform (DWT). A natural way to sample $a$ and $b$ is to use a logarithmic discretization of the scale and link this in turn to the size of steps taken between $b$ locations. This kind of discretization of the wavelet has the form

$$\psi_{\lambda,q}(t) = \frac{1}{\sqrt{a_0^{\lambda}}}\psi(\frac{t-qb_o a_o^{\lambda}}{a_o^{\lambda}}) \tag{2.3}$$

where the integers $\lambda$ and $q$ control the wavelet dilation and translation, respectively; $a_o$ is a specified fixed dilation step parameter and $b_o > 0$ is the location parameter. The general choices of the discrete wavelet parameters $a_o$ and $b_o$ are 2 and 1, respectively. This is known as dyadic grid arrangement.
Using the dyadic grid wavelet, the DWT can be written as

$$T_{\lambda,q} = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{a_o^{\lambda}}}\psi(\frac{t-qb_o a_o^{\lambda}}{a_o^{\lambda}}) \tag{2.4}$$

substituting $a_o = 2$ and $b_o = 1$, we get

$$T_{\lambda,q} = \int_{-\infty}^{\infty} x(t)2^{\frac{-\lambda}{2}}\psi(2^{-\lambda}t - q)dt \tag{2.5}$$

where $T_{\lambda,q}$ are the discrete wavelet transform values given on scale-location grid indexes $\lambda$ and $q$. For the DWT, the values $T_{\lambda,q}$ are known as wavelet coefficients or detail coefficients.

The decomposition of the dyadic discrete wavelet is also associated with the scaling function $\phi_{\lambda,q}(t)$ (Eq.2.6) which represents the smoothing of the signal and has the same form as the wavelet, given by (Addison, 2005).

$$\phi_{\lambda,q}(t) = 2^{\frac{-\lambda}{2}} \phi(2^{-\lambda}t - q) \tag{2.6}$$

The scaling function is orthonormal to the translation of itself, but not to the dilation of itself. $\phi_{\lambda,q}(t)$ can be convolved with the signal to produce approximation coefficients at a given scale as follows:

$$A_{\lambda,q} = \int_{-\infty}^{\infty} x(t)\phi_{\lambda,q}(t)dt \tag{2.7}$$

The approximation coefficients at a specific scale $\lambda$ are known as a discrete approximation of the signal at that scale. As proven in Mallat (1989), the wavelet function and the scaling function form multi-resolution bases resulting in a pyramidal algorithm. The decomposition methodology is schematically shown in Fig.2.1.

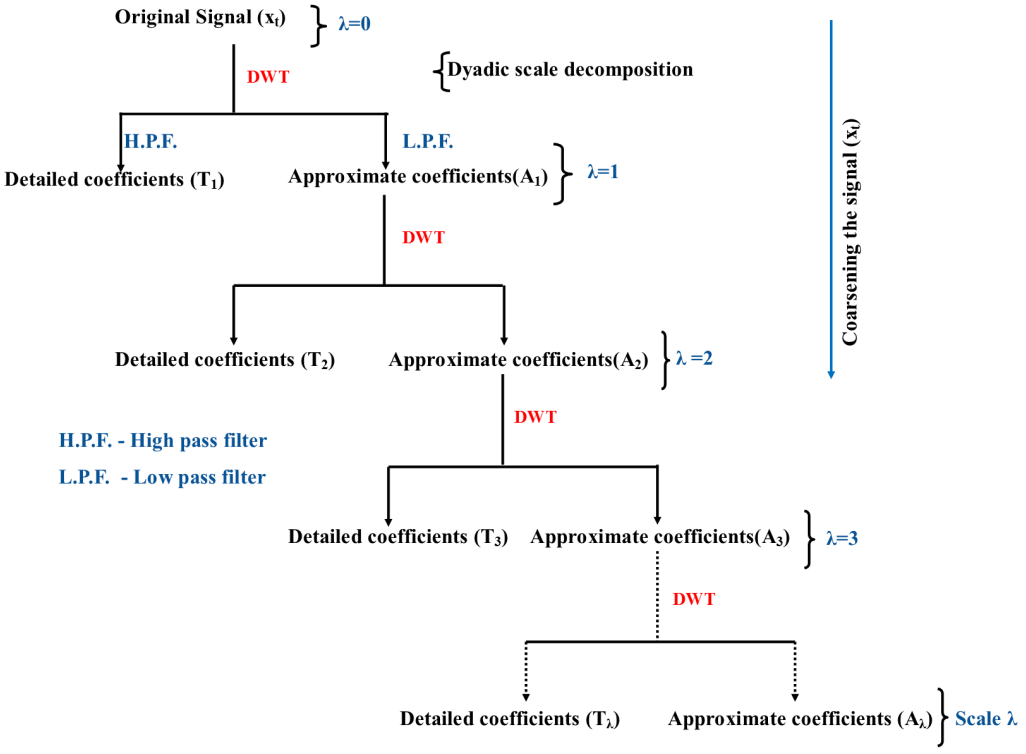



**Figure 2.1.:** Schematic showing the decomposition tree for signal $X_t$ using DWT.

In this study, to calculate the synchronization at multiple scales, we only consider the approximation coefficients (not detail coefficients) at that particular scale because the aim is to separate the effects of time-localized features and high frequency components from the signal.

For different $\lambda = 1, 2, 3, ...$, the approximation coefficients $A_\lambda$ correspond to the "coarse-grained" original signal after removal of the details at scales $\lambda, \lambda - 1, ..., 1$. In practical terms, considering a daily climatic time series at $\lambda = 0$, the time series represents the original observations. At $\lambda = 1$, $A_1$ represents the features beyond the 2-day scale (wavelet scale) which is obtained by extracting $T_1$ (2-day features) from the original time series. Similarly,

at $\lambda = 3$, $A_3$ represents the climatic variable beyond the 8-day scale and is obtained after removing $T_1$, $T_2$, and $T_3$ (2-, 4-, and 8-day features) from the original signal. In essence, $A_1$, $A_2$, $A_3$,...represent the original signal at different timescales. The schematic plot explaining the procedure and relationship between signal, approximate component, and detailed component has been shown in Fig.2.2.

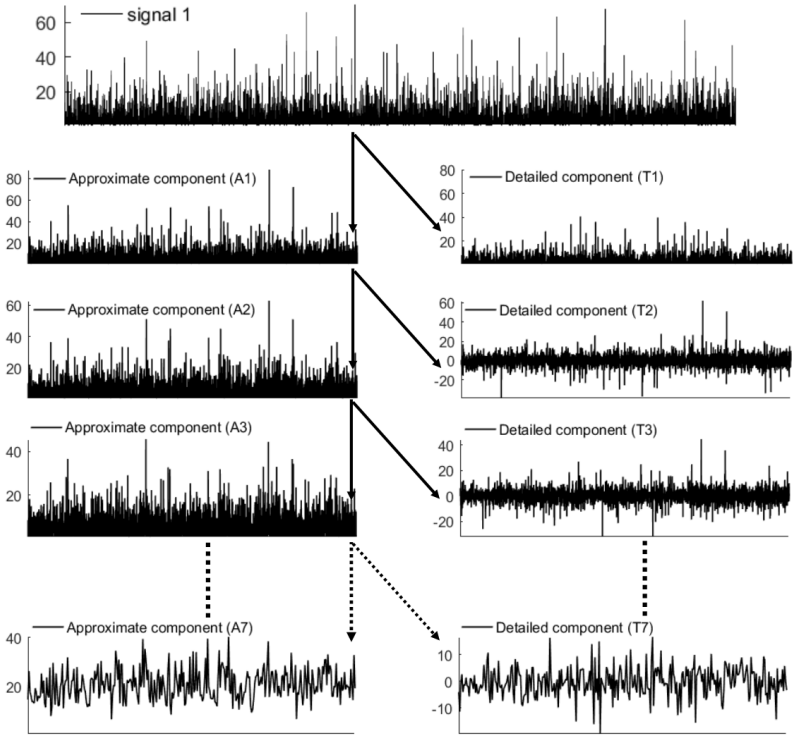



**Figure 2.2.:** Scheme of multi-scale decomposition of signals using discrete wavelet transformation (DWT). The relationship between signal, approximate component, and detailed component is shown.

For simplicity we denote the approximation coefficient $A_{\lambda,q}$ of the signal x(t) at scale $\lambda$ as $x_\lambda$.

### 2.2.2. Event synchronization

To quantify the synchronous occurrence of events in different time series, we use the event synchronization (ES) method proposed by Quiroga et al. (2002a). ES can be used for any time series in which we can define events, such as single neuron recordings, eptiform spikes in EEGs, heart beats, stock market crashes, or abrupt weather events, such as heavy rainfall events. However, ES is not limited to this definition of events. It could also be applied to time series which are pure event time series (e.g. heart beats). In principle, when dealing with signals of different characters, the events could be defined differently in each time series, since their common cause might manifest itself differently in each signal (Quiroga et al., 2002a). ES has advantages over other time-delayed correlation techniques (e.g. Pearson lag correlation), as it allows us to study interrelations between series of non-Gaussian data or data with heavy tails, or to use a dynamical (non-constant) time delay (Tass et al., 1998; Stolbova et al., 2014). The latter refers to a time delay that is dynamically adjusted according to the two time series being compared, which allows for better adaptation to the region of interest. Furthermore, ES has been specifically designed to calculate non-linear

linkages between time series. Various modifications of ES have been proposed, such as solving the problems of boundary effects and bias due to an infinite number of events (Tass et al., 1998; Stolbova et al., 2014; Malik et al., 2012).

The modified algorithm proposed by Stolbova et al. (2014), Malik et al. (2012), and Rheinwalt et al. (2016) works as follows: an event occurs in signals x(t) and y(t) at time $t_l^x$ and $t_m^y$, where $l = 1,2,3,4...,S_x$, $m = 1,2,3,4...,S_y$, and $S_x$ and $S_y$ are the total number of events, respectively. In our study, we derive events from a more-or-less continuous time series by selecting all time steps with values above a threshold ($\alpha = 95th$ percentile). These events in $x(t)$ and $y(t)$ are considered synchronized when they occur within a time lag $\pm\tau_{lm}^{xy}$ which is defined as follows:

$$\tau_{lm}^{xy} = \min\{t_{l+1}^x - t_l^x, t_l^x - t_{l-1}^x, t_{m+1}^y - t_m^y, t_m^y - t_{m-1}^y\} \tag{2.8}$$

This definition of the time lag helps to separate independent events, as it is the minimum time between two succeeding events. Then we count the number of times $C(x|y)$ an event occurs in $x(t)$ after it appears in $y(t)$ and vice versa ($C(y|x)$).

$$C(x|y) = \sum_{l=1}^{S_x} \sum_{m=1}^{S_y} J_{xy} \tag{2.9}$$

and

$$J_{xy} = \begin{cases} 1, & \text{if } 0 < t_x^l - t_m^y < \tau_{lm}^{xy} \\ \frac{1}{2}, & \text{if } t_x^l = t_m^y \\ 0, & \text{otherwise.} \end{cases} \tag{2.10}$$

$C(y|x)$ is calculated analogously but with exchanged $x$ and $y$. From these quantities we obtain the symmetric measure:

$$Q_{xy} = \frac{C(x|y) + C(y|x)}{\sqrt{(S_x - 2)(S_y - 2)}} \tag{2.11}$$

$Q_{xy}$ is a measure of the strength of event synchronization between signals $x(t)$ and $y(t)$. It is normalized to $0 < Q_{xy} < 1$, with $Q_{xy}=1$ for perfect synchronization (coincidence of extreme events) between signals $x(t)$ and $y(t)$.

Recalling Eq.(2.7), the scale-wise approximation at different scales 0,1,2,...,$\lambda$ for any given time series $x(t)$ is given by $x_\lambda = A_{\lambda,q}$ where $x_\lambda$ represents the approximation coefficients of signal $x(t)$ at scale $\lambda$. Now, to determine the synchronization between any two time series $x(t)$ and $y(t)$ at multiple scales, the event synchronization is estimated between the scaled versions of $x(t)$ and $y(t)$ for different $\lambda$ resulting in multi-scale event synchronization (MSES). The normalized strength of MSES between signals $x(t)$ and $y(t)$ at scale $\lambda$ is then defined as

$$Q^{x_\lambda, y_\lambda} = \frac{C(x_\lambda|y_\lambda) + C(y_\lambda|x_\lambda)}{\sqrt{(S_{x_\lambda} - 2)(S_{y_\lambda} - 2)}} \tag{2.12}$$

$Q^{x_\lambda, y_\lambda} = 1$ for perfect synchronization, and $Q^{x_\lambda, y_\lambda} = 0$ suggests the absence of any synchronization at scale $\lambda$ between $x(t)$ and $y(t)$.
Fig.2.3 shows the stepwise methodology of multi-scale event synchronization.

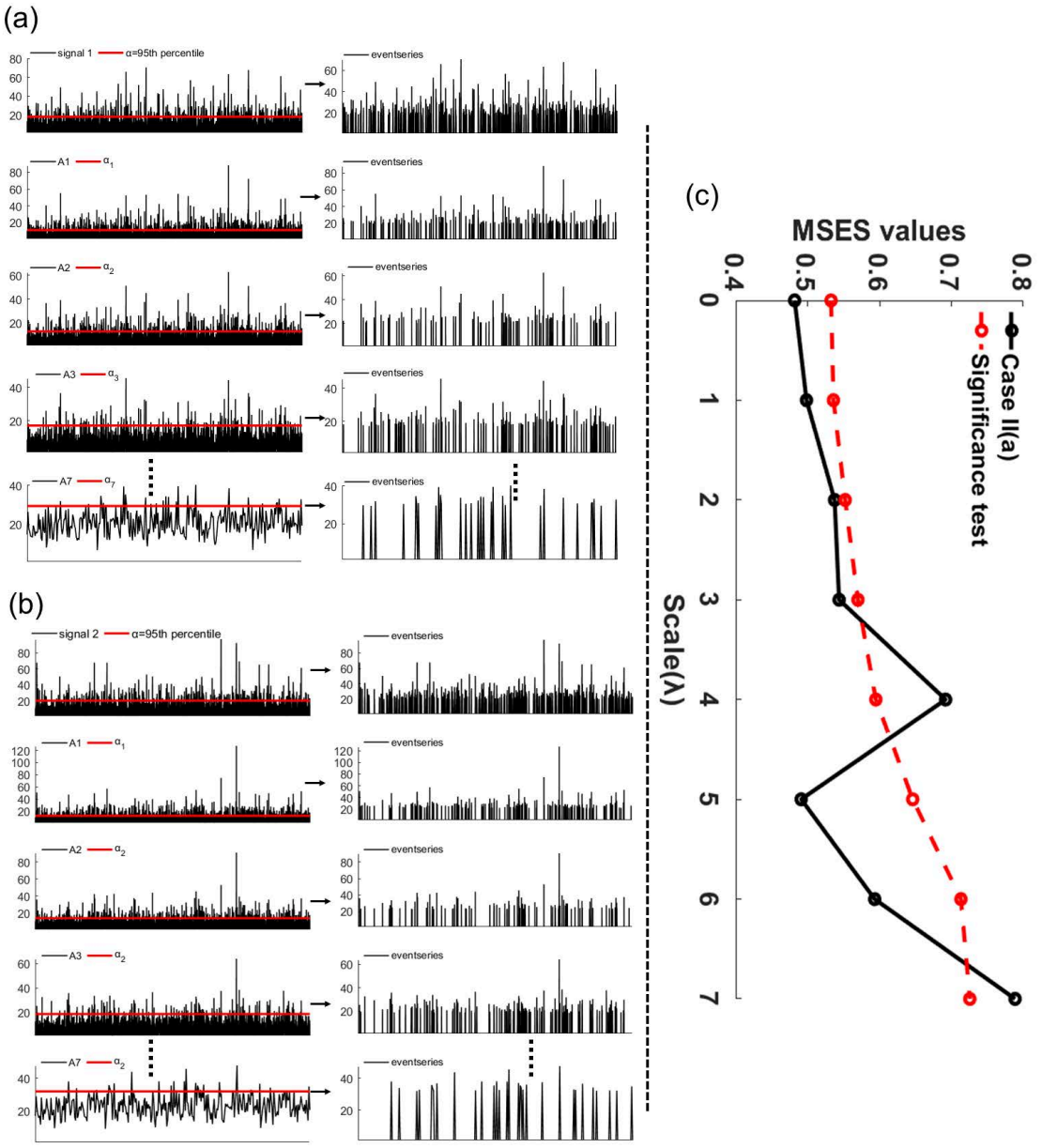



**Figure 2.3.:** Multi-scale event synchronization (MSES) stepwise methodology. (a) Signal 1 and its decomposed component along with corresponding event series after applying the (95th percentile) threshold. (b) Same for signal 2. (c) Event synchronization values corresponding to each scale.

### 2.2.3. Significance test for MSES

To evaluate the statistical significance of ES values, a surrogate test is used (Rheinwalt et al., 2016). We randomly reshuffle each time series 100 times (an arbitrary number). Reshuffling is done without replacement because estimating the expected number of simultaneous events in independent time series is equivalent to the combinatorial problem of sampling without replacement (Rheinwalt et al., 2016). Then, for each pair of time series, we calculate the MSES values for the different scales. At each scale, the empirical test distribution of the 100 MSES values for the reshuffled time series is compared to the MSES values of the

original time series. Using a 1% significance level, we assume that synchronization cannot be explained by chance, if the MSES value at a certain scale of the original time series is larger than the $99th$ percentile of the test distribution.

## 2.3 Data and study design to test MSES

The proposed method is tested using synthetic and real-world data. The aim of these tests is to understand whether MSES is advantageous, compared to ES, in understanding the system interaction and the scale-emerging natural processes.

### 2.3.1. Testing MSES with synthetic data

Following the approach of Rathinasamy and Khosa (2012b), Yan and Gao (2007), and Hu and Si (2016), we test MSES using a set of case studies including stationary and non-stationary synthetic data. The details of the case studies and the wavelet power spectra are given in Table 2.1 and Fig.2.4, respectively.

**Table 2.1.:** Details of synthetic test cases.

| Case | Mathematical expression | Other details | References & figures |
|---|---|---|---|
| I(a) | Sinusoidal stationary signal<br>$S1 = S +$ srong $noise_1$ | $S = \sin(\frac{2\pi t}{50}) + \cos(\frac{2\pi t}{60})$<br>$\frac{Noise_1}{signal} \approx 2.8$ | Rathinasamy et al. (2014) |
| I(b) | Sinusoidal stationary signal<br>$S2 = S +$ weak $noise_1$ | $S = \sin(\frac{2\pi t}{50}) + \cos(\frac{2\pi t}{60})$<br>$\frac{Noise_1}{signal} \approx 2.8$ | Rathinasamy et al. (2014) |
| II(a) | Stationary dataset (S1 and S2)<br>$S1 = S_{t_1} + y1 + y2 + y4$<br>$S2 = S_{t_2} + y1 + y3 + y4$ | Two AR1 processes $S_t = \phi S_{t-1} + \varepsilon_t$<br>$\varepsilon_t$=uncorrelated random noise<br>Parameter $\phi_1 = 0.60; \phi_2 = 0.70$ | Yan and Gao (2007)<br>Hu and Si (2016)<br>Fig.2.3:Panel I |
| II(b) | Stationary dataset (S1 and S2)<br>$S1 = S_{t_1} + y2$<br>$S2 = S_{t_2} + y4$ | $y1 = \sin(\frac{2\pi t}{16})$; $y2 = \sin(\frac{2\pi t}{32})$<br>$y3 = \sin(\frac{2\pi t}{64})$; $y2 = \sin(\frac{2\pi t}{128})$<br>where $t = 1,2,3,...401777$ | Yan and Gao (2007)<br>Hu and Si (2016)<br>Fig.2.3:Panel II |
| III(a) | Non-stationary dataset<br>$S1 = z1 + z2 + z3 + z4 + z5$<br>$S2 = S1$+random noise (uncorrelated)<br>$\frac{noise}{signal} \approx 2.781$<br>where $t = 1,2,3,...40177$ | $z1 = \cos(500\pi(\frac{t}{1000})\cdot 5)$<br>$z2 = \cos(250\pi(\frac{t}{1000})\cdot 5)$<br>$z3 = \cos(250\pi(\frac{t}{1000})\cdot 5)$<br>$z4 = \cos(125\pi(\frac{t}{1000})\cdot 5)$<br>$z4 = \cos(62.5\pi(\frac{t}{1000})\cdot 5)$<br>$z5 = \cos(31.25\pi(\frac{t}{1000})\cdot 5)$ | Yan and Gao (2007)<br>Hu and Si (2016)<br>Fig.2.3: Panel III |
| III(b) | Non-stationary dataset<br>$S1 = z4 + z5$<br>$S2 = S1$+random noise (uncorrelated)<br>$\frac{noise}{signal} \approx 2.781$<br>where $t = 1,2,3,...40177$ | $z4 = \cos(62.5\pi(\frac{t}{1000})\cdot 5)$<br>$z5 = \cos(31.25\pi(\frac{t}{1000})\cdot 5)$ | Yan and Gao (2007)<br>Hu and Si (2016)<br>Fig.2.3: Panel IV |

**Case I.** A single synthetic stationary time series ($S$) is generated and contaminated with two random white noise time series. Two sub-cases with different noise–signal ratios

are investigated (Table 2.1). This case allows understanding of how the synchronization between two series is affected by the presence of noise or high frequency features. For climate variables such situations can emerge when two signals originate from the same parent source or mechanism (e.g. identical large-scale climatic mode, identical storm tracks) but get covered by high-frequency fluctuations arising from local features.

**Case II.** Here we generate two stationary signals consisting of partly shared long-term oscillations and autoregressive ($AR1$) noise $S_t$ (see Table 2.1). The long-term oscillations $y1, y2, y3$, and $y4$ have periods of $16, 32, 64$, and $128$ units, respectively (Fig. 2.4, Panel I). The purpose of case II(a) is to test the ability of MSES to identify synchronization in processes which originate from different parent sources or different mechanisms (e.g. two different climatic process, different storm tracks) but have some common features ($y1$ and $y4$) at coarser scales.

**Case II(b)** presents two signals (Fig.2.4, Panel II) with no common features across all scales. Feature $y2$ in signal $S1$ and feature $y4$ in signal $S2$ represent a long-term oscillation of period 32 and 128 units, respectively. The idea is to investigate the possibility of overprediction of synchronization if we analyse at one scale only.

**Case III** Here, MSES is tested using non-stationary signals (Fig.2.4, Panel III and IV) generated as proposed by Yan and Gao (2007) and Hu and Si (2016). The signal encompasses five cosine waves ($z1$ to $z5$), whereas the square root of the location term results in a gradual change in frequency. Two combinations are generated of which case III(a) investigates the ability of MSES to deal with non-stationarity signals. Case III(b) examines the capability of MSES to capture processes emerging at lower scales (in this case at scales 5 and 6) in the presence of short-lived transient features. For both combinations, the signal is contaminated with white noise.

The time series of case III have features that are often found in climatic and geophysical data, where highfrequency, small-scale processes are superimposed on lowfrequency, coarse-scale processes (Hu and Si, 2016). Such structures are widespread in time series of seismic signals, turbulence, air temperature, precipitation, hydrologic fluxes, or the El Niño/ Southern Oscillation (ENSO). They can also be found in spatial data, e.g. in ocean waves, seafloor bathymetry, or land surface topography (Hu and Si, 2016).

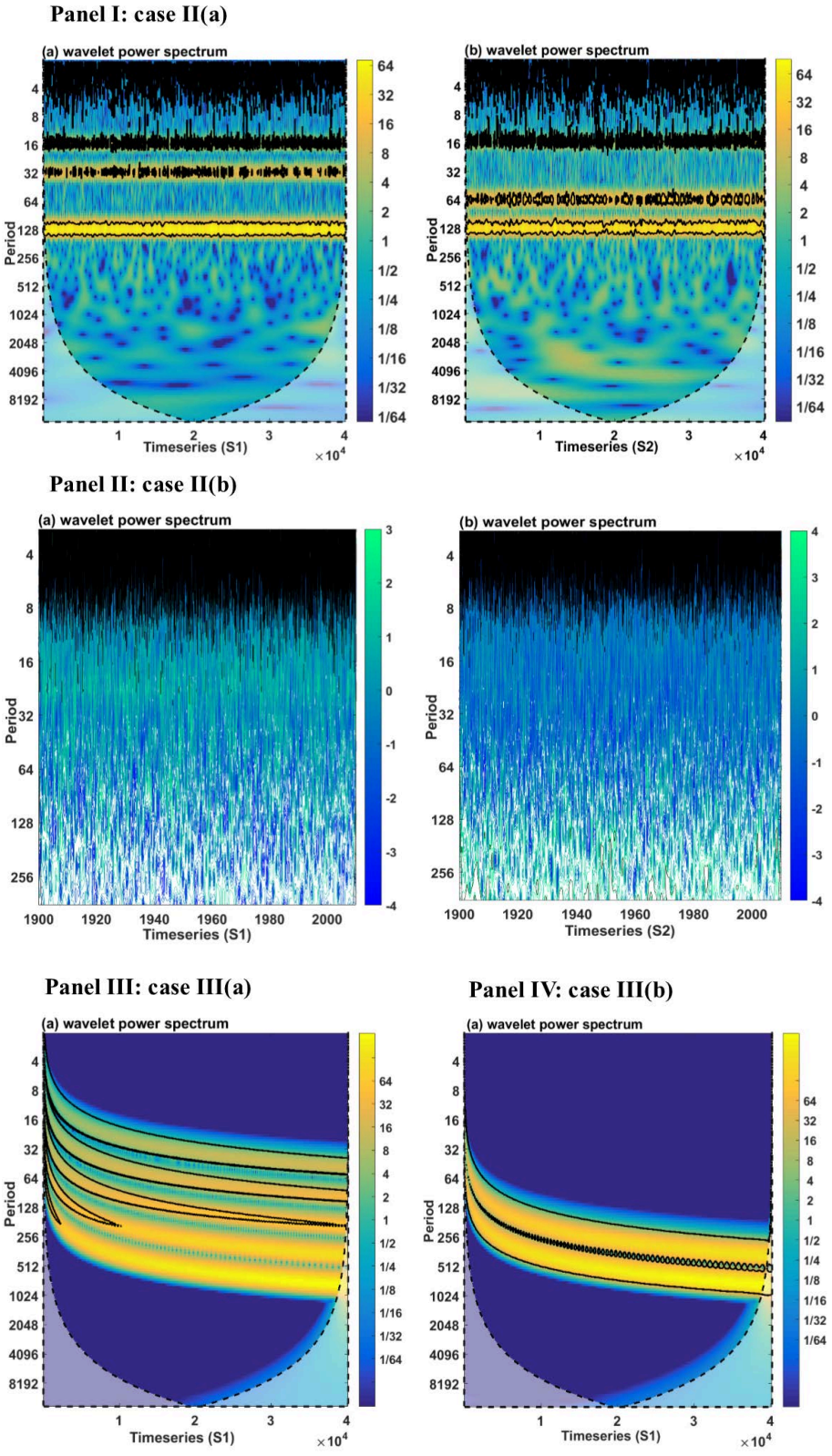



**Figure 2.4.:** Wavelet power spectra (WPS) of the test signals (Table2.1). Panel I: original signal $S1$ (left) and $S2$ (right), respectively, for case II(a); Panel II: original signal $S1$ (left) and $S2$ (right), respectively, for case II(b); Panel III: original signal $S1$ for case III(a); Panel IV: original signal $S1$ for case III(b). In all the panels, the y-axis represents the corresponding Fourier period=$2^{\lambda}$.

### 2.3.2. Testing MSES with real-world data

To test MSES with real-world data, we use precipitation data from stations in Germany (Fig.2.5): 110 years of daily data, from 1 January 1901 to 31 December 2010, are available from various stations operated by the German Weather Service. Data processing and quality control were performed according to Oesterle (2001).

**Case IV.** We use daily rainfall data from the three stations: Kahl/Main, Freigericht-Somborn, and Hechingen(station ID: 20009, 20208, and 25005). Considering Kahl/Main (station 1) as the reference station, the distance to the other two stations, Freigericht-Somborn(station 2) and Hechingen (station 3), are 14.88 and 185.62 km, respectively (Fig.2.5). Rainfall is a point process with large spatial and temporal discontinuities ranging from very weak to strong events within small temporal and spatial scales (Malik et al., 2012). This case explores the ability of MSES, in comparison to ES, to improve the understanding of synchronization given such time series features.

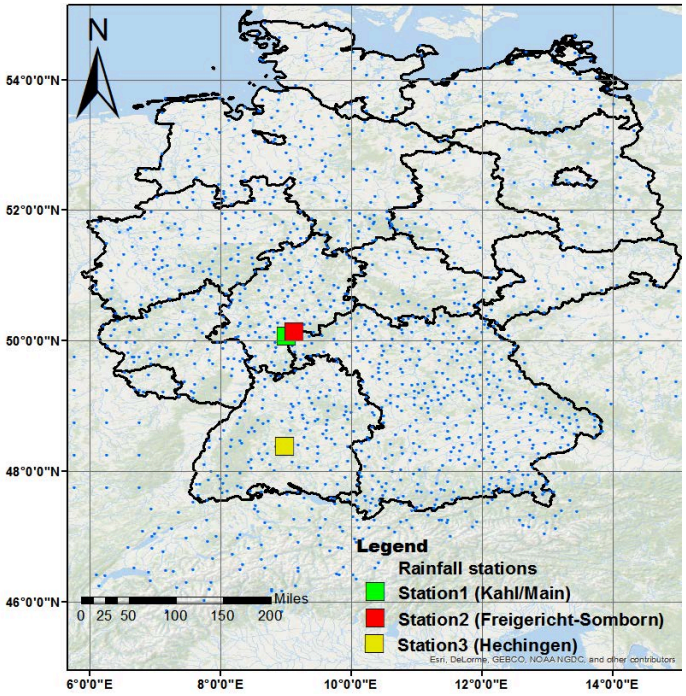



**Figure 2.5.:** Geographical locations of rainfall stations considered in case study IV.

## 2.4 Results

To evaluate the synchronization between two signals, which can be expressed in terms of events, at multiple scales, we decompose the given time series up to a maximum scale beyond which there is no significant number event. The number of events at a scale is a function of the nature of the timeseries and also the length of the time series under consideration. In most cases it was found that the number of events was significantly reduced after seven or eight levels of decomposition. We use the Haar wavelet as this is one of the simplest but most basic mother wavelets. There are several other mother wavelets which could be used for wavelet decomposition; however, it has been demonstrated that the choice of the mother wavelets does not affect the results to a great extent for rainfall (Rathinasamy et al., 2014).

In case I(a) the noise–signal ratio is quite high in the range of 2.7–3 (Table 2.1), such that the effect of the noise is felt up to scale 7 (Fig.2.6). Although both signals stem from the same parent source and hence ideally they should possess perfect synchronization ($ES \sim 1$) at all scales, the ES value at the observational scale ($\lambda = 0$) is moderate ($\sim 0.7$), leading to the interpretation that both signals are only weakly synchronized. In contrast, the proposed MSES approach is able to capture the underlying features (which were hidden in the original signal) at higher scales ($\lambda \geq 1$) by approaching ES values of 1, indicating the actual synchronization between these signals. At the scale ($\lambda = 0$) the ES measure is lower because of the heavy noise covering the underlying information. Considering higher scales, the effect of noise is removed through wavelet decomposition, allowing for a more reliable identification of the actual underlying synchronization between the signals. Interestingly, the slight decrease in the ES values at a high scale ($\lambda \geq 7$) (Fig.2.6) might indicate that the essential feature that is responsible for the synchronization at that scale gets removed in the form of a detail component (Fig.2.2). If features are present at a particular scale $\lambda$ and when we go up to the next scale ($\lambda + 1$), those features get removed in the form of the details and essentially the synchronization is lost at the scale $\lambda + 1$.

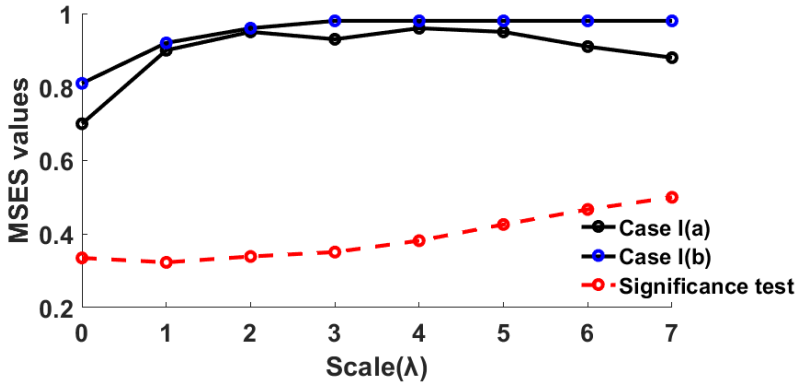



**Figure 2.6.:** MSES values for case I(a) and case I(b), including significance test values for the significance level of 1 %. The value at scale 0 is equal to the single-scale ES analysis.

While repeating the same analysis but with a lower noise–signal ratio (i.e. case I(b)), we find that the effect of noise is almost completely removed after ($\lambda > 3$) and the MSES values remain unaltered because of the same signal structure (Fig.2.6). These findings confirm that the MSES approach is able to capture the synchronization in the presence of noise.

The significance test (Sect.2.4) underlines the high level of synchronization as indicated by the quite high ES values (Fig.2.6). Based on this example we find that the MSES analysis captures the synchronization at multiple scales.

Case II(a) presents a system where synchronization between two signals exists at a common long-term frequency ($y_1$ and $y_4$). This is particularly relevant in studying the rainfall processes of two different regions, which are governed by different local climatic processes but similar long-term oscillations such as ENSO cycles. The MSES values ($\lambda =$ 0 to 7) are smaller than the confidence level, except for scales 4 and 7 (Fig.2.7 a). The synchronization emerging at scale 4 ($\lambda = 4$)and scale 7 ($\lambda = 7$) corresponds to features present at those scales shown in the wavelet power spectrum (Fig.2.5, Panel I). The thick contour in the WPS indicates the presence of significant features (at the 5% significance level) corresponding to $y_1$, $y_2$, $y_3$, and $y_4$ (Table 2.1). In the same figure, the dashed curve represents the cone of

influence (COI) of the wavelet analysis. Outside of this region edge effects become more influential. Any peak falling outside the COI has presumably been reduced in magnitude due to zero padding necessary to deal with the finite length of the time series. To test the statistical significance of WPS, a background Fourier spectrum is chosen (Addison, 2005; Agarwal et al., 2016a, b).

For case II(b), we would expect that the ES value should be zero or nonsignificant at scale $\lambda = 0$. However, we find that the synchronization between S1 and S2 at scale $\lambda = 0$ is significant (Fig. 2.7 b), although there is no common feature by construction (Fig.2.3, Panel II).
Interestingly, the MSES does not find significant synchronization at any scale ($\lambda > 0$). Moreover, the MSES values become zero after scale 4 because signals S1 and S2 have no common feature beyond these scales.
As seen clearly, the ES at only one scale overpredicts the actual synchronicity between the two series. This behaviour may be due to the integrated effect of all scales, and hence some spurious synchronization (although rather small but still significant) is indicated.

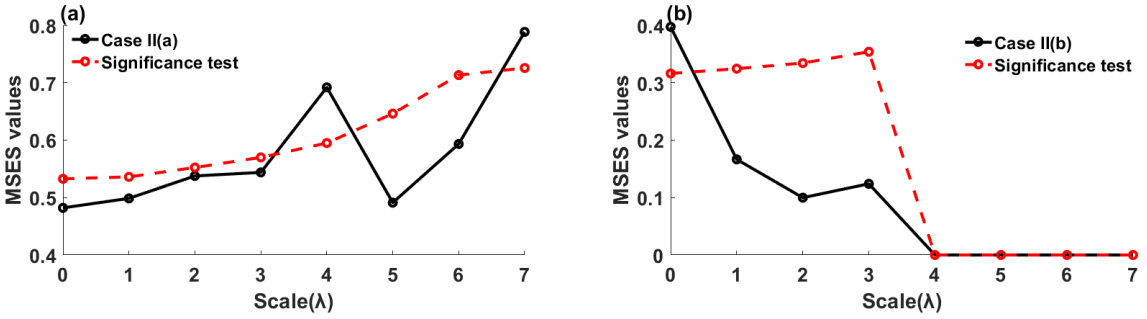



**Figure 2.7.:** (a, b) MSES and significance level (1%) values at different scales for cases II(a) and II(b). The value at scale 0 is equal to the single-scale ES analysis.

Case III(a) is used as an analogue of dynamics and features of natural processes (Table2.1). Its WPS (Fig.2.4: Panel III) shows non-stationary, time-dependent features at higher scales $2 \geq \lambda \geq 6$. ES values at lower scales $\lambda \leq 1$ are below the significance level, revealing that the two signals are not synchronized (Fig.2.8a). The ES for the signal components of the larger timescales reveals significant synchronization up to scale 6, which is expected because of the common features (scale 2 to scale 6) in S1 and S2. After scale $\lambda = 6$, the MSES value drops below the significance level as the features responsible for synchronization are removed in the form of the details component during decomposition. Results from this case show the wavelet's ability in capturing the underlying multiple non-stationarities that are common in both the time series which otherwise go unnoticed using ES at the observation scale.

The similar case III(b) is used to investigate the behaviour of MSES in a scale-emerging process in a non-stationary regime (Table2.1). As the wavelet spectrum of the signal reveals, only features at scales 5 and 6 are present (Fig.2.4: Panel IV). The corresponding MSES values are significant only at those scales (Fig.2.8b), revealing the synchronization at scales 5 and 6. This case illustrates that MSES reveals only the relevant timescales and does not mix them with the observation scale. In reality, there may be situations where the causative events act only at certain timescales and remain unconnected at other timescales. Under such situations MSES is useful for unravelling the relevant scale-emerging relationships.

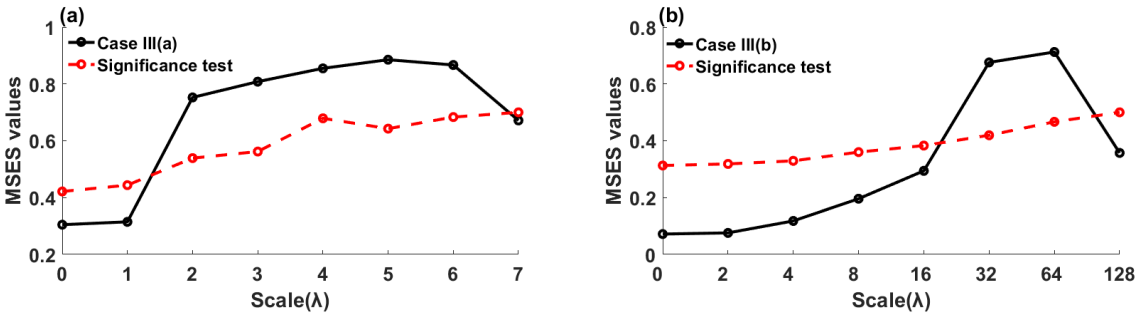



**Figure 2.8.:** (a, b) MSES values and significance level (1%) at different scales for cases III(a) and III(b). The value at scale 0 is equal to the single-scale ES analysis.

After testing the efficacy of the proposed MSES approach by using some prototypical situations, we apply the approach to real observed rainfall data (case IV). We find significant ES values between station 1 and station 2 at the scales $\lambda$=1,5,and 7 (Fig.2.9a) by tracking the features present in the WPS (Fig.2.9c, d, and e). The significant ES value at the observational scale ($\lambda = 0$) might be due to the integrated effect of features present at coarser scales ($\lambda = 1$, 5, and 7). In order to emphasize the features present in the data, we use the global wavelet spectrum (Fig.2.9f, g, and h) which is defined as the time average of the WPS (Agarwal et al., 2016a; Mallat, 1989).

Applying ES in the traditional way, i.e. analysing only at scale 0, we find synchronization. However, only when we consider multiple scales are we able to find that the synchronization is the result of high- and low-frequency components present at scales 1, 5, and 7.

For station 1 and station 3 synchronization is significant at scale 7 ($\lambda = 7$) (Fig.2.9b). However, evaluating the ES in the traditional way (i.e. $\lambda = 0$) leads to the conclusion that both stations are not significantly synchronized. Here, MSES plays a critical role in identifying synchronization at specific temporal scales. Hence, MSES provides further insights into the process, such as low-frequency features that are present and the dominating scales causing the significant synchronization at scale 0.

The results for the real-world case study suggest that proximity of stations (station 1 and station 2) does not necessarily indicate synchronization at all scales. For stations 1 and 3, which are comparatively far from each other, we find insignificant synchronization at the observational scale. However, considering the scales separately, MSES detects significant synchronization at scale 7 as both stations might be sharing some common climatic cycle at this scale.

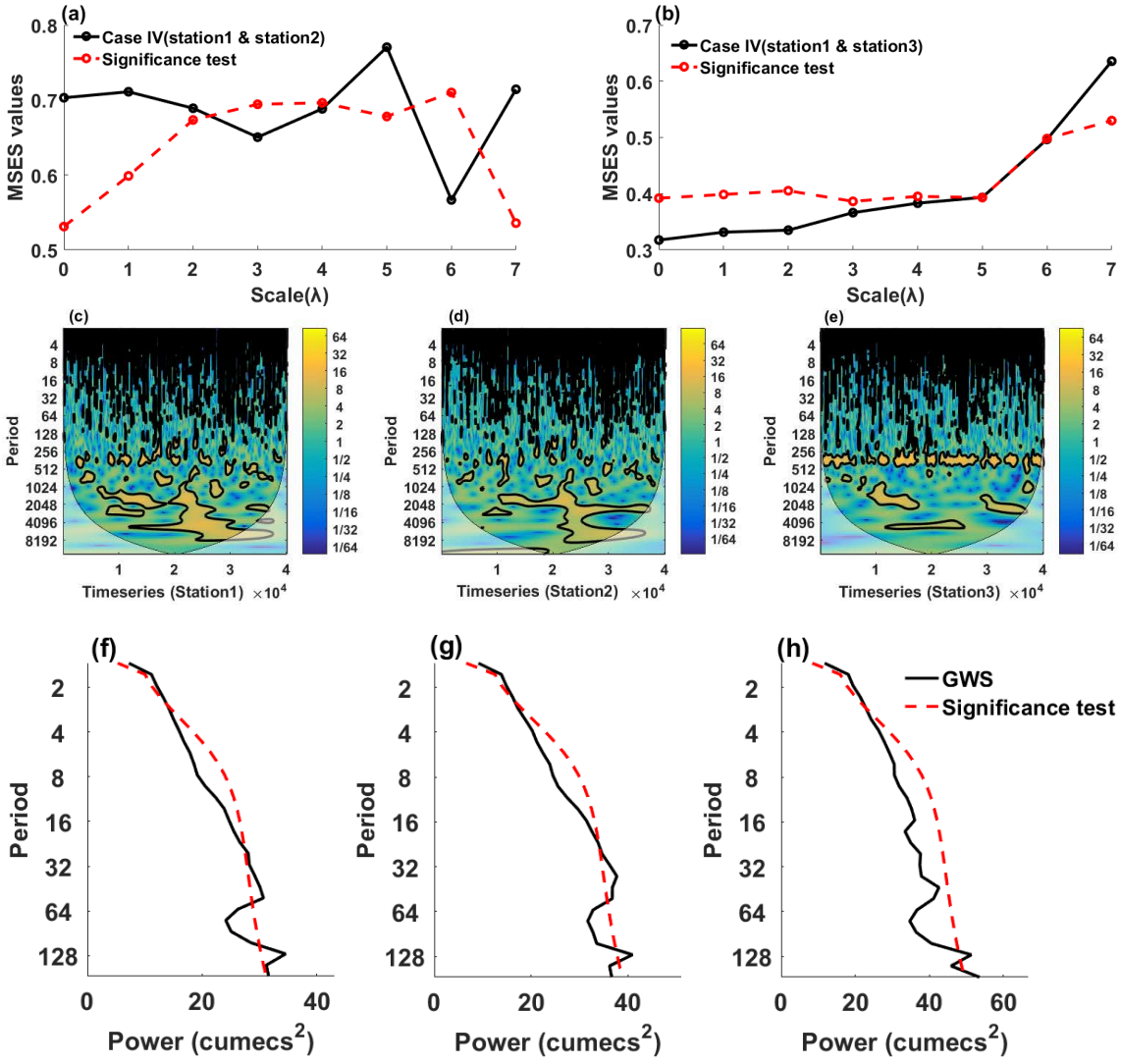



**Figure 2.9.:** (a, b) MSES and significance level (1%) values at various scales for stations 1 and 2 and stations 1 and 3, respectively; (c, d, e) WPS of precipitation of stations 1(c), 2(d), and 3(e) (station ID: 20009, 20208, 25005), respectively; (f,g,h) global wavelet spectrum of the same stations. In (c)–(h) the y-axis represents the corresponding Fourier period=$2^{\lambda}$.

## 2.5 Discussion

We have compared our novel MSES method with the traditional ES approach by systematically applying both methods to a range of prototypical situations. For test cases I and II we find that the ES value at the observation scale is influenced by noise, thereby reducing the ES values of two actually synchronized time series. When using MSES, the synchronization between the two time series can be much better detected even in the presence of strong noise. Another important aspect related to the analysis of these cases is that MSES has the ability to unravel synchronization between two stationary systems at timescales which are not obvious at the observation scale (scale-emerging processes). From these observations, it becomes clear that (i) event synchronization only at a single scale of reference is less robust, and (ii) the dependency measure of two given processes based on ES changes with the timescale depending on the features present in these processes.

Case study III illustrates that for a non-stationary system with synchronization changing over temporal scales, the single-scale ES is not robust. In contrast, MSES uncovers the underlying synchronization clearly. MSES is able to track the scale-emerging processes, scale of dominance in the process, and features present.

The real-world case study IV shows that the synchronization between climate time series can differ with temporal scales. The strength of synchronization as a function of temporal scale might result from different dynamics of the underlying processes. MSES has the ability to uncover the scale of dominance in the natural process.

Our series of test cases confirms the importance of applying a multi-scale view in order to investigate the relationship between processes that exist at different timescales. We suggest that investigating synchronization just at a single, i.e. observational, scale could give limited insight. The proposed extension offers the possibility of deciphering synchronization at different timescales, which is important in the case of climate systems where feedbacks and synchronization occur only at certain timescales and are absent at other scales.

## 2.6 Conclusions

We have proposed a novel method which combines wavelet transforms with event synchronization, thereby allowing us to investigate the synchronization between event time series at a range of temporal scales. Using a range of prototypical situations and a real-world case study, we have shown that the proposed methodology is superior compared to the traditional event synchronization method. MSES is able to provide more insight into the interaction between the analysed time series. Also, the effect of noise and local disturbance can be reduced to a greater extent and the underlying interrelationship becomes more prominent. This is attributed to the fact that wavelet decomposition provides a multi-resolution representation which helps to improve the estimation of synchronization. Another advantage of the proposed approach is its ability to deal with non-stationarity. Wavelets being made on local bases can pick up the non-stationary, transient features of a system, thereby improving the estimation of ES. Finally, it can be concluded that the proposed method is more robust and reliable than the traditional event synchronization in estimating the relationship between two processes.

## 2.7 Acknowledgments

This research was funded by the Deutsche Forschungsgemeinschaft (DFG) (GRK 2043/1) within graduate research training group Natural risk in a changing world (NatRiskChange) at the University of Potsdam (http://www.uni-potsdam.de/natriskchange) and RSF support (support by the Russian Science Foundation (grant no. 16-12-10198)). The third author acknowledges the research funding from the Inspire Faculty Award, Department of Science and Technology, India, for carrying out the research. Also, we gratefully acknowledge the provision of precipitation data by the German Weather Service.

## 2.8 Data availability

The authors used Germany's precipitation data which is maintained and provided by GermanWeather Service. The data is publicly accessible at https://opendata.dwd.de/. Further, preprocessing of the data was done by Potsdam Institute for Climate Impact Research (Conradt et al., 2012; Oesterle, 2001).

# Chapter 3

# Wavelet-based multi-scale similarity measure for complex networks

*Agarwal, A., Rathinasamy, M., Marwan, N., Caesar, L., and Kurths, J.: A wavelet-based multi-scale similarity measure for complex networks, European Physics Journal-B, https://doi.org/10.1140/epjb/e2018-90460-6.*

The application of complex networks in climate studies have enabled a compelling perspective and led to novel insights. However, the analysis and representation of a multi-scale complex relationship that exists in the global climate system are limited. A logical first step in addressing this issue is to construct multiple networks over different timescales. Therefore, we propose to apply the Wavelet Multi-scale Correlation (WMC) similarity measure which is a combination of two state-of-the-art methods, viz. wavelet and Pearson's correlation, for investigating multi-scale processes through complex networks. Firstly we decompose the data over different timescales using the wavelet approach and subsequently construct a corresponding network by Pearson's correlation.
The proposed approach is illustrated and tested on two synthetics and one real-world example. The first synthetic case study shows the efficacy of the proposed approach to unravel scale-specific connections which are often undiscovered at a single scale. The second synthetic case study illustrates that by dividing and constructing a separate network for each time window we can detect significant changes in the signal structure. The real-world example investigates the behavior of the global sea surface temperature (SST) network at different timescales. Intriguingly, we notice that spatial dependent structure in SST evolves temporally. Overall, the proposed measure has an immense potential to provide essential insights on understanding and extending complex multivariate process studies at multiple scales.

## 3.1 Introduction

Networks provide analytical capabilities to uncover structure and patterns of geophysical processes in which nodes denote geographic locations, and links indicate connections between these locations not by spatial proximity but rather on the basis of a similar dynamics in the process. The strength of the connection (link) is estimated using similarity measures that quantify the interrelation between the two nodes. In the past, several similarity measures have been investigated in the context of climate networks. Most prominent include Pearson's correlation coefficient (Donges et al., 2009a; Gozolchiani et al., 2008; Halverson

and Fleming, 2015; Rheinwalt et al., 2015; Steinhaeuser et al., 2012; Yamasaki et al., 2008), mean distance (Tsonis et al., 2006), event synchronization (Agarwal et al., 2018b; Agarwal et al., 2017; Stolbova et al., 2014), spike synchronization (Goldberg, 2004; Kreuz et al., 2015), mutual information (Butte and Kohane, 1999; Deza et al., 2015; Paluš, 2018; Rosvall and Bergstrom, 2007), and more recently multi-scale event synchronization (Agarwal et al., 2017). The selection of the similarity measure is generally a function of (a) characteristics of the data (Gaussian, non-Gaussian) and (b) nature of relationships (linear, nonlinear, or multi-scale, etc.) between the nodes in the networks. For example, nonlinear connections between nodes are better captured by mutual information than the linear counterpart correlation. Malik et al. (2012) showed that for signals having an event like structure (non-Gaussian data) such as extreme precipitation, event synchronization performs better than Pearson's correlation coefficient (PCC), as it statistically measures the coincidence of single events and uses a dynamic (not fixed) time delay. Most of the methods mentioned above have been successfully applied in the construction of networks in order to understand several complex processes (Abbasi and Hossain, 2013; Basaras et al., 2013; Boers et al., 2014; Boers et al., 2015; Halverson and Fleming, 2015; Marwan et al., 2014). However, the present set of network node similarity measures are limited to capture multi-scale relationships in Gaussian data that are prominent in many natural processes. In the past, several studies have reported that complex systems, in general, are governed by coupling mechanisms, and these couplings are typically exhibited at unique temporal and spatial scales (Looney et al., 2014; Steinhaeuser et al., 2012). For example, Molini et al. (2010) investigated the presence of causality across rainfall timescales and showed a significant influence of coarser time scales on the finer scaled rainfall variability. Similarly,Varotsos et al. (2013) showed the presence of information transfer from larger to smaller time scales in the global daily mean surface air temperature. Hatala et al. (2012) and Sturtevant et al. (2016) identified the presence of scale-emergent processes in wetland methane exchange processes. On similar lines, a study Casagrande et al. (2015) revealed the dynamics of the soil moisture-temperature coupling over a wide range of temporal scales (from days to several months). More recently, Paluš (2018) showed evidence of scale-specific relationship emerging between sea surface temperature and large-scale climate oscillations such as El Niño/Southern Oscillation (ENSO) and North Atlantic Oscillation (NAO). These studies strengthen the hypothesis of the presence of multi-scale processes in geophysical systems. Casagrande et al. (2015), Miralles et al. (2014), Okin et al. (2009), and Peters et al. (2007) argue that multi-scale processes can act as a triggering mechanism for extreme events, abrupt regime transition, and pattern formation. Therefore, it is essential to investigate the process at several dominant temporal scales for improved understanding and modeling.

Advancing towards these goals will necessitate the development of a more detailed measure which can unravel underlying multi-scale relationships and changes in network structure over time. This is achieved using wavelets that obtain a set of scales reflecting the underlying oscillations at different scale levels (Fernández-Macho, 2012; Lark and Webster, 2001; C. Yang et al., 2011). In the past, wavelets based multi-scale correlation and coherence have already been successfully used to characterize the temporal scales interaction between extreme events and large-scale climatic oscillations (Achard, 2006; Addison, 2005; Agarwal et al., 2016b; Livina et al., 2008; Molini et al., 2010; Podtaev et al., 2008; Rathinasamy et al., 2014; Richiardi et al., 2011; Shusterman and Feder, 1994).

In this study, we suggest applying a multi-scale similarity measure using wavelets and PCC for network construction. In wavelet multi-scale correlation (WMC) measure, wavelets

are used to isolate the time scales of variation and correlation to identify the node–node interactions at these scales. This would help in systematically (1) identifying the dominant processes and the timescales they act on, and (2) understanding how the network evolves with temporal scales. The multi-scale similarity measure is applied to two prototypical examples and a real-time case study of sea surface temperature networks. The results are finally compared with the results derived with simple Pearson correlation.

The essential background on wavelet is discussed in Section 3.2; Section 3.3 is devoted to testing the performance of the proposed approach in assessing scale-by-scale interaction in a synthetic network of known multi-scale coupling. Following the test on synthetic data, Section 3.4 deals with the multi-scale nature of sea surface temperature. Finally, a further discussion of wavelet cross-correlation strength and criticality is provided in section 3.5, together with concluding remarks and an outlook on future developments.

## 3.2 Methodology

In order to develop WMC as a similarity measure for network formation in analyzing complex systems, the methods of Wavelet analysis (Addison, 2005) and Pearson correlation coefficient are combined. This section provides an introduction to the basic concepts of the wavelet analysis relevant to this study. We use the Maximal Overlapping Discrete Wavelet Transform (MODWT) which is a modification of the Discrete Wavelet Transform (DWT, Percival, 2008; Rathinasamy et al., 2014). We prefer MODWT to DWT because pyramidal structures of coefficients limit DWT. More clearly, in DWT, decimation is carried out so that only half of the coefficients of the detailed component are left at the current level, and half of the coefficients of the smooth version are recursively processed using high pass and low pass filters for coarser resolution levels. Due to this decimation, the number of wavelet coefficients is halved with each move to a coarser level. This problem may be overcome by using the stationary Maximal Overlap Discrete Wavelet Transform (MODWT) where the resulting gaps are filled using redundant information obtained from the original series.

MODWT decomposes the time series into different time scales or frequency components. The wavelet decomposition is realized using the two basis functions known as father wavelet ($\phi(t)$) and mother wavelet ($\psi(t)$) . The general admissibility conditions for $\psi$ to be called a wavelet function are

$$\int_{-\infty}^{\infty} \psi(t)\mathrm{d}t = 0 \tag{3.1}$$

$$\int_{-\infty}^{\infty} \psi(t)^2\mathrm{d}t = 1 \tag{3.2}$$

Any function $f(t)$ can be expressed through these basis functions and their scaled and translated versions are

$$f(t) = \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\Phi_{J,k}(t) + \quad \sum_k d_{J-1,k}\Phi_{J-1,k}(t) + \ldots + \sum_k d_{1,k}\Phi_{1,k}(t) \tag{3.3}$$

where $J$ is the total number of scales to be analyzed, and $k$ is in the range of 1 to $l$ (length of the time series). The coefficients $S_{J,k}$, are the approximation coefficients and

$d_{J,k},\ldots,d_{1,k}$ are the wavelet transform coefficients at scales $J$ to 1, while the functions $\phi_{J,k}(t)$ and $\{\Psi_{j,k}|j=1,\ldots,J-1,J\}$ are the basis functions which are obtained through translation and dilation of the father of father ($\phi(t)$) and mother ($\psi(t)$) wavelets function respectively. The mother and father wavelet is scaled (or dilated) by a factor $j$ and translated (or shifted) by a factor $k$ to give

$$\phi_{j,k}(t) = 2^{\frac{-j}{2}}\phi(2^{-j}t-k) \tag{3.4}$$

$$\Phi_{j,k}(t) = 2^{\frac{-j}{2}}\Phi(2^{-j}t-k) \tag{3.5}$$

Further, the values of the wavelet transform coefficients at each of the scale and the approximation coefficients at scale $J$ are estimated by:

$$d_{j,k} \approx \int \Phi_{j,k}(t)f(t)dt, \quad j=1,\ldots,J-1,J \tag{3.6}$$

and

$$s_{j,k} \approx \int \phi_{J,k}(t)f(t)dt \tag{3.7}$$

where the scaling coefficients $s_{J,k}$ capture the smooth trend of the time series at the coarse scale $2^J$, which are also called smooth coefficients; and the wavelet coefficients $d_{j,k}$, also known as detail coefficients can detect deviations from the coarsest scale to the finest scale. The original series $f(t)$ can be reconstructed by summing up the detailed components and the smooth components:

$$f(t) = S_{J,k} + D_{J,k} + D_{J-1,k} + \ldots + D_{1,k} \tag{3.8}$$

where

$$S_{j,k} = \sum_k s_{J,k}\phi_{J,k}(t), D_{J,k} = \sum_k d_{J,k}\Phi_{J,k}(t), \ldots, D_{1,k} = \sum_k d_{1,k}\Phi_{1,k}(t). \tag{3.9}$$

### 3.2.1. Wavelet Multi-scale Correlation

Consider two time series $\{X(t)\}$ and $\{Y(t)\}$ with the same length $T$. The wavelet multi-scale correlation (WMC) measure between both time series can be estimated as (Fernández-Macho, 2012; Lark and Webster, 2001)

$$WMC = \rho_{X,Y}^{l_j} \equiv \frac{COV_{X,Y}^{l_j}}{Var_Y^{l_j} Var_X^{l_j}} \tag{3.10}$$

where

$$Var_X^{l_j} = \frac{1}{T_j}\sum_{t=M_j-1}^{T-1}[d_{j,t}^X]^2 \tag{3.11}$$

$$Var_Y^{l_j} = \frac{1}{T_j}\sum_{t=M_j-1}^{T-1}[d_{j,t}^Y]^2 \tag{3.12}$$

$$COV_{X,Y}^{l_j} = \frac{1}{T_j} \sum_{t=M_j-1}^{T-1} d_{j,t}^{X} d_{j,t}^{Y} \tag{3.13}$$

$d_{j,t}^{(.)}$ denotes the MODWT wavelet coefficient of variables $\{X,Y\}$ at scale $l_j$; $T_j' = T - M_j + 1$ stands for the number of coefficients unaffected by the boundary; $M_j = (2^j - 1)(M-1) + 1$ represents the length of the scale $l_j$ wavelet filter, and $M$ is the width of the wavelet filter. The multi-scale correlation measure $\rho_{X,Y}^{l_j}$ denotes the scale-wise correlation between $X$ and $Y$ at different $l_j$ scales. Like the Pearson correlation coefficient (PCC), the value of $\rho_{X,Y}^{l_j}$ ranges between $-1$ and 1. $\rho_{X,Y}^{l_j}$=0 implies the variables $X$ and $Y$ are not correlated at scale $l_j$. $\rho_{X,Y}^{l_j} = 1$ and $\rho_{X,Y}^{l_j} = -1$ indicate that the variables are perfectly correlated and anti-correlated, respectively. The values of $\rho_{X,Y}^{1l_j}.....\rho_{X,Y}^{l_j}$ indicate the strength of the relation between $X$ and $Y$ at different temporal scale bands.

### 3.2.2. Network Construction

In the context of the application of WMC as a network measure, consider a network $Z = \{N,E\}$ containing a set of $N$ nodes together with a set of $E$ edges $\{j,k\}$ which are 2 element subsets of $N$. Let each of the nodes be characterized by a time series of a certain variable of interest. The WMC between two nodes is estimated using Eq.3.10.

We begin the network construction at each temporal scale by finding the WMC measure between each pair of nodes. Since an inverse relationship is equally relevant in the present application, we set the edge weight to absolute values of WMC $|\rho_{X,Y}^{l_j}|$ (Steinhaeuser et al., 2011; Yamasaki et al., 2008). By doing so, we generate the $N \times N$ matrix of $\rho^{l_j}$ values for each scale $l_j$. We are only interested in the highly correlated links. Therefore we apply a threshold to the $|\rho^{l_j}|$ values which is the $\theta = 95th$ percentile of all $|\rho^{l_j}|$ values. This converts the $|\rho^{l_j}|$ matrix to a binary matrix, the adjacency matrix ($A$) of an undirected and unweighted network.

$$A_{m,n} = \begin{cases} 1, & \text{if } |\rho_{m,n}^{l_j}| \geq \theta^{l_j}. \\ 0, & \text{otherwise.} \end{cases} \tag{3.14}$$

Here, $\theta^{l_j}$= $95th$ percentile is a chosen threshold for each scale $l_j$, and $A_{m,n} = 1$ denotes a link between the $m^{th}$ and $n^{th}$ nodes and 0 denotes otherwise.

The application of the WMC as the network measure reveals the network topology at each of the $l_j$ scales and allows the investigation of the network at different temporal scales. In the following, we demonstrate the potential of this proposed method with prototypical examples and real world application with a finite number of nodes.

## 3.3 Testing WMC with synthetic data

The prototypical examples aim to test the efficacy of the proposed WMC method over the traditional Pearson's correlation as an effective network reconstruction measure while dealing with multi-scale processes. Following the approach by Agarwal et al. (2017), we test WMC using a set of case studies including stationary and non-stationary synthetic data. The different scales are included as periodic signals of different frequency. Correlated

noise is added to mimic real-world processes where stochasticity and autocorrelation are important features. The mathematical details of the case studies and wavelet power spectra are given in Table3.1 and Fig.3.1.

**Table 3.1.:** Details of synthetic models. Here, $f$ denotes a function.

| Mathematical expression | | | Other details |
|---|---|---|---|
| Periodic signal | $S1 = \cos(2\pi\omega_1)$ | $\omega_1 = .4t;$ | refer Fig.3.1 |
| | $S2 = \cos(2\pi\omega_2)$ | $\omega_2 = .09t$ | |
| | $S3 = \cos(2\pi\omega_3)$ | $\omega_2 = .01t$ | |
| Case I(a) ($t = 1:1000$) | $X_i t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t + Y_1$ | $1 < i \leq 10$ | $Y_1 = f(S1)$ |
| | $X_i t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t + Y_2$ | $11 < i \leq 20$ | $Y_2 = f(S2)$ |
| | $X_i t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t + Y_3$ | $21 < i \leq 30$ | $Y_3 = f(S1, S2)$ |
| Case II) ($t = 1:2000$) | $X_i t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$ | $1 < i \leq 10$ | $Y_1 = f(S1)$ |
| | $(Y_1)_{t=1:1000} + (Y_2)_{t=1001:2000}$ | | $Y_2 = f(S2)$ |
| | $X_i t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$ | $11 < i \leq 20$ | $Y_3 = f(S1, S2)$ |
| | $(Y_3)_{t=1:1000} + (Y_4)_{t=1001:2000}$ | | $Y_4 = f(S2, S3)$ |
| | $X_i t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$ | $21 < i \leq 30$ | $Y_5 = f(S1, S2, S3)$ |
| | $(Y_5)_{t=1:2000}$ | | refer Fig.3.3 |

**Case I.** We construct a synthetic network (Fig.3.2) of 30 nodes ($X_i(t); i = 1, 2, ...30$). For each of the nodes $X_i(t); i = 1, 2, ...30$) different time series are generated by combining a $2^{nd}$ order autoregressive process (AR) and periodic signals (Table3.1). The mathematical expression for the periodic signal and time series formulation are provided in the Table 3.1.

The mathematical construction of the time series (Table3.1) shows that 30 nodes of the network are classified into three group such that each group encompasses a particular cosine function (frequency). Group I: node 1 to 10; Group II: node 11 to 20 and Group III: node 21 to 30 encompasses the cosine functions $Y_1$, $Y_2$ and $Y_3$ with different frequencies, thus, different scales. The particular cosine functions (signal) are plotted in Fig.3.1(a)-(c) and the corresponding power spectra are shown in Fig.3.1(d)-(f).

The $\phi_1$=0.25 and $\phi_2$=0.4 are autoregressive constants, while $\varepsilon_t$ represent white noise with the signal/noise ratio of amplitude (SNR)= 0.90. The resultant node's time series have features that are often found in geophysical, hydrological and climatic data, where auto-correlated noise and high frequency, small-scale processes are superimposed on low frequency, coarse-scale processes (Agarwal et al., 2017; Hu and Si, 2016). Such structures are widespread in time series of seismic signals, turbulence, air temperature, precipitation, hydrologic fluxes, or the El Niño/Southern Oscillation (ENSO). They can also be found in spatial data, e.g., in ocean waves, seafloor bathymetry, or land surface topography (Hu and Si, 2016).

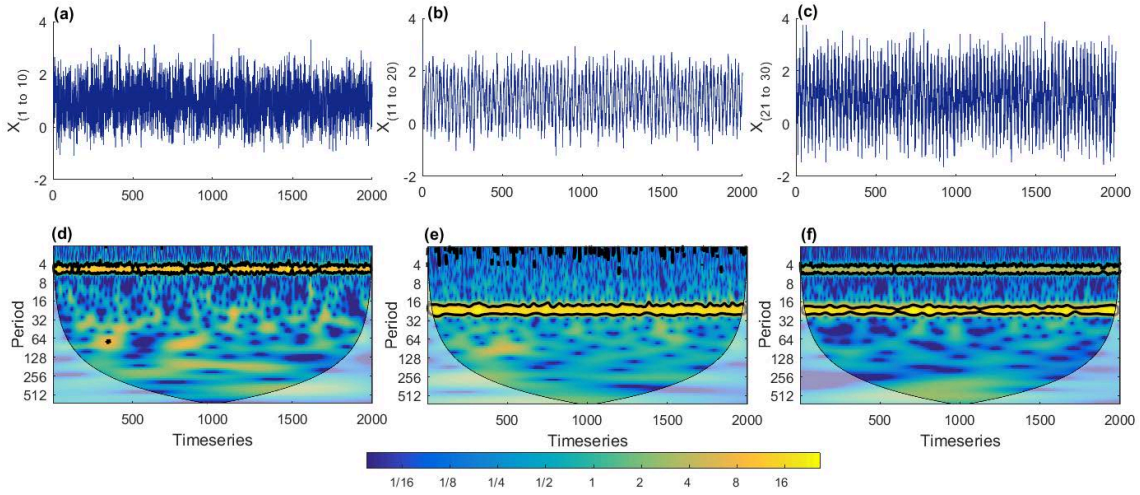



**Figure 3.1.:** The top row (a-c) shows the plot of the synthetic time series generated for network formation. Ten simulations were generated for each group of nodes. The bottom plot (d-f) shows the wavelet power spectra for one single time series from each of the group. The regions within the black contours denote the 95% significant values and the cones of influence denote the extent of the boundary distortion.

First, the network is constructed using absolute correlation values of PCC applying the $95th$ percentile threshold (Fig.3.2a). Even though, Group I (nodes 1 to 10, pink circles in Fig.3.2) and Group III (nodes 21 to 30, blue circles) nodes share a strong similarity owing to the common periodicity (signalS1, Table3.1) the network link does not reflect the expected similarity (strong connections) in the dynamics. Hence, it does not give an appropriate impression of the connection. A similar misleading result is inferred between the Group II and Group III nodes. Further, owing to the strong self-similarity between the nodes of Group III, we observe links connecting the nodes $21-30$ with each other. However, self-similar links are absent in the Group I and Group II because of the low signal to noise ratio, which masks the underlying correlation.

Next, for the same 30 nodes, the similarity measure WMC $|\rho^{l_j}_{X,Y}|$ is calculated at different timescales and the corresponding network is formed using a $95th$ percentile threshold (Fig.3.2b-h). Out of 7 scales considered, the links are present only at scales 2 (period $4-8$) and 7 (period $128-256$) owing to the common cosine function (frequency) between Group I & III (i.e. $Y_1$), and Group II & III (i.e. $Y_2$). Since at other timescales no significant common information is present,we do not see any linkage between the nodes as expected.

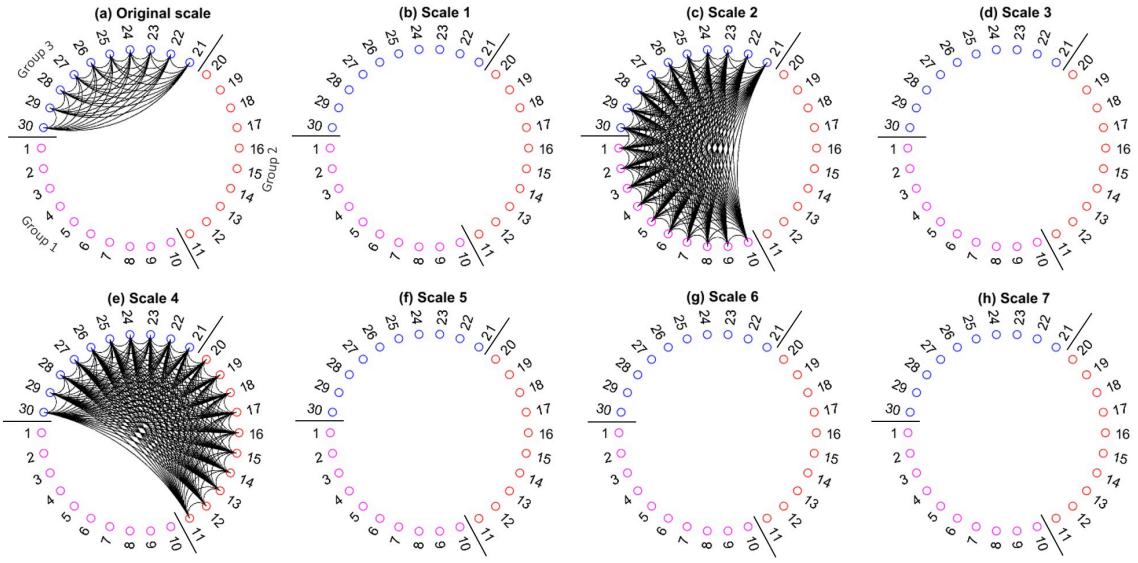



**Figure 3.2.:** Network structure and significant links obtained using the (a) PCC (original scale) and (b)–(h) WMC at scales $1-7$. A significant link is defined when the value of the correlation is higher than the 95th percentile.

**Case II.** Here, we attempt to test the efficiency of the proposed approach for a non-stationary signal wherein the similarity between two different variables is changing with time. In this case, we generate three groups of time series following a similar approach as in the previous case except that the groups have a non-stationary relationship in terms of time. The mathematical formulation used for the generation of the time series (case II) is given in Table3.1, and the common periodicities present between the three groups are shown in Fig.3.3 for easy visualization. Fig.3.3 shows that the common periodicity between all the 30 nodes is S1 for the time period $t = 1:1000$ and $S2$ for $t = 1001:2000$. Similarly, other details are intuitive from Fig.3.3. The particular cosine functions (signal) are plotted in Fig.3.4(a) to (c) and the corresponding power spectra are shown in Fig.3.4(d) to (f).

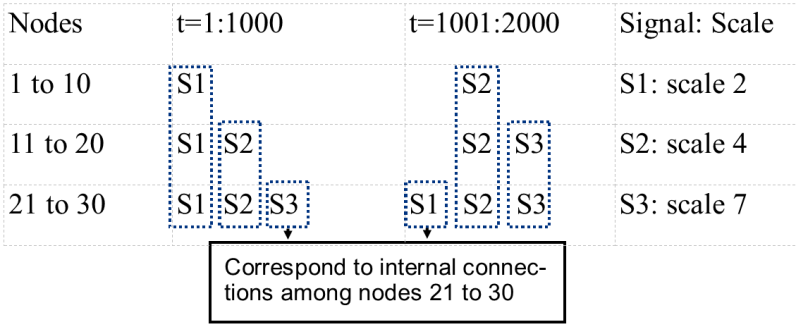



**Figure 3.3.:** Visualization of the common periodicities presents among the group of 30 nodes.

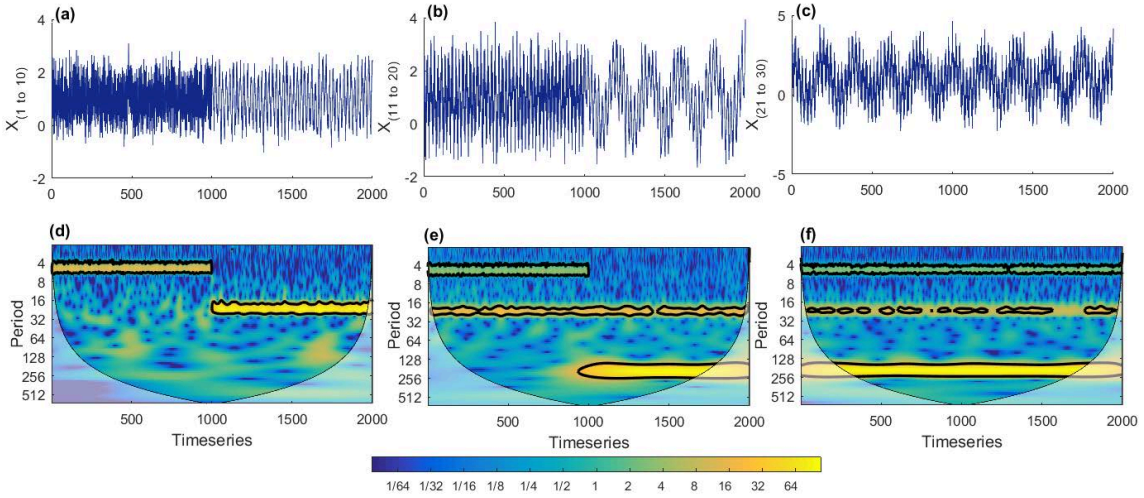



**Figure 3.4.:** The top row (a-c) shows the plot of the synthetic time series generated for network formation. Ten simulations were generated for each group of nodes. The bottom plot (d-f) shows the wavelet power spectra for one single time series from each of the group. The regions within the black contours denote the 95% significant values and the cones of influence denote the extent of the boundary distortion.

For all 30 nodes presented, initially, we create a network using absolute correlation values of PCC (Fig.3.5a) and then WMC (Fig.3.5(b)-(h)) by applying the $95th$ percentile threshold. We observe that PCC only uncovers the internal connection among a group of nodes (Fig.3.5a), whereas WMC captures the true relationship present in the signal of the 30 nodes (Fig.3.5(b)-(h)). For instance, in Fig.3.5c and 3.5e all 30 nodes are significantly correlated corresponding to the presence of the signal $S1$ and $S2$ (Fig.3.3) at scale 2 and 4 respectively. Similarly, nodes $11-30$ have a common periodicity ($S3$, Fig. 3) at scale 7 hence the same is captured in Fig.3.5h. As expected, all other temporal scales do not show any significant correlation among the nodes.

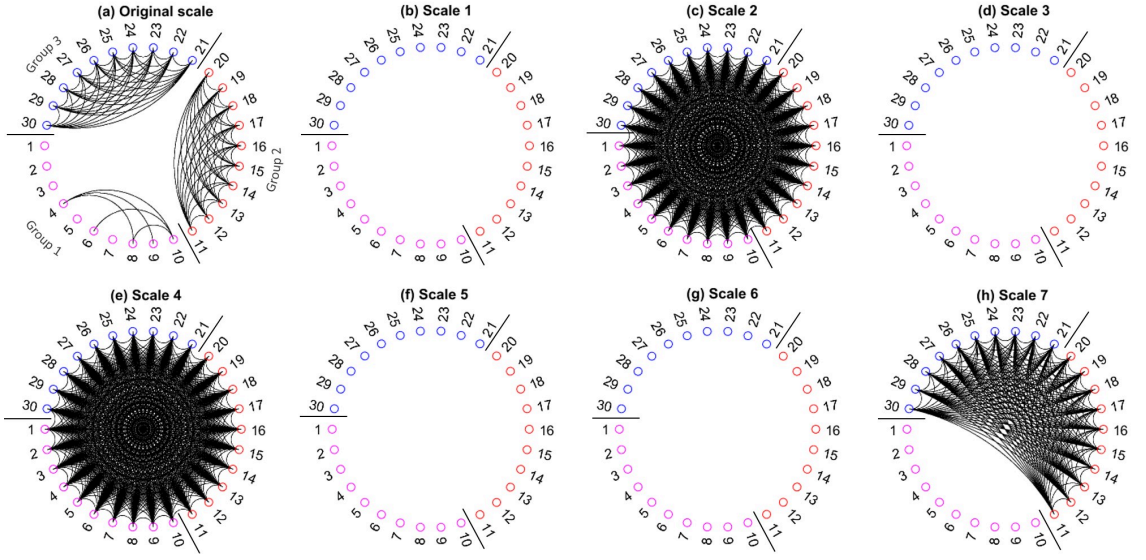


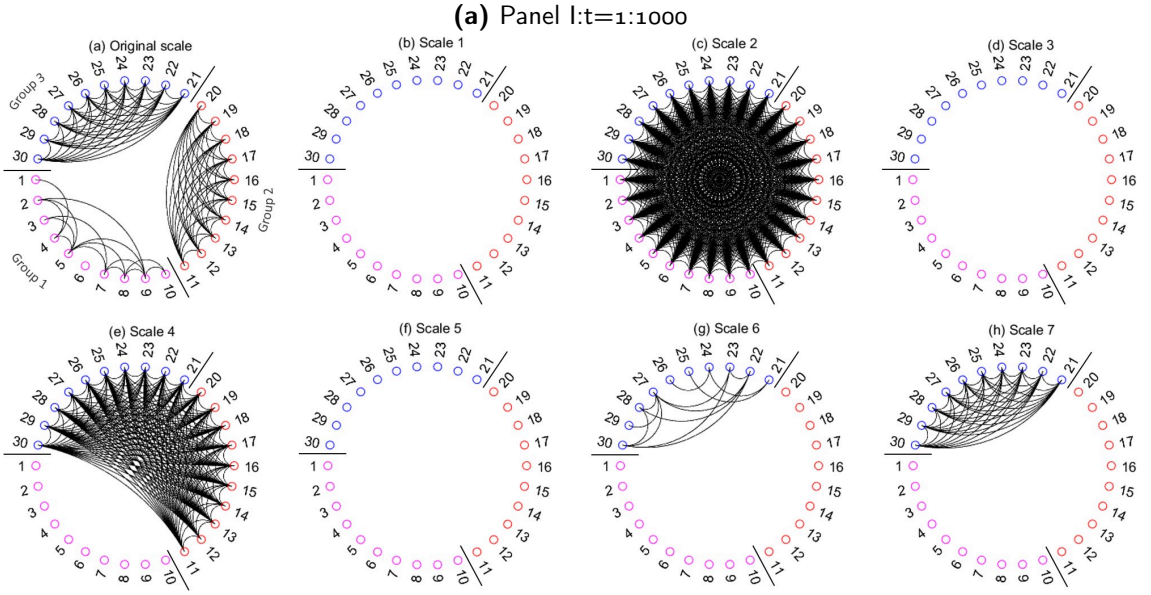

**Figure 3.5.:** Network structure and significant links obtained using the (a) PCC (original scale) and (b)–(h) WMC at scales $1-7$. A significant link is defined when the value of the correlation is higher than the $95th$ percentile.

Albeit, WMC is able to capture the dynamics of the non-stationary signal, still the time evolution pattern is not revealed completely. Hence, we propose to break the signal into two time windows $t = 1:1000$ and $t = 1001:2000$ respectively.

For the time period $t = 1:1000$, the network formed using PCC (Panel I, Fig.3.6) does not show any links between the node groups even though there is a significant common periodicity between the nodes. However, constructing a network using WMC clearly captures the significant relationships between the node groups. For example, at Scale 2, all the node groups are connected with each other owing to the common periodicity ($S1$). Similarly, at scale 4, only nodes belonging to Groups II and III are linked because of the common periodicity ($S2$). Further at scales 7 the significant links are present showing the intra group connection present in Group III owing to the periodicity ($S3$).

Now, looking at the network formed for the time period $t = 1001:2000$, the PCC based network (Panel II, Fig.3.6) does not clearly demarcate the changes in the signal, rather this network is very similar to what was observed during $t = 1:1000$ (Panel I, Fig.3.6). On the other hand, the network formed using the WMC measure clearly delineates the changes in the connection between the node groups. For instance, the common periodicity between all the node groups in $S2$ and that is clearly captured at Scale 4. Similarly, the significant links at scales 7 between the nodes belonging to Group II & III is due to the presence of the common periodicity ($S3$). The intra group links present between nodes $21-30$ at scale 2 is due two to the presence of periodicity $S1$ in that group of nodes.Further, few scattered links at scale 6 might be credited to the overlap/spill of information across scales in the wavelet transform. This phenomenon is more dominant at higher scales since moving from a finer to a coarser scale the width between the scales decreases (for more details refer Polikar (1996) and Addison (2005).

Thus from this case study, we conclude that the WMC measure clearly captures the non-stationary relationship between the nodes and also clearly unravels at which scale the changes occurred.



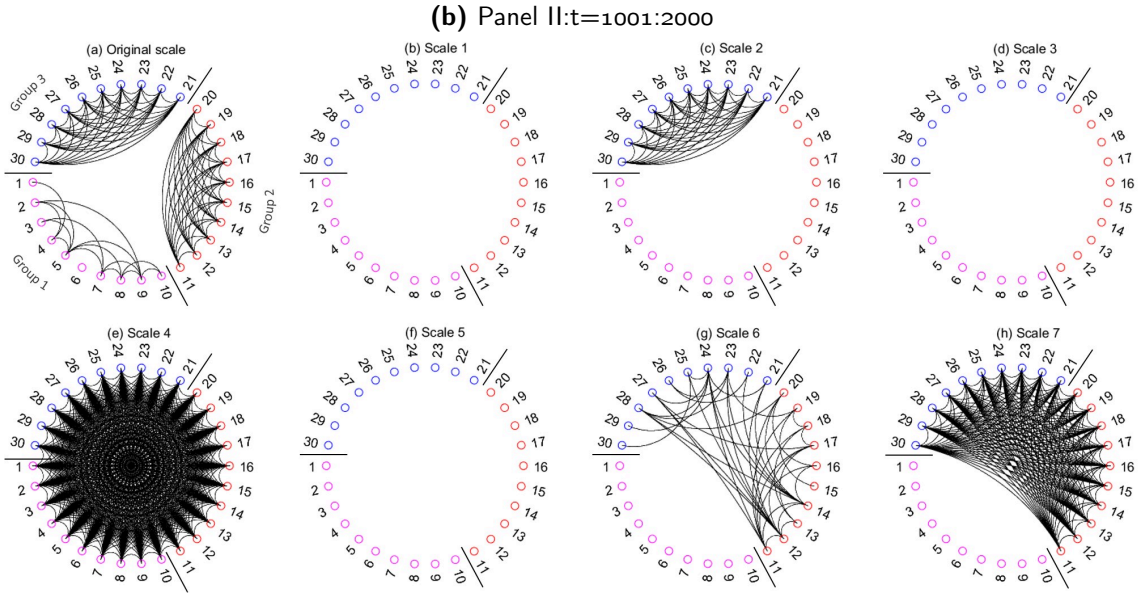



**Figure 3.6.:** Network structure and significant links obtained using the PCC (original scale) and WMC at scales $1-7$. A significant link is defined when the value of correlation is higher than the $95th$ percentile. Panel I corresponds to $t = 1:1000$ and Panel II corresponds to $t = 1001:2000$.Few scattered links at scale 6 might be credited to the overlap/ spill of information across scales in wavelet transform.

From the presented two case studies, it can be inferred that the PCC is unable to capture the significant multi-scale similarity between two variables, whereas the proposed WMC categorically captures the similarity at the specific time scales.

## 3.4 Real world application

To illustrate the applicability of the proposed multi-scale similarity measure on real-world data, we present the global monthly sea surface temperature (SST) network at different scales constructed using PCC and WMC measure. Here, we have used monthly gridded temperature data of $(2^\circ \times 2^\circ)$ grid resolution for the period $1979-2015$. The data is provided by the NOAA/OAR/ESRL PSD and freely available at https://www.esrl.noaa.gov/psd/. As a preprocessing step of the data we have removed all grid points of SST with missing values or gaps, hence a total of 9456 grid points are considered in this study; and further, we calculate anomaly series by subtracting the mean for each month from each time series.

Initially, we construct the network using absolute values of PCC similarity measure on an observed scale, and then we begin the network construction at another temporal scale based on the WMC measure between each pair of nodes. By doing so, we generate a similarity matrix $|\rho^{l_j}|$ of size $9456 \times 9456$ on a particular scale $l_j$ and apply the $\theta = 95th$ percentile threshold to obtain the adjacency matrix (A). In this study, the maximum scale J was taken as 7 owing to the distortion created due to the boundary effects of wavelet transform (Percival and Walden, 2000).

The comparison of the results obtained using PCC and WMC are shown in Fig.3.7a

and 3.7(b-h). We present the $95th$ percentile values of $|\rho^{l_j}|$ at each SST grid point to all other grid points, which is equivalent to the number of other nodes this node is highly correlated to (i.e., corresponds to the network measure degree). Additional plots (Figs.3.8 & 3.9) that show positive and negative values of $|\rho^{l_j}|$ separately are presented in the Appendix.

Analysis of the PCC based plot (Fig.3.7a) shows that regions around the Indian and the Pacific Ocean are correlating on the order of $\approx 0.60$ as well as regions in the North Atlantic, the plot does not reveal on what time scales these correlations are the most prominent. It also does not reveal whether there exists one link between all regions of a high correlation or whether there are several different links between these regions. This emphasizes the need for a more sophisticated analysis measure like the WMC as analyzing the $|\rho^{l_j}|$ values at different scales results in far more detailed findings in the SST patterns as described in the following.

Moving from lower to higher temporal scales we find that there is not much significant interaction between SST at the finer, intra-annual temporal scales (up to $4-8$ months). This can be expected as we are looking at monthly anomalies and thus removed the annual cycle from the data. However, at the $8-16$ months scale (Fig.3.7e), we observe two zones with a relatively large number of significant connections in the equatorial Pacific and the Indian Ocean that are in accordance with the pattern observed during the ENSO phase and the Indian Ocean Dipole (IOD) which are known to impact each other via the atmosphere (Luo et al., 2010; Stuecker et al., 2017; Yamagata et al., 2013). When looking at the signs of the correlation one can see that the ENSO tongue pattern west of South America is positively correlated with the SST pattern in the Indian Ocean (Fig.3.8e) and both are negatively correlated with the horseshoe pattern about the ENSO tongue (Fig.3.9e). This is the canonical SST pattern associated with an ENSO event and the corresponding IOD phase that can be excited by the ENSO event (Li et al., 2008). At inter-annual scale ($16-32$ months, Fig.3.7f, Fig.3.8f, Fig.3.9f), these patterns become more pronounced as at least ENSO events can be predicted and act on scales of up to two years (Chen et al., 2004).

At the scale of $32-64$ months (Fig.3.7g), the region of high correlation in the Eastern Pacific Ocean is fading away but coexist with a spot of high correlations on the Northern Pacific– forming together a pattern known as Pacific Decadal Oscillation (PDO) which has been in a warm phase since the late 1970s, but with time intervals of the cold phase, this variability is forced by Rossby waves (Newman et al., 2016) and can be seen in our analysis. Additionally, a new spot of high correlations around the subtropical part of the North Atlantic Ocean appears. The latter becomes more prominent on the $64-128$ months scale (Fig.3.7h). It resembles the SST patterns that are linked with the North Atlantic Oscillation (NAO) an atmospheric mode affecting SST (Visbeck et al., 2001). This becomes even clearer when analyzing at the positive correlations (Fig.3.8h). One can see that the region in the subpolar North Atlantic and the subtropical North Atlantic are positively correlated, a known response of the ocean surface temperatures to the NAO. When focusing at the regions that have a high negative correlation with the other nodes, the most prominent feature is the correlation between the Southern Ocean around Antarctica and the subtropical/tropical the Pacific Ocean, west of South America. This correlation indicates what has just been confirmed by Ferster et al. (2018): A link between SST in the Southern Ocean and the Southern Annular Mode (SAM) with teleconnections to ENSO events. As the SAM has a positive trend over the analyzed time period ($1979-2015$), this pattern emerges on the long time scales of several years. The positive states of SAM correlate with negative trends

(cooling) trends in SST in the high-latitude Southern Ocean and positive (warming) trends within the Southern Hemisphere sub-tropics and mid-latitudes (Ferster et al., 2018).

In Conclusion, it can be said that the changes in the correlation patterns in the SST field when viewing at the different scale regimes as well as the difference in patterns between the positive and negative correlation strength, shows the complexity of the climate system and emphasize the importance of detecting correlations at multiple time scales. The PCC based plot (Fig.3.7a) does not give the information needed to divide between the different modes of climate variability that are affecting SSTs in different regions and on different scales. But by applying the WMC measure we can re-confirm known physics-based associations such as well-known ENSO patterns in Pacific at interannual scale, the impact of the NAO on SST in the North Atlantic, but also the coupling of ENSO and IOD at interannual and interdecadal scale as well as the coupling between SAM, Southern Ocean SST and ENSO. This implicitly affirms the validity of our approach and furthermore shows the efficacy of the method in detecting a variety of climate modes on different temporal and spatial scale at one.
The proposed approach therefore clearly highlights the importance of advanced detecting methods, like the WMC, when dealing with such complex and big data, like global SSTs.

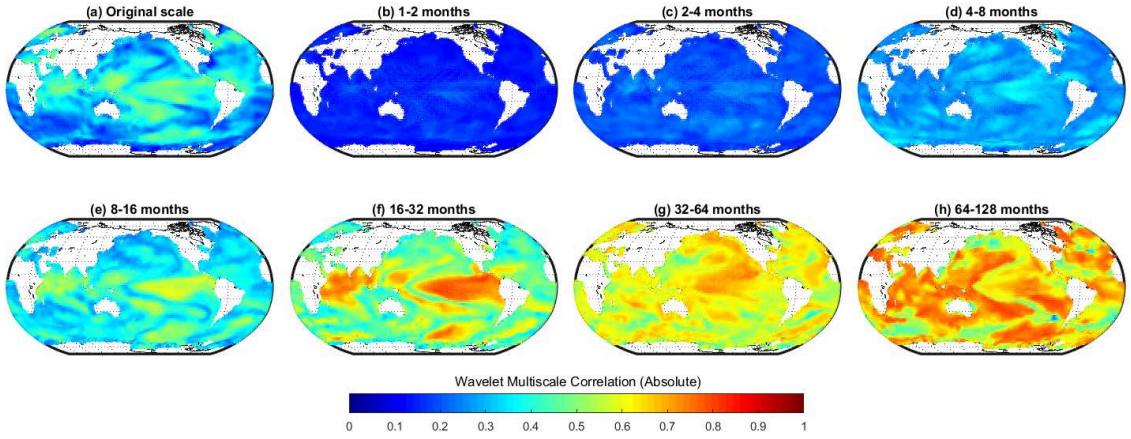



**Figure 3.7.:** 95*th* percentile of absolute wavelet multi-scale correlation (WMC) values for the SST network constructed on multi-scale using the PCC (original scale) and WMC at scales $1-7$. Color represents the strength of correlation on each scale.

## 3.5 Conclusion

We have introduced a novel methodology to infer multi-scale interactions from observations of dynamical climate systems that evolve over diverse temporal scales. The adopted metric — wavelet multi-scale correlation (WMC) — relies on the direct estimation of scale-by-scale correlations in the wavelet domain and is formulated here as an ensemble statistics across different temporal subsamples, with associated scale-dependent significance levels. Using a range of prototypical situations and a real-world case study, it is evident that the analysis at single observational scales using Pearson's correlation is not enough to capture the underlying multi-scale features of complex systems with many nodes, as it is often the case with geophysical data, and thereby resulting in poor network reconstruction.

On the other hand, the significance of the results using WMC shows its ability to unravel the hidden similarity structure at different temporal scales. The obvious advantages of the proposed approach over the other single scale measures are, that (i) we can study the cross-scale relationship among dataset and systems on different time scales, (ii) network obtained using the WMC measure offer a wider perspective to unravel the network scale varying dynamics and (iii) efficient in dealing with multi-scale non-stationary complex systems, since, as shown, the effect of non-stationary can be handled with the help of wavelets.

On investigating global sea surface temperatures at multi-scale, i.e., with the WMC measure, new insights were obtained which had remained hidden at the observational scale, i.e., a simple PCC. Our findings re-confirm the known climate modes that affect SST like the ENSO patterns in the Pacific Ocean at inter-annual scale and the NAO that affects SST in the North Atlantic on the intra-annual scale. We also detect the coupling of ENSO and IOD at inter-annual and interdecadal scale as well as the coupling between SAM, Southern Ocean SST, and ENSO. We thereby affirm the validity of the WMC measure to detect climate modes on different temporal and spatial scales.

The proposed approach offers a new paradigm and possibilities for applications in other natural processes where interactions at multiple time scales exist e.g. in neuroscience, ecology and economy. By studying on what time scale the factors are related and how the network properties emerge at that time scale, we may be in a better position to interpret the complex system and provide accurate predictions for future conditions. As a future work, the investigation could be extended to understand the causal effect of different climate drivers using the proposed WMC measure. This might give an understanding of the time scale relationships, which could be used to detect, attribute, and understand complex systems.

## 3.6 Acknowledgment

This research for authors (AA, NM & JK) was funded by Deutsche Forschungsgemeinschaft (DFG) (GRK 2043/1) within the graduate research training group Natural risk in a changing world (NatRiskChange) at the University of Potsdam (http://www.uni-potsdam.de/natriskchange). RM would like to thank the Humboldt Foundation for supporting this research through the Alexander Von Humboldt Fellowship award. Also, he would like to thank DST, India for the Inspire Faculty Fellowship Award. The authors gratefully thank Roopam Shukla (Department of Natural Resources, TERI University, New Delhi), and Niharika Tyagi (Teri University, New Delhi) for helpful suggestion and reading the paper.

## 3.7 Appendix

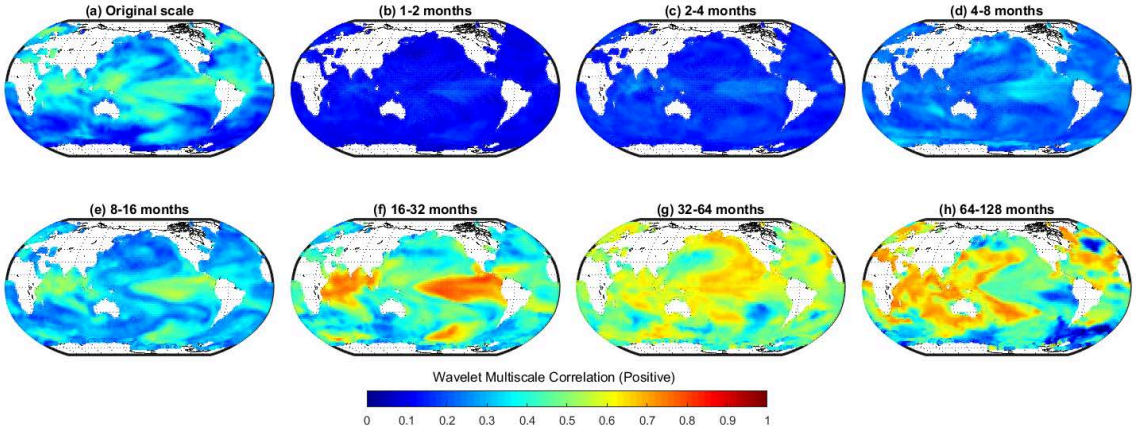



**Figure 3.8.:** 95*th* percentile of positive wavelet multi-scale correlation (WMC) values for the SST network constructed on multi-scale. Color represents the strength of correlation on each scale.

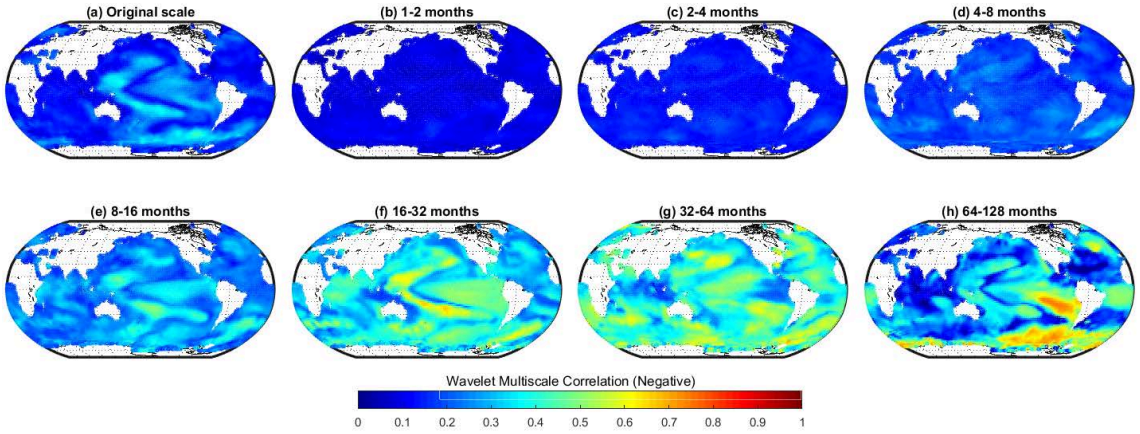



**Figure 3.9.:** 95*th* percentile of negative wavelet multi-scale correlation (WMC) values for the SST network constructed on multi-scale. Color represents the strength of correlation on each scale.

## 3.8 Supplementary figures

For clear visualization of the identified patterns in extratropics Northern Hemisphere (NH) and extratropics Southern Hemisphere (SH), we also have plotted similar spatial maps (Figure 3.7, 3.8, &3.9) in orthographic projection with North and south pole at the center. These plots are not included in the main text but it can give a deeper insight to interested readers. All the figures are available in the Appendix C.

# Part II.

# Application

# Chapter 4

# Detection of short- and long-range teleconnections in SST patterns on different timescales

*Agarwal, A., Caesar, L., Marwan, N., Rathinasamy, M., Merz, B. and Kurths, J.: Short and long-range Teleconnections in oceans and atmosphere. (Climate dynamics-in review).*

Sea surface temperature (SST) anomaly patterns can – as surface climate forcing – affect the weather at large distances. This is why following an El Niño event major global climate anomalies occur. This paper characterizes the links between the cells of a global SST grid data set at different temporal and spatial scales with the help of climate networks. The networks are constructed using wavelet multi-scale correlation. We identify and visualise the SST patterns that develop very similarly over time and distinguish them from those that have long-range teleconnections to other ocean regions. Our findings re-confirm accepted knowledge about known highly linked SST patterns like El Niño/ Southern Oscillation and the Pacific Decadal Oscillation, but also suggest new insights into the characteristics and origins of long-range teleconnections.

## 4.1 Introduction

The ocean covers more than two-thirds of the Earth's surface. Given the higher heat capacity of water as compared to air (by approximately a factor of 4) and greater mass in the ocean than the atmosphere, the ocean can store about 1000 times as much heat as the atmosphere (Solomon et al., 2007). Although it takes time to transport heat in or out of the ocean, and even the ocean mixed–layer has a larger heat capacity than the atmosphere and can, therefore, strongly influence the climatic conditions. As a consequence, the ocean is capable of altering atmospheric conditions on seasonal and longer timescales (Bigg et al., 2003; Ozturk et al., 2018b).

As the sea surface forms a direct boundary between ocean and atmosphere, the sea surface temperature (SST) is of particular importance for setting the atmospheric conditions. In the opposite direction, large-scale atmospheric circulation patterns (e.g., quasi–stationary Rossby waves) form atmospheric linkages between different oceanic regions around the globe and control the spatial extend of the interactions among oceans (Griggs and Noguer, 2002). These long distance interactions, referred as teleconnections are marked by a significant

correlation of the climatological variables. Teleconnections have periodicities ranging from months to several decades. Teleconnections have been identified to affect temperature, precipitation, storm tracks, severe weather, and drought (Mamalakis et al., 2018). An in-depth knowledge of the dynamics and patterns of the different teleconnections, therefore, is useful for advanced understanding in climate science.

Extensive research has analysed the teleconnections in the atmosphere and oceans considering a broad range of temporal and spatial scales, in both the tropics and the extratropics (Fovell et al., 2017; Hastenrath and Hastenrath, 1991; Huang and Shukla, 2005; Kucharski et al., 2010; Ramirez et al., 2017). Among the most well-known teleconnections are a number of defined patterns that are marked by a strong correlation of the SSTs at different places like the Pacific Decadal Oscillation (PDO) and the North Atlantic Oscillation (NAO) but also the link between El Niño/Southern Oscillation (ENSO) and the Asian monsoon system (Azad and Rajeevan, 2016; Behera et al., 1999; Huang and Shukla, 2005; Kucharski et al., 2010; Steinhaeuser et al., 2012; Tsonis et al., 2008; Zhou et al., 2015).

Recent studies have raised concern that our understanding of climate processes remains incomplete, because interactions in multi-scale systems, like the climate system, are typically exhibited at unique or at multiple temporal and spatial scales (Casagrande et al., 2015; Paluš, 2018; Steinhaeuser et al., 2012). Methods are required that help to unveil the interactions in the climate system at different scales, which usually remain hidden when looking at a particular scale only. Also, it is important to quantify the spatial distance of teleconnection variability between remote ocean regions at different timescales which could give us the quantitative understanding of short-range and long-range coupling between different oceanic regions.

To uncover such spatially and temporally variable interactions we use the complex network approach (Newman, 2003), which provides a powerful framework to dissect interdependence and connection strength, while retaining only the most significant network topology. In recent years complex network analysis has facilitated the identification of spatial patterns in various research fields (Gu et al., 2015). One such example is climate networks (Tsonis et al., 2006; Donges et al., 2009b) that are constructed from climate reanalysis data. Here the network nodes (or vertices) are the grid cells of climate data sets, and links between nodes (or edges) are established using a similarity measure between the times series associated with the different grid cells. Only the strongest links, filtered using either a predefined link density or correlation cutoff threshold, are retained in the final networks to help uncover the most important network characteristics. Linear correlation measures and several network measures (degree, clustering, betweenness, community structures) have been used to capture regional and global dependence structures within and among climate variables (Marwan et al., 2009; Donges et al., 2009a; Sivakumar and Woldemeskel, 2014), like the identification of regions that are most significantly affected by ENSO events (Paluš, 2018). In addition, nonlinear similarity measures have been used to explore uni- or multivariate pairwise relationships, such as mutual information (Agarwal et al., 2017; Stolbova et al., 2014; Donges et al., 2009a; Rheinwalt et al., 2016) and transfer entropy (Runge et al., 2015). To date, the complex network approach for the analysis of climate data has mostly been used to investigate patterns at one reference timescale only. However, the extension of the climate network to multiple timescales can reveal additional and important information as demonstrated recently (Agarwal et al., 2017).

In this paper, the wavelet-based correlation method is used by combining the wavelet transform and Pearson correlation. We present a well-accepted climate network constructed from global monthly SST data, with a focus on identifying dependence structures on different timescales over the entire globe. We, therefore, construct climate networks on eight different timescales, ranging from the monthly to interdecadal scale, using the wavelet-based correlation measure and interpret the detected short and long-range links. Our findings re-confirm accepted knowledge about highly linked SST patterns like El Niño/Southern Oscillation and the Pacific Decadal Oscillation, but also provide new insights into the characteristics and origins of long-range teleconnections.

## 4.2 Data characteristics

The climate networks developed are based on the global monthly sea surface temperature (SST) data provided by NOAA's Earth System Research Laboratory (ESRL). The data (Extended reconstructed SST V3b) has a spatial resolution of 2.0–degree latitude x 2.0–degree longitude and is given for the time period $1979-2015$. The data is freely available at https://esrl.noaa.gov/psd/. As a preprocessing step we have removed 1056 grid points (out of total 10512) with missing values or gaps, hence in total 9456 grid points are considered. To avoid artefacts due to autocorrelation and seasonality, we removed the seasonal cycle and normalized the data. Specifically, we calculate for every month (i.e., separately for all Januaries, Februaries, etc.) the long-term mean and standard deviation. Each data point is then normalized by subtracting the mean and dividing by the standard deviation of the corresponding month at that grid cell. This normalization significantly reduces temporal autocorrelation in the time series.

### 4.2.1. Wavelet multiscale correlation

To construct the SST network, each grid cell is considered as a node, and links between each pair of nodes are set up based on a statistical relationship between them. The similarity measure used is the wavelet multiscale correlation (WMC) between the monthly anomaly series proposed by Agarwal et al., (Resubmitted with minor revisions). The proposed WMC method is a combination of two well established methods "maximum overlap discrete wavelet transformation" and Pearson's correlation. All the mathematical details are presented in the Appendix D.

### 4.2.2. Network construction

Formally, a network or graph is defined as an ordered pair $G = \{N, E\}$, containing a set of nodes $N = \{N_1, N_2, \ldots N_N\}$ together with a set $E$ of edges $\{i, j\}$ which are 2-element subsets of $N$. In this work, we consider undirected and unweighted graph ($G$), where only one edge can exist between a pair of nodes and self-loops of the type $\{i, i\}$ are not allowed. Hence, edges simply show connections between nodes, and each edge can be traversed in either direction. This type of graph can be represented by the symmetrical adjacency matrix (Stolbova et al., 2014)

$$A_{i,j} = \begin{cases} 1, & \{i, j\} \in E \\ 0, & \{i, j\} \notin E. \end{cases} \tag{4.1}$$

To construct a climate network (Fig. 4.1), each grid cell of SST data is considered as a node of the network and undirected and unweighted edges are created between all possible

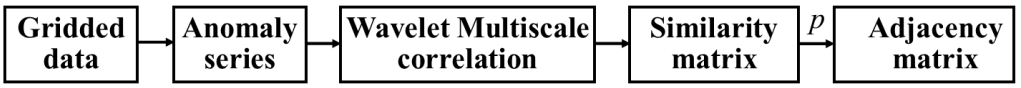



**Figure 4.1.:** Schematic of network construction. Each grid stations of SST dataset is considered as node and similarity between each pair of nodes is calculated using WMC measure. Finally, by applying the $95th$ percentile threshold along with multiple testing, a link between each pair of nodes is setup.

**Table 4.1.:** Overview of sea surface temperature network threshold values ($90th-, 95th-, 99th$ and FDR) at all timescales. In this study we have selected threshold values obtained using False discovery rates (FDRs) method.

| Percentile threshold | Original scale | 1-2 months | 2-4 months | 4-8 months | 8-16 months | 16-32 months | 32-64 months | 64-128 months |
|---|---|---|---|---|---|---|---|---|
| $90th$ | 0.32 | 0.12 | 0.16 | 0.24 | 0.33 | 0.48 | 0.55 | 0.67 |
| $95th$ | 0.41 | 0.17 | 0.22 | 0.32 | 0.44 | 0.61 | 0.66 | 0.76 |
| $99th$ | 0.72 | 0.26 | 0.32 | 0.43 | 0.65 | 0.86 | 0.80 | 0.84 |
| Multiple testing | Original scale | 1-2 months | 2-4 months | 4-8 months | 8-16 months | 16-32 months | 32-64 months | 64-128 months |
| FDR | 0.62 | 0.51 | 0.56 | 0.63 | 0.74 | 0.84 | 0.85 | 0.85 |

pair of nodes based on their similarity. We use WMC measure to quantify the similarity between anomalies SST timeseries, yielding a square similarity matrix of $9456 \times 9456$ (total number of grid cell considered). We generate an adjacency matrix from similarity matrix by applying significance-based pruning which is explained in the next section.

### 4.2.3. Multiple testing

In literature, a number of criteria have been proposed to generate an adjacency matrix from a similarity matrix, such as fixed amount of link density (Agarwal et al., 2018a; Donges et al., 2009a; Stolbova et al., 2014) or fixed thresholds (Jha et al., 2015; Sivakumar and Woldemeskel, 2014). Here, we consider a 5% link density combined with multiple testing. Multiple testing attempts to avoid false links by controlling the type I error or adjusting $p$-values to give only significant links (Benjamini and Yosef, 1995; Reiner et al., 2003). This is a very active area of statistics, and hence many different methods have been proposed. Although the varied approaches have the same goal, they go about it in fundamentally different ways. Here, we use false discovery rate (FDR) to control $p$–values (for more details refer Benjamini and Yosef (1995)). This may seem like a stringent requirement (5% link density combined with multiple testing), but a large number of edges satisfies this criterion and are therefore retained in the network. To give a clear understanding of the final selected threshold, we tabulate all the threshold values ($90th-, 95th-, 99th$ and FDR) in Table A.1.

## 4.3 Results and Discussion

We present the results in three parts, first discussing the scale-specific spatial patterns obtained over different timescales, second investigating the occurrence of short- and long-

range linkages between the identified scale-specific spatial patterns over different timescales, and finally providing a 3–D global visualization of the link distributions.

### 4.3.1. Spatial patterns over different timescales

Using the above-explained similarity measure and pruning method, we construct a separate SST network for each timescale (in total eight scales including the observed timescale, see Fig. 4.2). The networks reveal, separately for each timescale, how well the temporal evolution of the SSTs of each gridd cell is linked to that of the other grid cells, i.e., it shows the number of significantly correlated edges ($\sum A_{ij}$, also called links) that each grid cell (or node) has. This gives us a basis for understanding what kind of teleconnections act at a certain timescale.

Considering only the positively correlated links (Fig. 4.2), it is obvious that at all scales the majority of the grid cells show a low link number that is larger than zero. We identify this as the background link number at a certain scale that is explained by the near-field spatial correlations of the grid cells (Scarsoglio et al., 2013; Steinhaeuser et al., 2012). In addition, there are distinct regions with substantially higher number of links.

At the original, i.e., unprocessed scale (Fig. 4.2a) we observe a highly linked region in the mid to eastern equatorial Pacific accompanied by regions with a medium number of links in the southern Pacific, the western equatorial Pacific and the Indian Ocean. When looking at the scale-specific networks (Fig. 4.2b-h), it becomes clear that links at the observational scale are aggregated results of the coupling between scale-specific features of global SST data present at different timescales. Moving from lower to higher temporal scales, it becomes apparent that – with the annual cycle removed from the data – there are hardly any significant correlations between the SST evolution in different regions at timescales of less than a year (Fig. 4.2b-d).

However, at the 8–16 months scale (Fig. 4.2e), there are two zones with a relatively large number of significant connections in the mid to eastern equatorial Pacific and the southern Pacific. There is also a region with an above global average number of links in the Indian Ocean. While the number of links in this region increases when moving to the larger timescale of 16–32 months (Fig. 4.2f), the number of links from and to the Southern Ocean decreases.

At the scale of 32–64 months (Fig. 4.2g), the number of links in the mid to eastern Pacific Ocean decreases and a well linked region in the form of a tilted horseshoe emerges in the northern Pacific, as well as a region with a medium number of links in the subtropical North Atlantic. On even longer timescales (64–128 months) the highly linked region in the equatorial Pacific reemerges, as well as several smaller regions with a medium number of links in all three ocean basins (Fig. 4.2h). Fig. 4.3 shows the number of negatively correlated links. In contrast to the positive correlations (Fig. 4.2), there is no global background link number due to the fact that the near-field correlation is in general positive. On the original scale, regions whose SSTs evolution are negatively correlated to that of other regions can mainly be found in the Pacific Ocean (Fig. 4.3) centered about the equator as well as in the Southern Ocean. The separation in the different timescales again reveals that the observed patterns are aggregates of the timescale-specific feature. Again there are only a few highly linked regions on timescales of less than a year (Fig. 4.3b-d) with the exception of the subtropical South Atlantic (Fig. 4.3d). On longer timescales of about one to two years the link number in the South Atlantic stays rather stable and highly linked regions in the Pacific

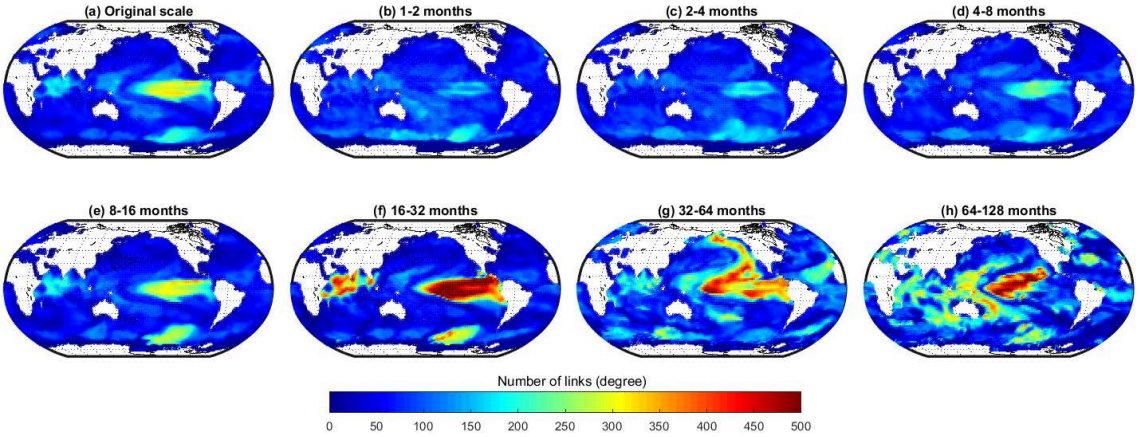



**Figure 4.2.:** Number of links of each grid cell in a global SST network constructed at different timescales by considering only positive correlations.

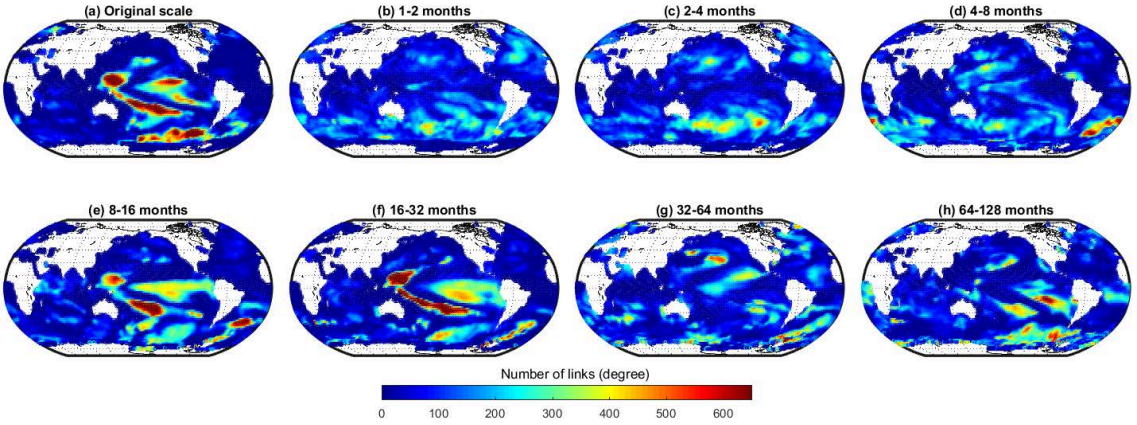



**Figure 4.3.:** Number of links of each grid cell in a global SST network constructed at different timescales by considering only negative correlations.

emerge (Fig. 4.3e-f), namely in the mid to eastern equatorial Pacific and the equatorial to south-western Pacific. The latter becomes more linked on the 16–32 months timescale, while the link numbers of the former decrease. Moving to interannual timescales the regions with a large number of significantly negatively correlated edges globally decrease but a few regions in the South Pacific remain (Fig. 4.3h).

### 4.3.2. Occurrence of short- vs long-range links

We further investigate the length distribution of the significantly correlated edges (both positive and negative) at the different scales by creating histograms showing the number of links plotted over the geographical length link (Fig. 4.4). The geographical link length (shortest path between two points on a sphere) was computed using Haversine's formula which determines the great circle distance between two points on the sphere (Van Brummelen, 2013).

For all scales (Fig. 4.4a-h) these frequency distributions of the link lengths show that the dominant factor controlling the significance of a correlation between two SST time series is the near-field spatial correlation. This property is embodied by Tobler's First Law of

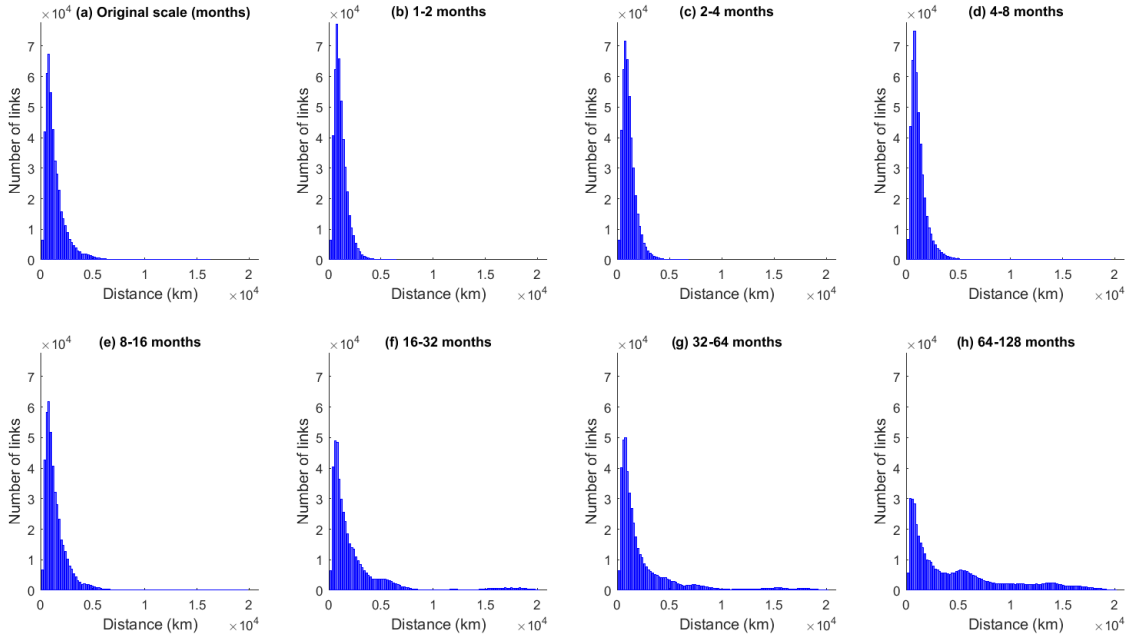



**Figure 4.4.:** Frequency distribution of the link lengths in SST networks at different timescales. Short-distance links can be interpreted as near-field correlations, while long-distance links suggest long-range spatial dependencies or teleconnections.

Geography, which states that 'everything is related to everything else, but near things are more related than distant things' Tobler, 1970. Nevertheless, the number of links decreases again at very short distances. There are two reasons for that: First, because the nodes are evenly spaced angularly on the sphere, neighboring grid points near the equator are much further apart ($> 500$km) than those near the poles. Second, the maximum number of possible links depends on the geographical link length due to the spherical shape of the Earth and the land ocean distribution on it; in general, the number of possible links increases with the link length until a maximum at a length of about $10,000$km and then decreases again. The second noticeable feature is a 'fat tail' for the longer timescales (16–32 months and longer, Fig. 4.4f-h) that suggests the presence of teleconnections, that is, a similar time-dependent behavior in locations that are geographically separated.

To better understand the nature of these teleconnections we divide the entire globe into three standard regions, namely the extratropical Northern Hemisphere (NH), Tropical Hemisphere (TH) and extratropical Southern Hemisphere (SH), and analyse separately the frequency distributions of the link lengths of edges that connect nodes only within in these regions for timescales larger than 16–32 months (Fig. 4.5). It can be seen that short-range links appear in all regions on all the considered timescales (i.e. larger than 16 months) with the smallest number of short-range links occurring at the very long timescales (64-128 months) for all three hemispheres. When looking at the long-range links, the three hemispheres show a different behaviour.

In the tropics, teleconnections appear on timescales from 16–32 months and on the longer timescales of more than 64 months. At these timescales, there are also long-range links in the two extratropical hemispheres that start developing at the 32–64 months timescale at which the tropics show no long-range links and additionally a notable reduction in

short-range links. The geographical lengths of the teleconnections in the tropics can be longer than those in the extratropics due to Earth's spherical shape as grid points in the tropics have larger distances. Hence, we conclude that the length distribution of links varies between the different decadal timescales for the global and the within-region analysis. As most of the teleconnections that we are interested in act across the three zones, we

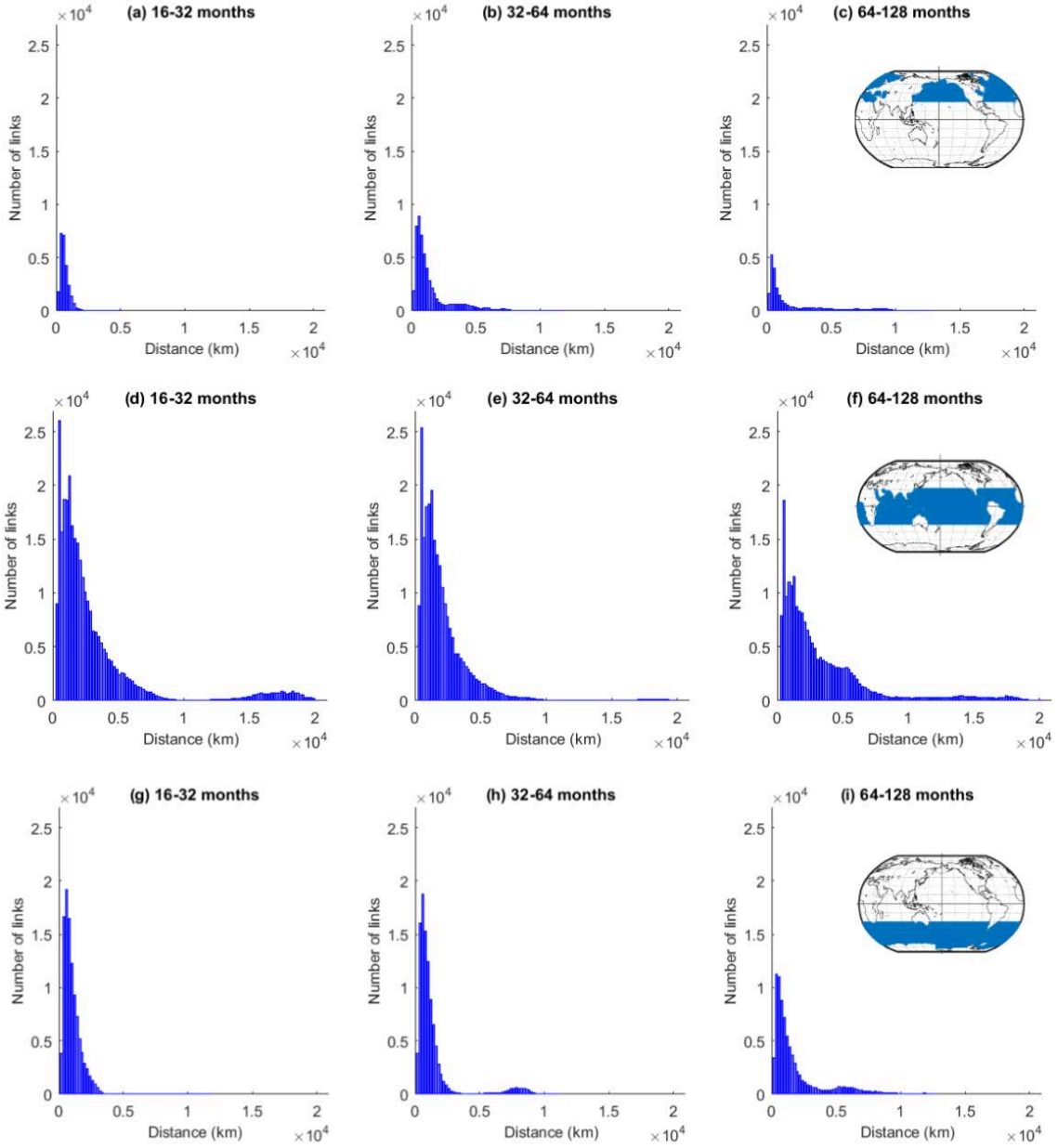



**Figure 4.5.:** Frequency distribution of the edge lengths in SST networks at different timescales within three standard zones, namely NH, TH and SH (within-region links). Short-distance links show near-field spatial correlation, while long-distance links suggest long-range spatial dependencies or teleconnections within the same zone.

further analyse the link length frequency distribution from one region (NH, TH,and SH) to the rest of the world (including within-region links, Fig. 4.6). This will help us to understand the interactions between tropics and extratropics which cannot be revealed by Fig. 4.5.

Fig. 4.6 confirms that the extratropics (NH & SH) do not possess teleconnections at the interannual scale (16–32 months) both within the same region (shown in Fig. 4.5) and with the rest of the world (Fig. 4.6). We also notice that the absolute number of short-range connections (Fig. 4.6) at this scale remains almost unchanged in all three regions except for slight increments in the extratropical SH which depict the near-field interactions between SH and TH. These links connect regions with a distance of about 5,000km, which is visible in Figs. 4.2 and 4.3.

The first interesting observation at the intradecadal scale (32–64 months) is the significant increase in short-range connections in the extratropics (NH and SH) in contrast to almost no significant changes in the tropics compared to the interannual scale (Fig. 4.6). However, Fig. 4.5e already confirmed that the tropics do not have much within-region short-range connections at the 32–64 months scale compared to the 16-32 month scale (Fig. 4.5d). Also the increase in short-range connections in NH is mainly credited to internal short-range connections (Fig. 4.5b). These observations suggest that SH exhibits a significant number of short-range links with the tropics at the intradecadal scale. Moreover, we observe that long-range connections appear in the extratropics (NH & SH) which are completely absent at the interannual and other finer scales (Fig. 4.6a, c) or when looking only at the within-region links (Fig. 4.5b,h). This observation indicates that interactions in the climate system occur at specific timescales and are absent at other timescales, which undermines the necessity of multi-scale analyses.

At the decadal scale (64–128 months) and for all three zones, we observe a significant reduction in beyond-region, short-range connections compared to the intradecadal scale. For NH it is the same reduction as is seen for within-region links, whereas the reduction is larger for SH and TH. This observation indicates the decline in near-field connections within NH and in-between TH and SH. The rise in NH and SH at distances of about 10,000 to 15,000km shows the significant coupling of the two hemispheres to the other regions. The fact that for both extratropical hemispheres the number of very long links increases, while it stays rather constant for TH, indicates the existence of very long-range links between SH and NH.

These analyses show that (i) at the interannual scale the tropics show the largest number of significant correlations acting over both short and long distances and the existence of links with a length of about 5,000km between SH and TH (Fig. 4.6d, g); (ii) at the intradecadal scale the main characteristics are long-range links between NH and TH with a length of about 5,000km and 7,000km (Fig. 4.6b, e), as well a small number of links between SH and TH with a length of about 15,000km, yet there are no long-range links within TH; (iii) at the intradecadal scale long-range connections between NH and SH to the tropics as well as interhemispheric (SH to NH) links exist.

### 4.3.3. 3-D visualization of the link distributions

The results listed above identify the ocean regions that are most closely linked when looking at the variability of SST at different timescales going from the intra- to interannual and even decadal scale (Fig. 4.2 and 4.3). The analysis of the frequency distributions of the link lengths gives further insight into the occurrence of teleconnections (Figs. 4.4–4.6). What is missing is detailed knowledge about the actual regions that are linked at the different timescales.

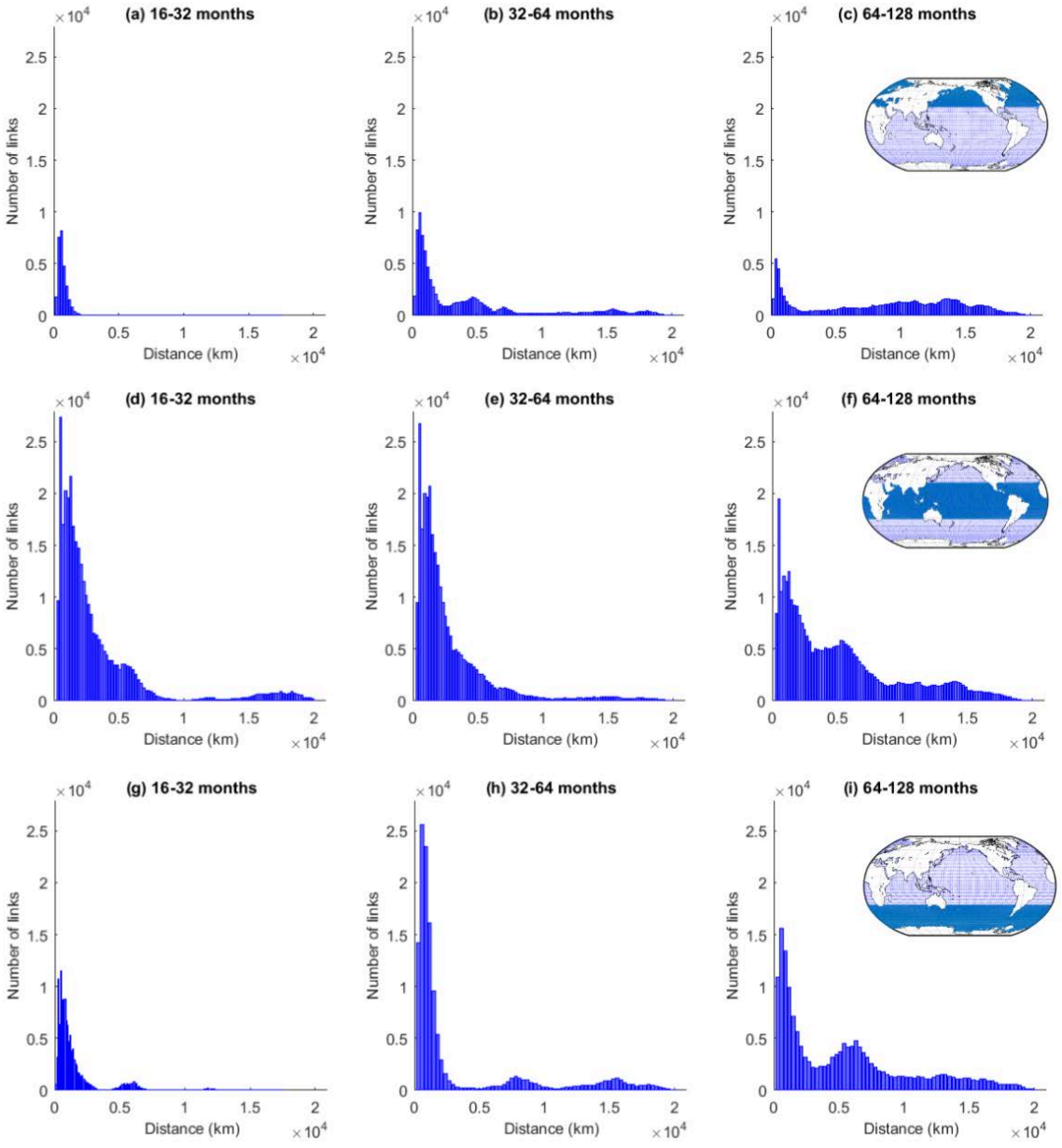



**Figure 4.6.:** Frequency distribution of the edge lengths in SST networks at different timescales constructed for standard zones, namely NH, TH, and SH, considering the interaction with the whole globe (beyond-region links). Short-distance links show near-field correlation, while long-distance links suggest long-range spatial dependencies or teleconnections.

We, therefore, visualize the climate networks on spherical 3-D node-link maps (Fig. 4.7) constructed at different timescales, where the node positions are fixed, and links are plotted using the software GTX tool (Buschmann et al., in prep.). To best visualize the structures of the global climate network, especially the connectivity of different regions either via short- or long-range links , we do not plot all links but focus on the regions that were identified as being highly linked. The 3-D projection of the networks reveals patterns consistent with the 2-D visualization (Fig. 4.2 and 4.3). However, it contains additional information about short- and long-range connections, for example even though the 2-D visualization

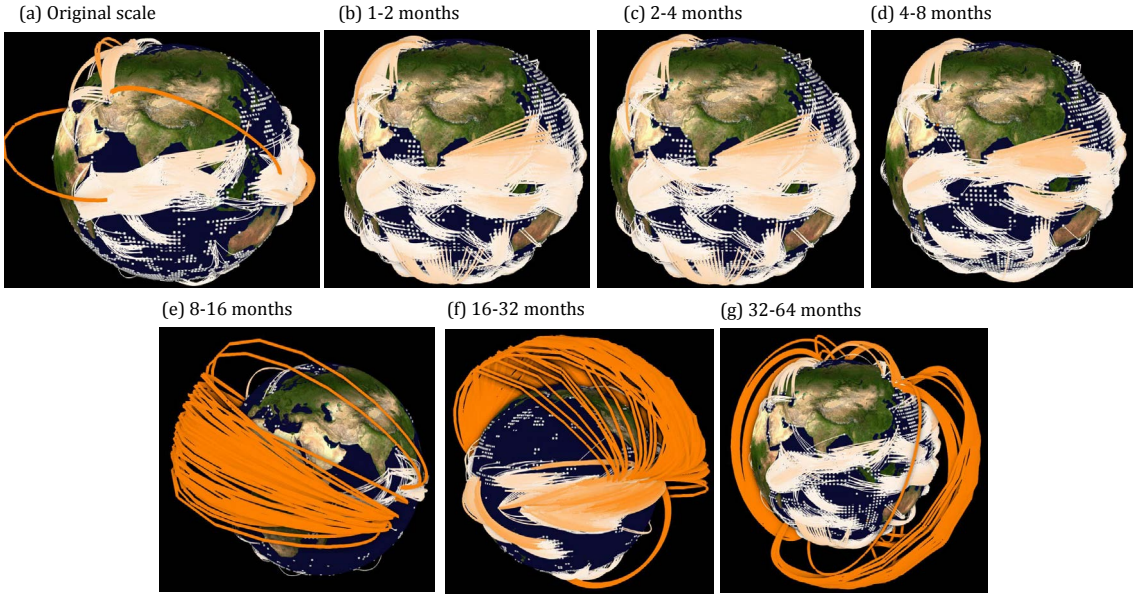



**Figure 4.7.:** Spherical 3-D globe representation of the short- and long-range connections in SST networks at different timescales. Only nodes beyond a pre-selected betweenness centrality (BC) value are plotted. For instance, in subplot (a-d),(e-f) and (g) nodes are plotted for BC values greater than $90K$, $57K$ and $38K$, respectively. Edge color represents the geographical link length. The definition and interpretation of BC is presented in the Appendix B.

at the original scale (Fig. 4.2a and 4.3a) depicts approximately the same patterns as the 16–32 months scale (Fig. 4.2e and 4.3e), the 3-D visualizing clarifies that there are hardly any long-range connections present at the original scale (Fig. 4.7a) compared to the 16–32 month scale (Fig. 4.7e). This visualization of the connections helps in attributing the interrelationships between selected regions at certain timescales . Thus, the node-link maps add valuable information to the discussion and interpretation of the results obtained in Sect. 4.3.1 and 4.3.2.

## 4.4 Discussion

In the following we link the results from our analysis to the known modes of climate variability and teleconnections and highlight the new findings of this study.

The link maps (Fig. 4.2 and 4.3) clearly reveal the known modes of SST variability on the various time-scales. The highly linked region in the eastern equatorial Pacific can be identified as El Niño/Southern Oscillation (ENSO), a periodic fluctuation in SST (accompanied by a fluctuation in air pressure) across the equatorial Pacific, that occurs with an average time between peak events of about four years (Chen et al., 2004) and can therefore be seen best at the 16–32 month scale (Fig. 4.3f). At the same timescale, the region of variability of the Indian Ocean Dipole (IOD) shows a large link number, IOD event act on about the same timescale like ENSO (Feng and Meyers, 2003). Additionally ENSO and IOD are known to impact each other via the atmosphere (Luo et al., 2010; Stuecker et al., 2017; Yamagata et al., 2013) and the node-link map reveals that the links between the two are mainly between the northern part of the ENSO tongue and the equatorial Indian Ocean (Fig. 4.7e and f). Furthermore, the diagram reveals that the arrow shaped pattern

of negatively correlated links about the western equatorial Pacific (Fig. 4.4f) that is known to occur at the same time as El Niño is neither linked to ENSO nor IOD.

The last prominent feature at the intraannual timescale is the highly linked area in the Southern Ocean (e.g., Fig. 4.2f and 4.3f). Our 2–D visualization confirms here what has just been analysed by Ferster et al.Ferster et al., 2018: a link between SST in the Southern Ocean to ENSO events via the Southern Annular Mode (SAM), i.e. the north–south movement of the westerly wind belt that circles Antarctica. Positive states of SAM correlate with a cooling of SSTs in the high-latitude Southern Ocean, and a warming within the Southern Hemisphere sub-tropics and mid-latitudes (Ferster et al., 2018); this negative correlation can be seen in in our 2-D visualization (Fig. 4.3f). Our 3-D visualization further reveals a teleconnection between the very eastern part of ENSO and the region in the Southern Ocean near Western Antarctica (Fig. 4.7). Viewing the 2–D maps we can say that this teleconnection has to be positively correlated (Fig. 4.2f).

Going to longer timescales (32–64 months) we find the majority of the highly linked regions in the northern Pacific. Form and location of these regions resemble strongly the pattern of the Pacific Decadal Oscillation (PDO), a pattern in SST and sea-level pressure that varies between positive and negative phases on periods of 20–30 years. Yet we find high link numbers at the intradecadal scale. A detailed look shows that this is not a contradiction to the known facts: Viewing the 3-D visualization (Fig. 4.7g) we find that most of the links are not within the PDO pattern (these indeed should act on a 20–30 year scale), but are teleconnections to the Southern Ocean. These can act on shorter timescales, likely via a western Pacific ocean-atmosphere pathway, whereby SST anomalies can propagate from the southern to the northern hemisphere (Mamalakis et al., 2018).

But not all teleconnections work via the atmosphere. The 3–D visualization (Fig. 4.7f) shows a number of links going from the North Atlantic to the South Atlantic (this can also be seen in Fig. 4.3g). This negative correlation likely shows the so-called see saw response (Rahmstorf, 2002) that is due to the transport of heat from the Southern Ocean to the North Atlantic via the Atlantic Meridional Overturning Circulation (AMOC). If the AMOC is stronger (as it has been in the period after year 2000 compared to the years before (Caesar et al., 2018)) more heat is transported towards the North which leads to a cooling in the Southern Ocean and a warming in the subpolar North Atlantic.

More general information on where and at what distances the teleconnections act is given by the histograms of the frequency distributions of the link lengths. Within one region we find teleconnections mainly in the tropics (Fig. 4.5d) acting on a length scale of about $18,000$km. There are also a few long-range links in both the extratropics NH and SH, but here on all timescales the short-range links dominate. This is different when looking at the links from one region to the others (Fig. 4.6). Here a number of long-range links are detected, especially on the $64-128$ month scale on spatial scales of about $6,000$–$7,000$km as well as about $10,000-15,000$km. This is likely due to the fact that teleconnections often work via Rossby waves that link the tropics to the extratropics. The wavelengths of these Rossby waves depend on the strength of the westerly winds, but are typical about $7,000$km for weak westerly flow in the mid-latitudes and increase for stronger winds and more polar regions (Hoskins and Karoly, 1981).

In general, we can conclude that our method enables us to distinguish between the regions

that are highly linked through significantly correlated short-range edges and those that have long-range teleconnections. While the former are typical for proximity-based correlation on very short timescales as well as for correlations within the region where a climate mode takes place (like El Niño or the Indian Ocean dipole) and are mainly due to the advection of heat by the oceanic surface, the long-range links correspond to teleconnection patterns that either act via the atmosphere, e.g., through Rossby waves, or through the major open circulation systems like the AMOC. The latter warrant even more investigation as they are of particular importance for the understanding and forecasting of lasting weather events like droughts, flooding and temperature extremes.

## 4.5 Conclusions

Ocean-atmospheric teleconnections are intriguing phenomena in the climate system as they connect very remote regions. It is thus clear that the throughout diagnosis of their characteristics will improve our description and understanding of the climate system. Here, we propose a unified framework to investigate and unravel the short-range and long-range teleconnection at different timescales and spatial distances. The study helps to identify the highly linked regions in the ocean at different scales along with the identification of the link type, i.e. short-range link or teleconnection. We can therefore identify the climate modes that are highly linked within themselves and distinguish them from the regions that have long-range teleconnections to other ocean regions. This adds valuable information to the understanding of the workings of these teleconnections.

## 4.6 Acknowledgments

This research was funded by Deutsche Forschungsgemeinschaft (DFG) (GRK 2043/1) within the graduate research training group Natural risk in a changing world (NatRiskChange) at the University of Potsdam http://www.uni-potsdam.de/natriskchange.

## 4.7 Data availability

The data is open access and freely available at https://www.esrl.noaa.gov/psd/.

# Chapter 5

# Unraveling the spatial diversity of Indian precipitation teleconnections

*Agarwal, A., Marwan, N., Maheswaran, R., Krishnan,R., Kurths, J., and Merz, B.: Unravelling the spatial diversity of Indian rainfall teleconnections. (Scientific Reports-In review).*

A better understanding of precipitation dynamics in the Indian subcontinent is required since India's society depends heavily on reliable monsoon forecasts. We introduce a nonlinear, multi-scale approach, based on wavelets and event synchronization, for unraveling teleconnection influences on precipitation. We consider those climate patterns with highest relevance for Indian precipitation. Our results suggest significant nonlinear influences which are not well captured by the wavelet coherence analysis, the state-of-the-art method in understanding linkages at multiple time scales. We find substantial variation across India and across time scales. For instance, El Niño/Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD) mainly influence precipitation in the southeast at interannual and decadal scales, respectively, whereas the North Atlantic Oscillation (NAO) has a strong connection to precipitation particularly in the northern regions. Our results provide an exciting perspective for capturing the dynamics of precipitation and improving precipitation forecasting.

## 5.1 Introduction

Understanding the spatial patterns, frequency and intensity of precipitation in the Indian subcontinent is an active area of research due to its essential impact on life and property (Pai et al., 2015). The Indian monsoon is the pulse and lifeline of over one billion people, and the socio-economic development in this part of the world heavily depends on reliable predictions of the monsoon (Goswami and Krishnan, 2013). A report by Paliwal (2017) argues that 'India struggles to predict weather over its lands,' and that India 'has a lot to lose if it does not make better sense of weather,' especially for its farmers and disaster management (Shukla et al., 2017). In recent years, India has witnessed country-wide crop yield losses due to unreliable seasonal precipitation forecasts. Farmers claimed that they would not have suffered these losses if the forecast had been accurate and communicated $10-20$ days in advance (Paliwal, 2017).

Numerous studies have emphasized the importance of understanding the influence of large-scale climatic patterns on precipitation for improving forecast accuracy (Feng et al.,

2016). A significant number of studies have analyzed the relationship between precipitation and climatic patterns for India. This research has shown that relevant patterns are the El Niño/Southern Oscillation (ENSO, Kumar et al., 2006; Mokhov et al., 2012), the Indian Ocean Dipole (IOD, Behera et al., 1999; Krishnan and Swapna, 2009), the North Atlantic Oscillation (NAO, Bharath and Srinivas, 2015; Feliks et al., 2013), the Pacific Decadal Oscillation (PDO, Dong, 2016; Krishnan and Sugi, 2003), and the Atlantic Multidecadal Oscillation (AMO, Goswami et al., 2006; Krishnamurthy and Krishnamurthy, 2016). An in-depth understanding of the associations and their spatiotemporal diversity, between these climatic patterns and precipitation is crucial for better short-term and long-term precipitation forecasts. A case in point is the substantial increase in the monsoon forecast accuracy for India by tracking and including the IOD as one of the predictors in the dynamical monsoon forecasting model (Ashok et al., 2001). Another example is the 'National Monsoon Mission' (NMM) which the Government of India has launched with a vision to develop a state-of-the-art dynamical prediction system for monsoon precipitation on different time scales. According to the Indian Institute of Tropical Meteorology (IITM), India's skill of seasonal and long-range forecasts has become more reliable after launching NMM (Rao, 2015). However, a recent study Pillai et al. (2018) concludes that the present-day seasonal prediction systems need to be improved in their representation of the tropical SST–monsoon teleconnections, which should further increase the seasonal prediction skill of the Indian summer monsoon. Hence, reliable seasonal precipitation forecasting for India remains a grand challenge, notwithstanding the notable progress made in the prediction of tropical climate during the past decades.

Over the years, linkages between climatic patterns and precipitation have been investigated by a range of statistical methods, such as correlation (Abid et al., 2018), principal component analysis (Luterbacher et al., 2006), empirical orthogonal functions (Hannachi et al., 2007), regression and canonical analysis (Xoplaki et al., 2004; Shukla et al., 2018; R. Shukla et al., 2018). Recently, wavelet coherence has become the state-of-the-art method for studying the influence of climatic patterns on precipitation at different temporal scales. For example, Tan et al. (2016) analyzed the relations between Canadian precipitation and different global climate indices. Similar studies reported in other parts of the world (Araghi et al., 2017; Hu and Si, 2016; Tan et al., 2016). In climatic systems energy is stored and transported differently on different temporal scales, resulting from interactions of intertwined sub-components across a wide range of scales (Miralles et al., 2014; Peters et al., 2007). Multi-scale interactions have therefore received extensive attention in the field of climate dynamics (Paluš, 2018; Peters et al., 2007; Steinhaeuser et al., 2012; Casagrande et al., 2015) and have been proposed as a mechanism for triggering extreme events (Okin et al., 2009; Paluš, 2014a; Peters et al., 2004) and abrupt transitions (Peters et al., 2007). Therefore, the multi-scale analysis holds the promise of better understanding the system dynamics compared to analyzing processes at one time scale only.

Recently, event synchronization (ES) has emerged as a powerful nonlinear tool for disentangling the temporal organization of extreme precipitation events or anticipating the monsoon onset (Boers et al., 2014; Malik et al., 2012; Stolbova et al., 2016) or droughts (Konapala and Mishra, 2017). Boers et al. (2015) analyzed the spatial patterns of extreme precipitation events of the South American monsoon and their relation to ENSO and large-scale atmospheric moisture divergence. Conticello et al. (2018) used ES to investigate the influence of the climate state, represented by the atmospheric vapor transport and geopotential height patterns, on precipitation extremes in the Lazio region, Italy. Very

recently, Sun et al. (2018) demonstrated the potential of the ES approach to understanding the spatiotemporal covariation patterns of precipitation and soil moisture. Rooted in neurosciences, ES has the advantages that is quantifies nonlinear connections and that it automatically classifies pairs of events arising at two locations as temporally close (and, thus, possibly statistically – or even dynamically – interrelated) without the necessity of selecting an additional parameter in terms of a fixed tolerable delay between these events (Conticello et al., 2018; Ozturk et al., 2018a).

Based on the encouraging results reported by the above studies, we use ES to quantify the (possibly nonlinear) linkages between large-scale climatic patterns and precipitation across India. More specifically, we analyze the linkages between the 95-percentile extreme events, extracted from gridded Indian precipitation data at monthly resolution, and the climate patterns ENSO, IOD, NAO, PDO, and AMO which have been shown to be of significant relevance for precipitation in India. We combine ES with the wavelet transform, as proposed recently by Agarwal et al. (2017). This combination, termed MSES (Multi-scale Event Synchronization), allows studying nonlinear connections between times series at different temporal scales. To consider the spatial variation across India, we sub-divide India into homogeneous regions that share similar precipitation characteristics and identify a representative grid cell for each region. The homogenous regions and the representative grid cells are obtained using the concept of complex networks following Agarwal et al. (2018a).

The novelty of this paper is the integration of (1) a nonlinear method, i.e., ES, for quantifying the linkages between large-scale climate patterns and precipitation at (2) multiple time scales, considering (3) the spatial variation of these linkages. To our knowledge, this combination (nonlinear – multiple time scales – spatial variation) has never been implemented, neither for India nor for any other region. We argue that it allows unraveling the spatio-temporal diversity of Indian precipitation teleconnections, offering a compelling perspective for capturing the dynamics of precipitation and improving precipitation forecasting.

## 5.2 Study area and data

### 5.2.1. Study area

Our study area is the Indian subcontinent which shows a significant variation in climate characteristics. India extends over an area of $3,287,263 km^2$. Its climate regimes are classified as arid (northwestern India), semi-arid (northern lowlands and central peninsular India), humid (coastal lowlands, southwestern and northeastern highlands) and alpine (Himalayan mountains in the north). The spatio-temporal variation of precipitation, as well as temperature, is significant over the country (Bharath and Srinivas, 2015). The entire country receives 80% of its total precipitation during the southwest monsoon, from June to September (Bharath and Srinivas, 2015). During the northeast monsoon (October to December), the precipitation is considerable but confined to the southeastern part of the country. The temperature is low from November to February, with lower values recorded in northern India and higher values in the southern part. Temperature rises from March to May, with highest temperatures recorded in the central and northwestern part. With the onset of the southwest monsoon, the temperature drops and becomes fairly uniform over the major part of the country (Bharath and Srinivas, 2015; Stolbova et al., 2014).

### 5.2.2. Gridded precipitation data

We use the high-resolution ($0.25° \times 0.25°$) daily gridded precipitation data set for the period 1901–2013, developed by the Indian Meteorological Department (IMD) for the spatial domain of 66.5°E to 100°E and 6.5°N to 38.5°N, covering the mainland region of India (Pai et al., 2014). The gridded data has been generated from the observations of 6995 gauging stations across India (Pai et al., 2014). The dataset captures well the spatial distribution of precipitation over the country. For our study, out of 17415 grid cells, 4631 cells lying inside the boundaries of India were identified. This dataset contains continuous daily precipitation values for 63 years (Jan 1951 to Dec 2013) without any missing values.

### 5.2.3. Time series of global and regional climate indices

For understanding the linkages between climate patterns and precipitation, we use time series of global and regional climate indices for the same period, i.e., 1951–2013. We have selected those indices for which earlier studies have shown a relation to Indian precipitation. The selected climate indices and the respective studies are: ENSO (Mokhov et al., 2012), IOD (Ashok et al., 2001), NAO (Kakade and Dugam, 2000; Feliks et al., 2013), PDO (Dong, 2016; Krishnan and Sugi, 2003) and, AMO (Goswami et al., 2006; Krishnamurthy and Krishnamurthy, 2016). For detailed information on these climate indices and the data sources, we refer to https://www.esrl.noaa.gov/.

## 5.3 Methodology

To investigate the nonlinear, multi-scale linkages between climate patterns and precipitation we propose the methodology shown in Fig.5.1. First, we construct a precipitation network of 4631 grid cells of the precipitation dataset using event synchronization. We further pool grid cells with similar precipitation characteristics into homogenous regions and identify a representative grid cell for each region as proposed by Agarwal et al. (2018a). The linkages between the precipitation time series of the representative cells and the teleconnection indices are analyzed by the Multi-scale Event Synchronization (MSES) method developed by Agarwal et al. (2017). Finally, the proposed methodology is compared to the state-of-the-art method in multi-scale time series analysis, the wavelet coherence analysis (WC).

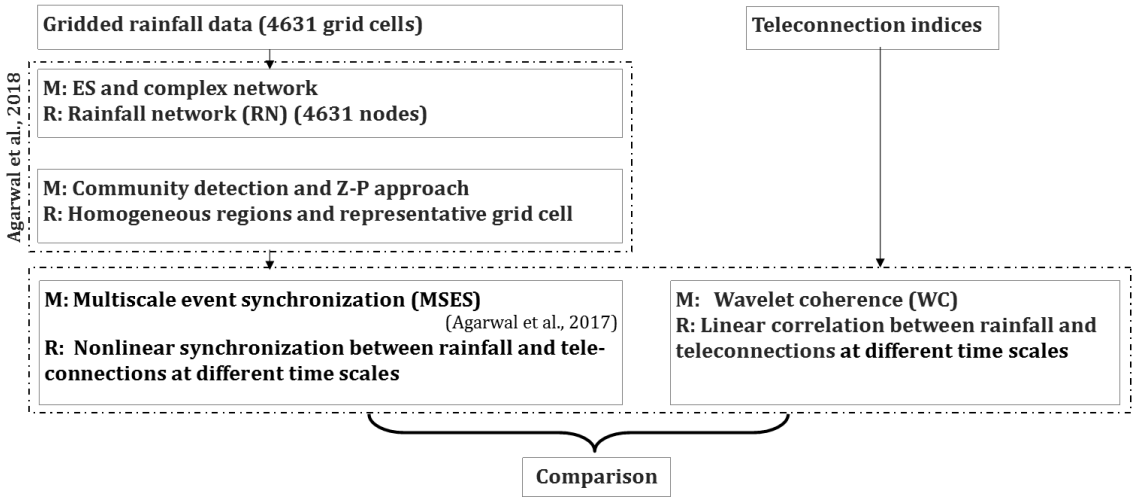



**Figure 5.1.:** Schematic of the methodology to investigate the linkages between climate patterns and precipitation (M stands for method and R for result). Methods such as community detection and Z-P approach has been discussed in Appendix A.

### 5.3.1. Event synchronization and network construction

Event synchronization (ES) measures the nonlinear synchronization for point processes (Quiroga et al., 2002a). Each grid cell of the precipitation data set serves as a network node, while the daily precipitation estimates at each cell provide the time series for that node. Following Agarwal et al. (2018a), we define heavy precipitation events at each node as those days with precipitation larger the $95th$ percentile at that grid cell. Then ES is used to calculate the strength of synchronization ($Q$) between all possible pairs of grid cells. A link between two grid cells is set up if their heavy precipitation occurrence is strongly synchronized, which we define as having a $Q$ value greater than the $95th$ percentile (for more details see Agarwal et al. (2018a)). We repeat the procedure for all possible pair of nodes to construct a precipitation network which is mathematically represented as (Agarwal et al., 2018a)

$$A_{i,j} = \begin{cases} 1, & \text{if } Q_{i,j} \geq \theta^{Q}_{i,j}. \\ 0, & \text{otherwise.} \end{cases} \tag{5.1}$$

Here, $\theta^{Q}_{i,j} = 95th$ percentile is a chosen threshold, and $A_{i,j} = 1$ denotes a link between the $i^{th}$ and $j^{th}$ nodes and 0 denotes otherwise.

### 5.3.2. Community detection and Z-P approach

The linkages between climate indices and precipitation are evaluated on a regional scale. The study area is subdivided into homogeneous regions with similar characteristics of heavy precipitation events using the concept of complex networks, as proposed by Agarwal et al. (2018a). Several studies have reported superior performance of complex networks in identifying homogeneous regions compared to more traditional methods, such as the hierarchical clustering algorithm or the information-theoretic algorithm (Harenberg et al., 2014).

Further, for each community, we identify a representative grid cell using Z-P space approach as proposed by Agarwal et al. (2018a). The cell with the highest number of intracommunity

links is considered as representative (Halverson and Fleming, 2015), based on the argument that this cell shows the strongest synchronization within the community. We expect its climatological properties, such as the linkage to large-scale climate patterns, to have the highest similarity to the properties of the other cells in the community. We could also use a composite, e.g., by normalizing the grid cell time series and defining the time series of the mean of the normalized series as representative. However, this definition would reduce the variability and could mask existing connections to climatic patterns.

### 5.3.3. Multi-scale Event Synchronization

In this study, we use the multi-scale event synchronization (MSES) measure (Agarwal et al., 2017) to quantify the relationship between precipitation and climate indices. MSES combines the wavelet transform and event synchronization. The MSES values between precipitation and climate indices are estimated in the following manner:

(a) The climate indices and precipitation values at monthly resolution are decomposed using the maximum overlap discrete wavelet transformation (MODWT; supplementary information E) to obtain the scalewise detail components. These components represent the features of the signal at the different time scales.

(b) After fixing a 95% threshold for each of the components of precipitation and climate indices, the event synchronization values are estimated. The 95% threshold values are estimated for each scale component separately, ensuring a reliable estimation of the synchronization between the events. Since the time series of the climate indices represent anomalies, both strongly positive and strongly negative values are considered in the derivation of event time series.

(c) The estimated ES values are considered significant if they are higher than the ones obtained from a significance test (Agarwal et al., 2017).

(d) These steps (a-c) are repeated for all combinations of climate indices and precipitation for the different regions.

### 5.3.4. Wavelet coherence

We benchmark the MSES results against the wavelet coherence (WC) analysis, because wavelet coherence is the state-of-the-art method in evaluating linkages between hydroclimatological variables at multiple time scales (Peters et al., 2004; Tan et al., 2016). We use the Grinsted Toolbox (Grinsted et al., 2004) for calculating the WC between precipitation of the representative grid cells and the climatic indices (supplementary information E). It is important to note that WC uses the complete, continuous time series for quantifying the linkages between precipitation and climate patterns, whereas MSES first derives events at the different time scales, and then uses the synchronization between these events to identify the linkages.

## 5.4 Results and Discussion

### 5.4.1. Homogeneous regions and representative grid cells

To reduce the number of pairs of precipitation and climate index time series for finding mutual associations, we pool precipitation grid cells with similar heavy precipitation

event characteristics into homogeneous regions. These regions and their main physical characteristics are given in Fig.5.2. A more detailed discussion of these regions is provided in Agarwal et al. (2018a). For each community, we identify a representative grid cell (black dots in Fig.5.2) using the Z-P space approach. Next, we investigate the nonlinear linkages between the precipitation time series of the representative cells and the climate indices.

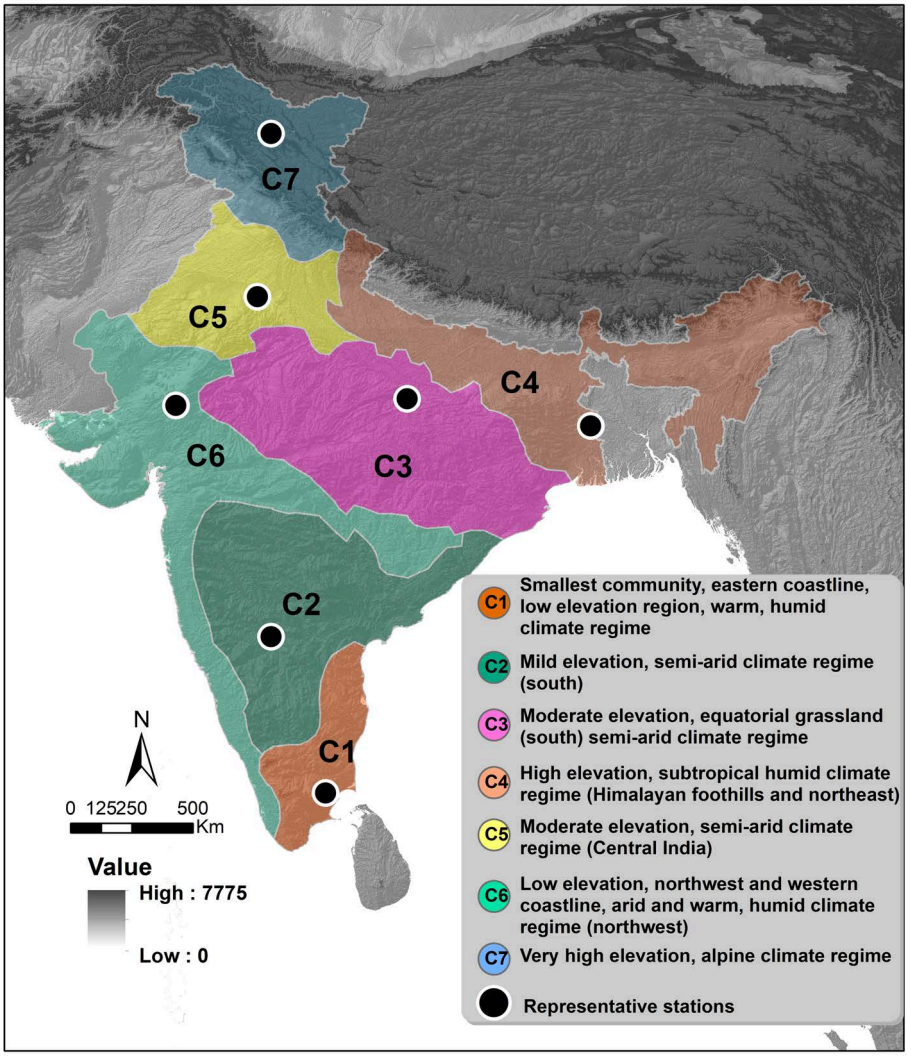



**Figure 5.2.:** Spatial distribution/ extent of the seven regions, or communities, with similar heavy precipitation event characteristics across India. Black dots indicate representative grid cells for each of the community identified using the Z-P space. Terrain characteristics of the Indian subcontinent is shown using the SRTM DEM (in background).

### 5.4.2. Linkages between precipitation and climatic patterns at multiple time scales

Figs.5.3(a-e) and 5.4(a-e) show the nonlinear synchronization, in terms of MSES values, and the linear coherence, in terms of WC values, between precipitation and the climatic patterns, respectively. MSES and WC are given for the five most relevant climatic patterns for the Indian subcontinent and precipitation in each of the representative grid cells of the seven homogeneous regions. We limit the analysis to scale 7, i.e., 16 years, due to

the distortion created by the boundary effects of the wavelet decomposition (Percival and Walden, 2000).

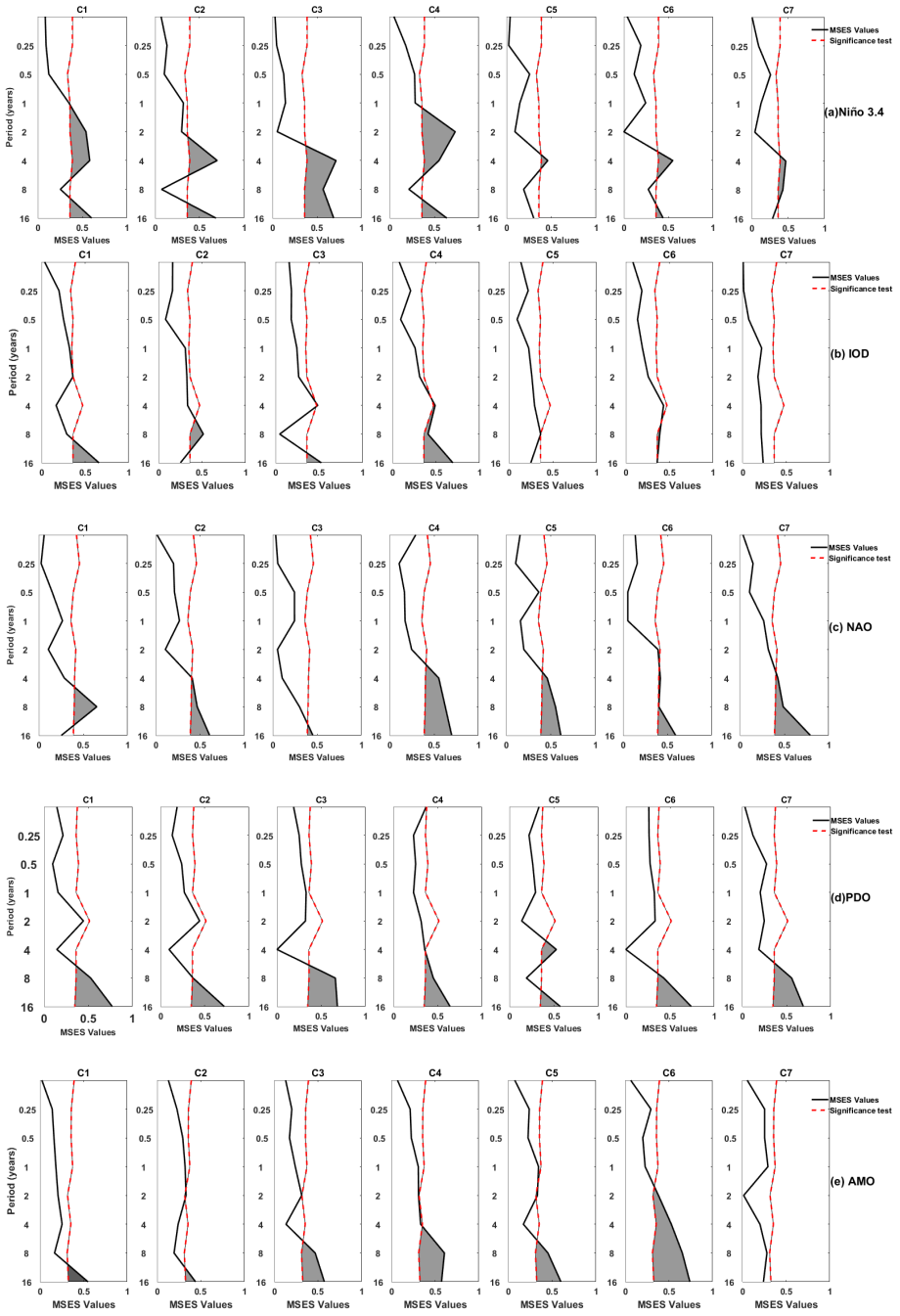



**Figure 5.3.:** Multi-scale event synchronization (MSES) between precipitation and climate indices. From top to bottom: Nino3.4, IOD, NAO, PDO, and AMO. From left to right: community 1 to community 7. MSES values are shown as solid lines, and significant connections (at the 95% significance level) are marked in grey.

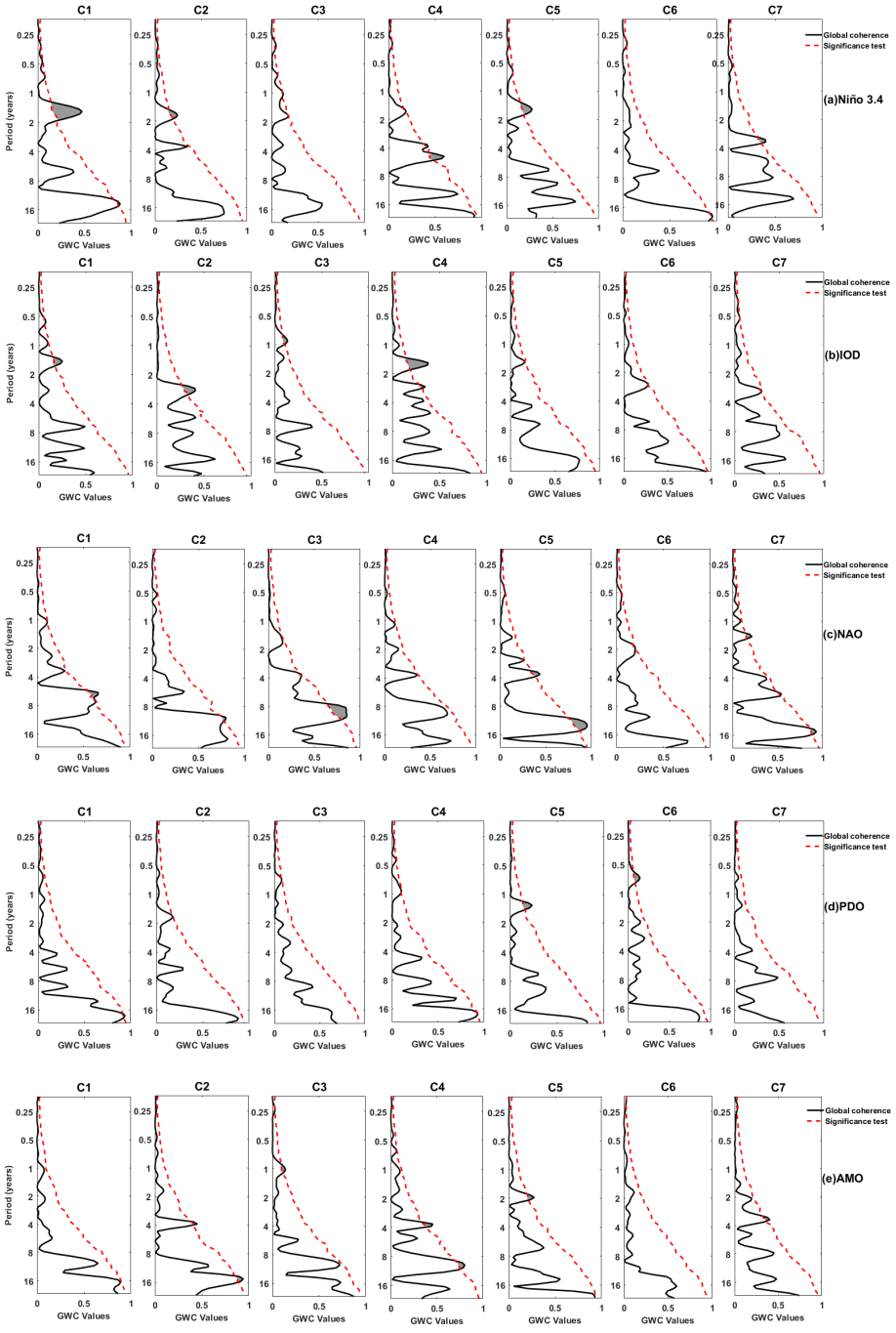



**Figure 5.4.:** Global Wavelet Coherence (GWC) between precipitation and climate indices. [Top to bottom]: Nino3.4, IOD, NAO, PDO, and AMO. From left to right: Community 1 to community 7. WC values are shown as solid lines, and significant connections (at the 95% significance level) are marked in grey.

Fig.5.3a shows a significant association between El Niño/Southern Oscillation (ENSO) and precipitation in all regions of India at the interannual scale. Its strength varies in space and with temporal scale. It is stronger for the southeastern peninsular ($C1$,$C2$,$C3$ and $C4$) and decreases notably in the northwestern Himalayan ($C5$,$C6$ and $C7$) regions. In the southeastern peninsula, the highest synchronization for the low ($C1$), mild ($C2$) and moderate ($C3$) elevation regions occurs at the 4-year scale, and at the 2-year scale for the

high elevation ($C4$) region. For the southeast regions of India, we observe a significant synchronization at the decadal scale ($8-16$ years) which is counterintuitive given the interannual time scale of ENSO (D'Arrigo, 2005; McGregor et al., 2013). The analysis based on WC (Fig.5.4a) shows substantially less correlation between precipitation and ENSO in all regions.
Overall, the association between ENSO and precipitation at the interannual scale is coherent with the general understanding that extreme precipitation and in India are associated with ENSO (Rajeevan and Pai, 2007). However, the variation of this linkage across India has not been reported. We find stronger linkages for the regions close to the ocean (southeastern peninsular comprising $C1$ to $C4$) compared to the inland regions with higher elevation (northwestern India comprising $C5$, $C7$). The results mentioned above are in congruence with the findings by Guhathakurta et al. (2017) and Mishra et al. (2012). The spatial heterogeneity in the strength of the relationship between ENSO and precipitation may be a result of the tropical convection during the ESNO events (Bansod, 2011). Other studies have confirmed that there is a decrease in the strength of the relationship between precipitation and ENSO events with distance from the ocean. A similar pattern observed in Mexico where the Nino3.4 teleconnection is weaker, if not opposite in sign, in northern versus southern Mexico (Hu and Feng, 2002). This observation leads us to the understanding that the ENSO teleconnection is strong in regions of climatologically strong convection.
Interestingly, an association between ENSO and precipitation at the decadal scale has not been reported for India so far. This association might be a consequence of the interdependencies between ENSO and IOD at the decadal scale (Luo et al., 2010). Recently, Izumo et al. (2010) demonstrated that IOD events tend not only to co-occur with ENSO events but also to lead them through tropospheric biennial oscillation (Pillai and Mohankumar, 2010). These interdependencies are vital for enhancing prediction skills. MSES has the potential to capture such interdependencies when applied directly to such indices. However, this is beyond the scope of the study.

The synchronization and coherence between the Indian Ocean Dipole (IOD) and precipitation are given in Fig.5.3b and Fig.5.4b, respectively. The nonlinear dependence measure points to a significant synchronization at time scales of $8-16$ years in the southeastern regions $C1-C4$. The rest of the country seems to be unaffected by IOD. The WC analysis obtains a similar spatial pattern, however, the significant associations occur at shorter time scales ($1-4$ years, Fig.5.4b). Interestingly, with both methods, we cannot find any coupling in the Himalayan region ($C7$). The results obtained by MSES are in accordance with the general understanding that IOD plays a vital role in the Indian monsoon system in the southeast regions, i.e., in close proximity to the Indian Ocean, at interannual and decadal scales (Krishnan and Swapna, 2009). This result can be explained by the fact that one of the general conditions for Indian precipitation is the Tropical Easterly Jet and Tropical Westerly Jet (Rai and Dimri, 2017). In case of occurrence of IOD, the pressure dipole generated between the Tibetan plateau and the Madagascar Island either strengthens the southeastern Indian monsoon (positive IOD) or weakens it (negative IOD, Jiang and Ting, 2017). However, the reason for the association at the decadal scale is not apparent and needs further investigation.

Unlike IOD, North Atlantic Oscillation (NAO) demonstrates significant synchronization with precipitation across the entire subcontinent (Fig.5.3c). The association in the northern regions $C4$,$C5$ and $C7$ are strong and significant at interannual and decadal scales, whereas the southern regions $C1$,$C2$,$C3$ and $C6$ show weaker linkages. Overall, the strength of syn-

chronization is higher at the decadal scale than at the interannual scales. The comparison of the results obtained by MSES (Fig.5.3c) and WC (Fig.5.4c) reveals that the nonlinear method shows an increase in the association particularly in the northeastern Himalayan foothill region ($C4$). For some regions, MSES detects linkages which are not found by WC. For example, in the Himalayan region ($C7$), MSES shows a significant association at time scales of $4-16$ years, whereas WC shows only a signal just at the significance level at the scale of 16 years.
The overall MSES results are in congruence with other studies (Bhatla et al., 2016; Feliks et al., 2013; Goswami et al., 2006), but so far space and scale variation in the associations between NAO and Indian precipitation has not gained attention. The linkages between precipitation and NAO in the northern part of the country might be due to westerly influences from the Eurasian region which are, in turn, strongly affected by NAO. Another explanation by Goswami et al. (2006) suggests that the linkage of NAO and Indian precipitation at higher scales (decadal and beyond) in the northern part of India results from the interdependency of NAO and AMO.

In the case of Pacific Decadal Oscillation (PDO), we infer a robust decadal synchronization across the entire subcontinent (Fig.5.3d). The strength of synchronization varies in space, and reaches values of around 0.7 for several regions. On the contrary, WC (Fig.5.4d) does not reveal significant associations at the decadal scale except for the eastern coastline ($C1$) and Himalayan foothills ($C4$), where values at the boundary to significance are found. The MSES results agree with Krishnan and Sugi (2003) who demonstrate a strong relationship between PDO and precipitation across the country. The interannual synchronization might be a pseudo influence because of the interdependency of PDO and ENSO (Rathinasamy et al., 2017; Krishnan and Sugi, 2003).

Fig.5.3e shows the coupling between Atlantic Multidecadal Oscillation (AMO) and Indian precipitation. The highest strength of synchronization is observed in the northwestern and central regions ($C3-C6$). Weaker associations are detected in the south ($C1$,$C2$), whereas no significant synchronization is found for the Himalayan alpine region ($C7$). The linkages are most prominent at the decadal scale; in some regions also significant synchronization at interannual scales is found. In contrast, WC shows only weak associations (Fig.5.4e).
Our MSES results confirm the assertion given by Zhang and Delworth (2006) and Zhang and Delworth (2005) who found an in-phase relationship between Indian precipitation and AMO. A study by Goswami et al. (2006) also unravel a link between AMO and multidecadal variability of Indian precipitation. However, our study is the first to observe that the strength of the coupling between AMO and precipitation varies according to the different climate regions and is strongest at the decadal scale.

In summary, our findings re-confirm known physics-based associations, thus implicitly affirming the validity of our approach, but also provide new insights into Indian precipitation teleconnections. We find substantial variation in the significant linkages across India and for different time scales (Fig.5.5). MSES reveals an appreciable increase in the association between climate patterns and precipitation in most regions when compared to WC. In some regions, the values increase by $40-50\%$. The much higher skill of MSES in detecting associations suggests the presence of nonlinear and threshold relationships which can not be captured by WC which is limited to linear, Gaussian processes.

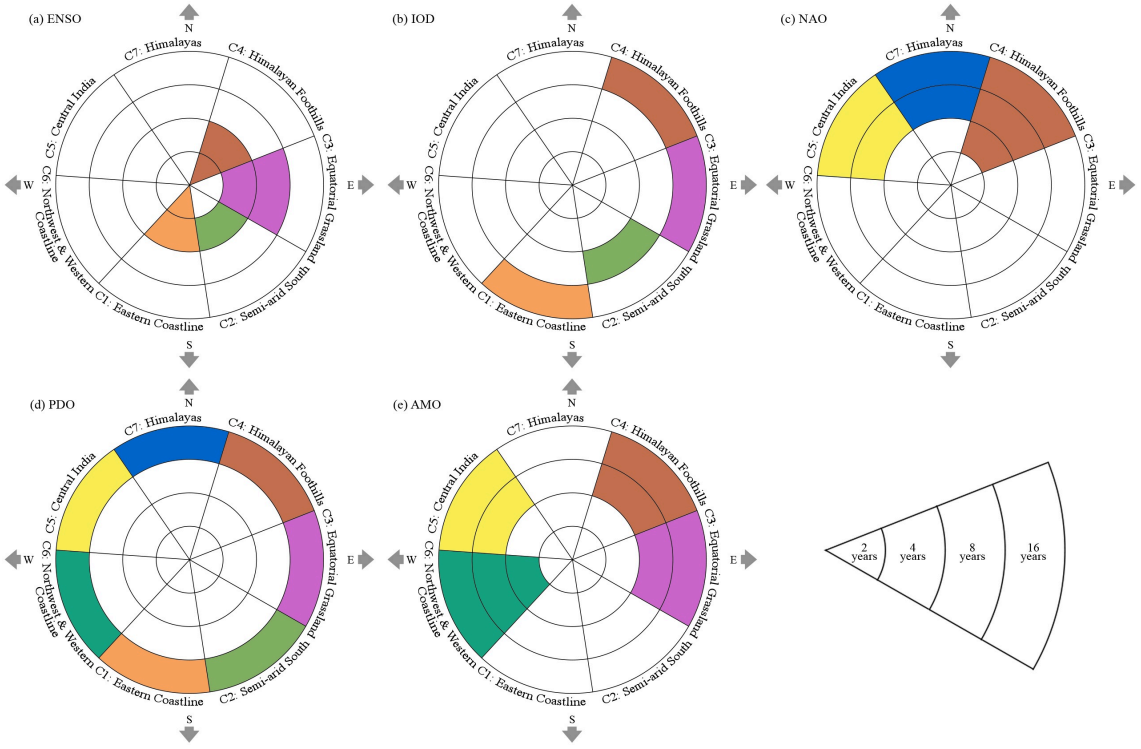



**Figure 5.5.:** Schematic map of spatial diversity of Indian precipitation teleconnections at different time scales. (a) Nino3.4, (b) IOD, (c) NAO, (d PDO, and (e) AMO. Color are consistent with the community shown in the Fig.5.2. Presence of color (irrespective of magnitude of synchronization) in community segment indicates significant synchronization between teleconnection and Indian precipitation. Every single segment of circle shows the temporal scale. Cardinal direction have been projected in the background of each circle.

## 5.5 Conclusions

A novel nonlinear, multi-scale approach (MSES) based on wavelets and event synchronization is used for unraveling teleconnection influences on Indian precipitation at multiple time scales. The analysis considers those climate patterns with highest relevance for Indian precipitation. To understand the spatial heterogeneity, India is sub-divided into homogeneous regions using complex networks. The comparison with wavelet coherence analysis (WC), the state-of-the-art method in understanding linkages at different time scales, shows a much higher skill for MSES in detecting associations between climate patterns and precipitation. This suggests that there are significant nonlinear linkages which are not well captured by linear approaches such as WC.

The application of MSES to the homogeneous regions allows unraveling the spatial diversity in the teleconnection patterns over India. ENSO has a strong influence on precipitation in the southeastern parts of the country. These regions are also affected by IOD, however, the IOD influence is much weaker compared to ENSO. NAO has a strong connection to precipitation particularly in the northern regions. The effect of PDO stretches across the whole country, whereas AMO influences precipitation particularly in the arid and semi-arid regions. The substantial variation of precipitation teleconnections across India and across time scales that is unraveled by the proposed method provides an exciting perspective for rainfall forecasting for India and for making better sense of its weather.

## 5.6 Acknowledgments

This research was funded by the Deutsche Forschungsgemeinschaft (DFG) (GRK 2043/1) within the graduate research training group "Natural risk in a changing world (NatRiskChange)" at the University of Potsdam (http://www.uni-potsdam.de/natriskchange). The authors gratefully thank the Roopam Shukla and Dr. Bedartha Goswami for helpful suggestion.

# Part III.

# Synthesis, conclusion and outlook

# Chapter 6

# Synthesis, conclusion and outlook

## 6.1 Synthesis

With the increasing intensity and frequency of extreme events (like floods and droughts) around the globe and the subsequent damages to life and property, it is necessary to develop strategies and methods for investigating the climate and other geophysical processes. With strong evidences of multi-scale behavior that triggers the abrupt and extreme events in the natural processes, it is urgent to develop strategies that incorporate a reliable framework to quantify multi-scale variability. It is plausible that these multi-scale interactions get camouflaged in the analysis using the traditional approach of analyzing the climatic interactions at single time scale (Paluš, 2014b). Therefore, at this point, the important and pertinent question to be answered is **"how to quantify and visualize multi-scale coupling and patterns in univariate or multivariate climate process to enhance our understanding about global climate system?"**
This chapter highlights the summary of the thesis along with a brief discussion on (i) salient contribution made through this research, (ii) conclusion, and, (iii) outlook.

Unraveling spatiotemporal patterns and interactions among climate variables has always been an important task for geoscientists in general, and for climatologists in particular, mostly because it contributes significantly to better prediction and forecasting. However, complexities are intrinsic to natural systems, and for this reason, the task of identifying patterns and interactions has always been challenging. Coupled with the existing challenges of global warming-induced climate change, these patterns and interactions become further unusual, unexpected and unpredictable. Recognizing these challenging realities, climate science studies, globally, recognize that climatic and other geophysical processes are intrinsically nonlinear and carry multi-scale features along with influences that are in general of time-varying nature. This thesis was largely prompted by these realizations and an implied aspiration to develop advanced methods with a capability to capture these aforementioned features.

Some of the observations that resulted in, as well as resulted from, this research are following:

1. In the investigation and quantification section, titled “Multi-scale event synchronization analysis for unraveling climate processes (chapter 2)”, different synthetic and real-world case studies were introduced, and the related processes were analyzed for intrinsic, but characteristic feature content.

The prominent observations that emerged from the study are as follows:

(a) Our investigation on various prototypical examples using wavelet confirms that climate-related processes show variability at a range of temporal scales due to the presence of noise, local disturbances, scale-emergent phenomena, non-stationarity, localized features and transient features of a system.

(b) Our series of test cases confirm that investigating coupling and synchronization on just at a single scale could give limited insights since, feedbacks and synchronization within a climate system might occur only at certain timescales and remain absent at other scales. In addition, single measures such as correlation and ES, are effective only for the stationary system, in other cases, they may underestimate or overestimate the strength of the coupling.

(c) MSES is superior compared to traditional ES, and provide greater insights of the interactions in the analyzed time series. Also, the effect of noise and local disturbances can be reduced to a greater extent (attributed to wavelet potential), and the underlying relationships become more prominent.

(d) MSES is able to deal with non-stationarity because wavelet can pick up the non-stationary and transient features of a system upto some extent, thereby improving the estimation of relationship.

(e) MSES on a real-world case study (precipitation data) suggests that proximity of stations does not necessarily indicate relationship at all scales. Also, we notice far stations shows high synchronization at higher scales. This observation puzzled us and motivated us to investigate in-depth analysis on the scale-distance relation of a climate process, which we have addressed in chapter 4.

All the above observations were interesting and formed a strong foundation for our convictions that multi-scale analysis of climatic processes holds the promise of a better understanding of the system dynamics that may be missed when analyzing process at a single timescale. However, it is to be noted that wavelets are limited in visualizing and representing large spatial connection and interactions. **Hence, the next challenge is to determine how to visualize the spatial interactions in large spatial dataset?**

2. To answer this question we employed a complex network in the chapter titled "Wavelet-based multi-scale similarity measure for complex network (chapter 3)" which provides analytical capabilities to uncover structure and patterns of the large spatial dataset and helps to visualize that.

   Here, our first aim was to evaluate the ability of Pearson's correlation coefficient (PCC) to capture multi-scale interactions and patterns among large spatial data, second, to foster a new method in cases where PCC is limited in capturing interactions over different timescale, third, to develop a network over different timescales to visualize and represent large spatial interactions and patterns, and fourth, its application on real dataset and their physical significance. The important outcomes of the study are as follows:

   (a) The analysis of the results from network constructed on stationary and non-stationary spatial datasets showed that PCC alone is not enough to capture the underlying multi-scale scale features of a large spatial climate system with many

nodes and hence, resulting in a poor estimation of interactions between spatial dataset.

(b) On the other hand, the proposed "Wavelet multi-scale correlation (WMC)" measure was capable of unravelling scale-specific interactions among large spatial data which are often undiscovered at a single scale.

(c) The network constructed based on WMC was efficient in (i) identifying the dominant processes and the timescale they act on, and (ii) understanding how the network (spatial interaction and patterns) evolve with temporal scales.

(d) Further, the application of the proposed approach when applied on global sea surface data captures patterns similar to the known climatic patterns such as ENSO patterns in the Pacific Ocean at inter-annual scale and the NAO in the North Atlantic on the intra-annual scale. The analyses also revealed a strong coupling of ENSO and IOD at inter-annual and interdecadal scale as well as the coupling between SAM, Southern Ocean SST, and ENSO.

This study clearly affirmed the validity of the WMC measure to detect climate modes on different temporal scales. It is to be noted that, so far, we have shown the application of WMC only on temporal scales (Fig.1.1). **Would it be worth investigating how the interactions and patterns change with spatial distances?** For example, how spatial coupling (short-range and long-range connections) between Indian and Pacific Ocean changes with time scales? This also gives us a chance to recall an unanswered question of the first study, i.e., how spatial relationship changes with different temporal scales?

3. To answer the questions raised in the previous study I used the same approach (WMC measure and complex network) and SST data to further investigated the length distribution of the significantly correlated edges (both positive and negative) and offer a 3–D visualization. This study (chapter 4) titled "Detection of short- and long-range teleconnections in SST patterns on different timescales" bridges the two climate scales, spatial and temporal (Fig.1.1) with the following objectives to (i) investigate scale-specific patterns over different timescales, (ii) investigate occurrence of short and long-range linkages between identified scale-specific patterns over different timescales, and (iii) offer a 3–D global visualization of the link distributions.
The outcomes of the study reveals that:

(a) Climate network constructed using SST dataset exhibits a timescale varying behavior and shows various high variability and co-variability regions which are well–known modes of SST variability (also discussed in chapter 3). Our analysis offer additional insights for example, although ENSO and IOD are known to impact each other via the atmosphere (Luo et al., 2010; Stuecker et al., 2017; Yamagata et al., 2013), but our presented node-link diagram reveals that links are mainly concentrated between the northern part of the El Niño tongue and the equatorial Indian Ocean (Fig.4.7e & f). Furthermore, the diagram reveals that the arrow shaped pattern of negatively correlated links (located in the western equatorial Pacific (Fig.4.4f)), known to occur at the same timescale as ENSO are neither linked to El Niño nor IOD. Our study also confirmed the interannual interactions between Southern Ocean and ENSO via Southern Annular Mode (SAM), and intradecadal interactions between Northern Pacific known as Pacific Decadal Oscillation (PDO) and Southern Ocean.

(b) Histogram of frequency distributions of the link lengths, shows short and long range teleconnections with varying spatial distance over different timescales. For instance, at interannual scale the tropics show the largest number of significant correlations acting over both short as well as long-ranges (teleconnections) with the lengths of the links of about 5,000 km between extratropics Southern Hemisphere (SH) and Tropical Hemisphere (TH) at this scale (Fig.4.6d &g). At the interannual to decadal scale mainly long-range links between extratropics Northern Hemisphere (NH) and TH with a length of about 5,000 km and 7,000 km exists (Fig.4.6b &e). In addition a small number of links between SH and TH with a length of about 15,000 km are also observed. There are no long-range links within the TH at intradecadal timescale. At the decadal scale, the existence of long-range connections between NH and SH to the tropics as well as interhemispheric (SH to NH) links are found.

(c) Our analysis further reveals that not all the teleconnections influence via the atmosphere. The 3-D visualization highlight the see saw response (Rahmstorf, 2002) at intradecadal scale between North Atlantic and Southern Ocean. This is due to the transport of heat from the Southern Ocean to the North Atlantic via the Atlantic Meridional Overturning Circulation (AMOC; Caesar et al. (2018)).

In general, the outcomes of this study are vital to distinguish between the regions that are highly linked through significantly correlated short-range edges and those that have long-range teleconnections. By understanding the characteristics of the teleconnections we can incorporate remote regions into climate models as their significant influence to the climatic system are revealed with this study. **In the next course of study we aim to investigate the causative effect of these oceanic patterns on extreme rainfall using the proposed multi-scale measure (MSES).**

4. In chapter 5 titled "Unraveling spatial diversity of Indian rainfall teleconnections", the multivariate interactions and patterns between extreme Indian precipitation and climate patterns over different time scales were investigated. Such studies have a vital application since it is believed that many extreme events (e.g. extreme precipitation, heat waves) are resultant of different scale interactions between climate variables (Goswami et al., 2010). We selected India as a study region since India is believed to have a mysterious monsoon resultant of multi-scale interactions of several climate patterns (Suhas et al., 2012). In this study, the dominant and relevant climatic patterns for India, represented by climatic indices such as El Niño/Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), North Atlantic Oscillation (NAO), Pacific Decadal Oscillation (PDO) and Atlantic Multi-decadal Oscillation (AMO) were considered. Few salient features and observations from the study are listed below.

   a) The uniqueness of the studies lies in the fact that it involves all the three-climate scales- temporal, spatial and topological (Fig.1.1). First, we identify homogeneous communities (for more details see Appendix E) and later compute non-linear linkages between the identified communities (spatial regions) and dominant climatic patterns, represented by climatic indices such as El-Nino Southern Oscillation, Indian Ocean Dipole, North Atlantic Oscillation, Pacific Decadal Oscillation and Atlantic Multi-decadal Oscillation.

   b) The study unravels the spatial diversity in the teleconnection patterns over India and shows that ENSO has a strong influence on precipitation in the southeastern

parts of the country. These regions are also affected by IOD, however, the IOD influence is much weaker compared to that of ENSO. NAO has a strong connection to precipitation, particularly in the northern regions. The effect of PDO stretches across the whole country, whereas AMO influences precipitation particularly in the arid and semi-arid regions.

c) Substantial variation of precipitation teleconnections across India and across time scales, unraveled by the proposed method refers to a recurring and persistent, large-scale pattern of pressure and circulation anomalies that spans over India. Such teleconnection patterns are also referred to as preferred modes of low-frequency (or long time scale) variability and are a naturally occurring aspect of our chaotic atmospheric system. An in-depth understanding of such precipitation teleconnections is imperative for better weather predictions.

d) The results from MSES when compared with the state-of-the-art method, wavelet coherence (WC) showed that MSES has better ability to detect linkages between the Indian monsoon system and climate patterns. For example, in some regions, the values increase by $40-50\%$ suggesting the presence of nonlinear and threshold relationships which weren't captured by WC.

## 6.2 Conclusion

This thesis titled "Unraveling spatio-temporal climatic patterns via multi-scale complex networks" fosters four main advances to the climate science.

1. It sheds light on the definition and understanding of interactions of scales within the climate systems.

2. It investigates climate process at multiple scales and proposes advanced linear and non-linear methods to quantify them.

3. It applies proposed methods on univariate global sea surface temperature data to investigate spatial interactions and patterns at different timescales and offer a 3-D visualization for advanced understanding.

4. Last but not least, it narrows down the focus and unravels the spatial diversity of Indian precipitation teleconnections by quantifying multivariate relationships among identified SST patterns (represented by climate indices) and extreme Indian precipitation at different timescales.

Findings of the thesis highlight the need to consider the multi-scale variability and dynamics of the geophysical process, in general and climate process, in particular. All proposed new methods such as MSES, WMC, and identified climatic patterns, will appeal to the broader society of Earth scientists and modelers given the problems they face in understanding the spatio-temporal variability of the climate variables, scale-distance relationships, and coupling between the teleconnections and rainfall.

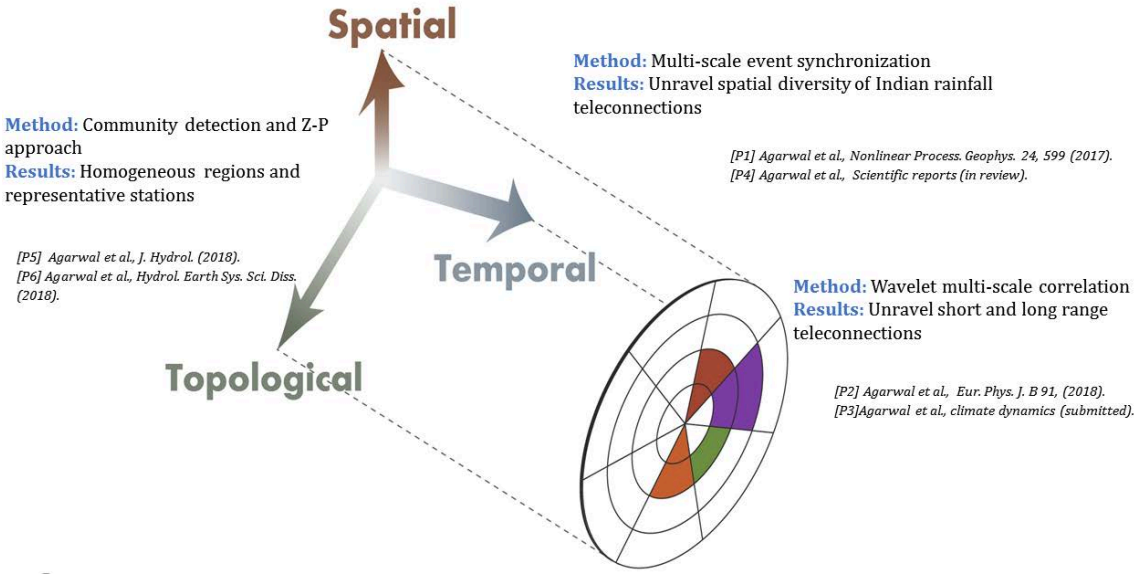



**Figure 6.1.:** Figure highlights the scientific contributions and how this thesis has bridged the existing research gaps by studying the climatic processes at different scales.

## 6.3 Outlook

In this thesis, we showed the improvement in understanding the climate system when considering multi-scale characteristics of climate interactions and patterns. However, the interactions in univariate and multivariate process might also occur at cross-scale (Fig.6.2). Hence, it would be worth exploring the *cross-scale* interactions using the proposed measures.

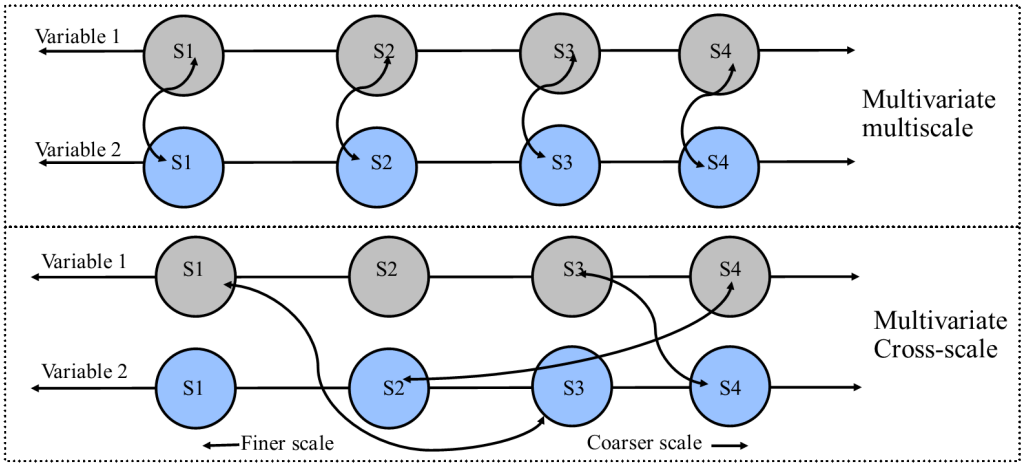



**Figure 6.2.:** Schematic illustrations of multivariate multi-scale and multivariate cross-scale interactions. In the multivariate multi-scale case, possible interactions are analyzed at the same scales, whereas the multivariate cross-scale study allows investigating interactions between any combinations of scales.

Our study on sea surface temperature (chapter 4) and extreme Indian precipitation (chapter 5) showed that another important domain of research is to understand the *network–of-network* structure of multivariate climate variables. It is imperative to investigate one step further and quantify the multivariate cross-scale synchronization and causality across

time-scales between the identified SST patterns, such as the ENSO pattern, and extreme PCP for the entire globe. Such multivariate network-of-network investigation at multi-scale or cross-scale could better guide us to understand "how do large scale climatic patterns and abrupt transitions trigger the extreme events? "

For Instance, Rheinwalt et al., (2015) used the network-of-network approach to investigate the influence of sea surface temperature variability on monsoon system and revealed a pronounced precipitation dipole between Southeast Asia and the Afghanistan-Pakistan region, which is supposed to be controlled by an interplay of the Madden-Julian oscillation, and the African-Arabian jet stream. Further, Donges et al., (2011) studied the topology of interacting climate variables such as three-dimensional geopotential heights and uncovered known as well as novel features of the atmosphere's vertical stratification and general circulation. Although, the results of both the studies are interesting in climate science, however they are limited to single scale only.

Another crucial aspect that emerges from this thesis is to investigate the *temporal evolution* of climate network. We showed in the case study II of chapter 3 that by generating a climate network at different timescales for different time periods we might be able to capture the dynamical mechanism for major climate shifts. Such investigations could shed lights on the mysterious events in recent climate history of climate shift in the mid-1970s.

# Part IV.

# Appendix

# Appendix A

# Quantifying the role of single station within homogeneous regions using complex network analysis

*Agarwal, A., Marwan, N., Maheswaran, R., Merz, B. and Kurths, J.: Quantifying the roles of single stations within homogeneous regions using complex network analysis, Journal of Hydrology, doi:10.1016/j.jhydrol.2018.06.050, 2018.*

Regionalization and pooling stations to form homogeneous regions or communities are essential for reliable parameter transfer, prediction in ungauged basins, and estimation of missing information. Over the years, several clustering methods have been proposed for regional analysis. Most of these methods are able to quantify the study region in terms of homogeneity but fail to provide microscopic information about the interaction between communities, as well as about each station within the communities. We propose a complex network-based approach to extract this valuable information and demonstrate the potential of our approach using a rainfall network constructed from the Indian gridded daily precipitation data. The communities were identified using the network-theoretical community detection algorithm for maximizing the modularity. Further, the grid points (nodes) were classified into universal roles according to their pattern of within- and between-community connections. The method thus yields zoomed-in details of individual rainfall grids within each community.

## A.1 Introduction

Reliable and accurate information about precipitation is essential for most hydrological studies. For example, precipitation observations are required for the design of hydraulic structures, flood estimation and forecasting, assessment of water availability, or climate impact studies. However, in most situations, raingauges are scarce, requiring knowledge about how precipitation characteristics at neighboring stations are related. These interrelationships can be viewed in a statistical sense (e.g. by applying correlation analysis), in a physical sense (as in dynamical meteorology), or in a topological sense (as in complex network analysis). Knowledge of these interrelationships will be crucial for various purposes, including (1) applying interpolation/extrapolation techniques to generate rainfall at locations where raingauge measurements are not available (Yang et al., 2015), (2) filling gaps in historical rainfall records using available rainfall observations at neighboring stations

(Jha et al., 2015), (3) determining the optimal density and locations for the installation of new raingauges (Mishra and Coulibaly, 2009; Pardo-Igúzquiza, 1998), and (4) analysing regional flood frequency (Hassan and Ping, 2012; Smith et al., 2015; Zrinji and Burn, 1994; Zrinji and Burn, 1996).

Even though there is a plethora of methods available for identifying homogeneous regions, such as clustering algorithms (Hsu and Li, 2010; Agarwal et al., 2016b), principal component analysis (Darand and Mansouri Daneshvar, 2014), region-of-influence approach (Zrinji and Burn, 1994; Zrinji and Burn, 1996), or multiple regression (Sivakumar et al., 2015), there are some important challenges which need to be addressed.

A common assumption in studies (Razavi and Coulibaly, 2013; Salinas et al., 2013) dealing with interpolation/extrapolation, missing values and prediction in ungauged basins (PUB) is that the variables of interest, such as precipitation characteristics, at nearby points are more closely related than those at distant points, as described by Tobler (1970) in his 'First Law of Geography'. This assumption is also the foundation of geostatistics, which in turn is fundamental to many classical approaches to spatial data analysis and interpolation throughout hydrology and other geoscientific disciplines. While this assumption is often reasonable, it may not hold in every situation, especially in regions with complex topography (Jha et al., 2015). In such areas, statistics of rainfall recorded at neighboring stations can significantly vary due to the high topographic gradients and, hence, changes in rainfall patterns between them (Berndtsson, 1988; Niu, 2013; Özger et al., 2010).

A significant disadvantage of these methods is that the selection of factors for identifying the similarity in rainfall patterns is highly subjective. They rely on the preconceived notion of the existence of linear relationship between the factors that influence the precipitation in a region. For instance, in PCA method the subjectivity is introduced in terms of extraction method, rotation method, number of compoenents to be retined etc. For more details refer to Saxena et al. (2016).

More importantly, the traditional methods for pooling stations within homogeneous regions are not capable of unraveling the role of each raingauge station within the community. This includes the interactions within the community, the role of the stations, and the strength and number of inter- and intra-community connections.

The main aim of this chapter is to address this last point by proposing a network-based approach for unravelling the role of each node in a community. This microscopic analysis is essential to understand the role of each of the member stations of the community and is very useful in many applications. For example, by knowing the connections and their strength, it is possible to reduce the uncertainty of predictions at ungauged locations by including only those stations that have strong connections in that community. Similarly, the reliability of filling gaps in observational time series can be improved by identifying the stations that share strong connections with that particular station. The relative importance of the stations in the community will also help in understanding the connection between the communities and is particularly useful for selecting stations that share characteristics with more than one community.

In the context of connections within rainfall systems, recent developments in network theory, especially regarding complex networks, have been found useful for identifying the

spatial connections in rainfall (Malik et al., 2012). Steinhaeuser et al. (2010) explored the utility of complex networks to analyze climate data, i.e., air temperature, pressure, relative humidity and perceptible water. They used the WalkTrap community detection algorithm to identify communities. They concluded that these communities have a climatological interpretation and that alterations in community structure can be an indicator of climatic events. Tsonis et al. (2011) applied complex networks and modularity based community detection to observed and simulated model data and concluded that the complexity of the system condenses into small interacting components called communities. This approach provided information about the nature of different climate subsystems. Jha et al. (2015) demonstrated the use of the clustering coefficient, a complex network based measure (Stolbova et al., 2014), on two rainfall networks in Australia. They attempted to relate the strength of spatial connections in rainfall to topographic and rainfall properties, towards identifying dominant factors governing spatial connections and for offering a better physical interpretation on spatial rainfall variability. Eustace et al. (2015) identified community structures by proposing local community neighborhoods ratio algorithm and showed that the algorithm detects well-defined communities in networks by a wide margin. Conticello et al. (2018) applied the Louvain community detection algorithm to identify clusters of rainfall stations using the concept of event synchronization and Self Organizing Maps. Even though the study of Halverson and Fleming (2015) on streamflow regionalization is not directly relevant for rainfall, it showed that the choice of the community detection algorithm does not strongly impact the community structure.

All above-mentioned studies have used complex network based community detection algorithm to identify homogenous regions but little attention has been paid to the different characteristics or roles of each of the member stations of a community. Although Halverson and Fleming (2015) have identified the high priority stations, based on high betweenness centrality values, but have not discussed the role of other stations. This study shows that the microscopic analysis of homogeneous regions provides additional insights into the behavior and dynamics of single stations within the homogeneous region, which can be vital for many engineering and water management purposes.

This study builds on emerging ideas in the very fast-evolving field of complex network theory and contributes to work in hydro-monitoring system design (Mitra et al., 2017a). Although studies in different fields, such as physics (Quiroga et al., 2000; Quiroga et al., 2002a) or neurology (Rubinov and Sporns, 2011; Zhou et al., 2007), have seen immense use of complex network theory, event synchronization, and Z-P space, our study is the first combined application of these methods in hydrology to date. It clearly demonstrates the large potential of these methods in hydrology.

As advancement to the research in the application of complex networks in rainfall network analysis, we use a network-based measure to provide a comprehensive analysis of the stations in a community and their roles. For this, we apply the concept of cartographic representation of networks by Guimerà and Amaral (2005). The proposed approach is demonstrated using the synthetic network and then applied to the Indian Precipitation gridded precipitation dataset. The paper is organized in the following manner. Section A.2 describes the basic aspects of network construction, network measurements and the methods used in the study. The application of the methodology and the subsequent results obtained are discussed in detail in Section A.3. The conclusions are reported in Section A.4.

## A.2 Methods

### A.2.1. Network definition

A network or a graph is a collection of entities (nodes, vertices) interconnected by lines (links, edges) as shown in Fig.A.1. These entities could be anything from humans defining social networks (Arenas et al., 2008), computers in web networks (Zlatić et al., 2006), neurons of the brain (Pfurtscheller and Lopes da Silva, 1999; Zhou et al., 2007), streamflow stations defining hydrological networks (Halverson and Fleming, 2015; Sivakumar and Woldemeskel, 2014) to raingauge stations defining climate networks (Stolbova et al., 2014; Ozturk et al., 2018a; Steinhaeuser et al., 2011).

Formally, a network or graph is defined as an ordered pair $G = \{N, E\}$, containing a set of nodes $N = \{N_1, N_2, \ldots N_N\}$ together with a set $E$ of edges $\{i, j\}$ which are 2-element subsets of $N$. In this work, we consider undirected and unweighted graph ($G$), where only one edge can exist between a pair of nodes and self-loops of the type $\{i, i\}$ are not allowed. Hence, edges simply show connections between nodes, and each edge can be traversed in either direction. This type of graph can be represented by the symmetrical adjacency matrix (Stolbova et al., 2014)

$$A_{i,j} = \begin{cases} 1, & \{i, j\} \in E \\ 0, & \{i, j\} \notin E. \end{cases} \tag{A.1}$$

Figure A.1 is a simple example of an undirected and unweighted network. In general, large graphs with non-trivial topological characteristics, used to represent real systems, are called complex networks. To define whether a link between two nodes exists, any similarity measure can be used, such as correlation (Donges et al., 2011; Jha et al., 2015), synchronization (Conticello et al., 2018; Stolbova et al., 2014; Ozturk et al., 2018a) or mutual information (Paluš, 2018). Depending on the topological structure of the network, groups of nodes can be pooled together forming communities (Jha et al., 2015).

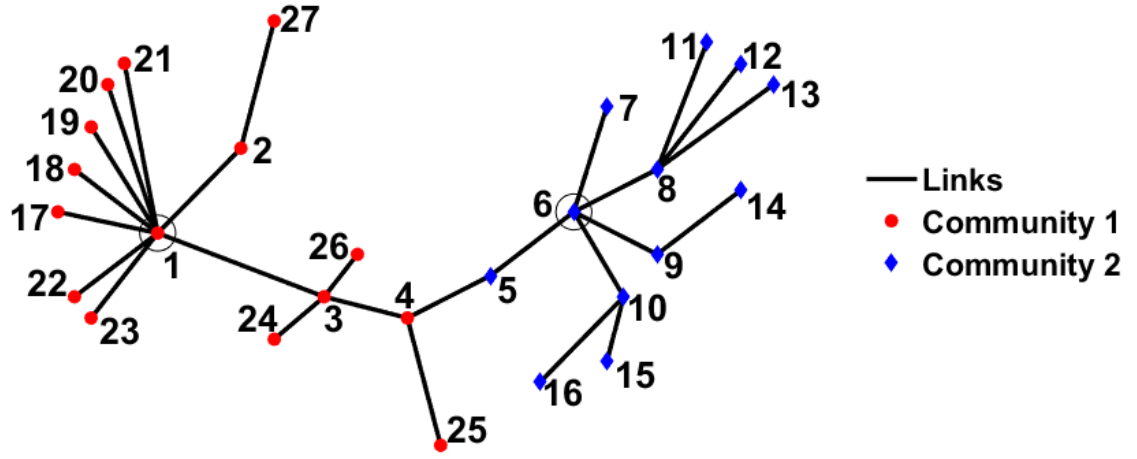



**Figure A.1.:** The topology of the sample network used to explain the network construction and universal role of a node. Different colors represent different communities, i.e., community 1 (red) and community 2 (blue). Nodes 4 and 5 are the hybrid nodes connecting their community to the other community. Nodes 1 and 6 are the hubs of their respective community.

### A.2.2. Event synchronization

We use event synchronization to define whether a link between two nodes exists. Event synchronization (ES) has been specifically designed to calculate nonlinear relations between timeseries with events defined on them. A simple algorithm proposed by Quiroga et al. (2002b) can be used for any time series for which we can define events, such as single-neuron recordings, epileptiform spikes in electroencephalograms (EEG), heartbeats, stock market crashes, or rainfall events. When dealing with signals of a different character, the events could be defined differently in each time series, since their common cause might manifest itself differently in different time series. ES has advantages over other time-delayed correlation techniques (e.g., Pearson lag correlation), as it uses a dynamic (not fixed) time delay (Agarwal et al., 2017). The latter refers to a time delay that is adjusted according to the two time series being compared, which allows its application to different situations. Another advantage of ES is that it can be applied to non-Gaussian data (Stolbova et al., 2014). Having its roots in neuroscience, ES only considers events beyond a threshold and ignores the absolute magnitude of events, which could be a challenge to incorporate in future work.

A number of modifications have been proposed to the basic algorithm, considering various issues such as boundary effects or bias toward the number of events (Agarwal et al., 2017; Rheinwalt et al., 2016). The modified algorithm proposed by Rheinwalt et al. (2016) can be explained as follows: An event above a threshold $\alpha$ percentile occurs in the signals $x(t)$ and $y(t)$ at times $t_l^x$ and $t_m^y$ where $l = 1,2,3,4...S_x$, $m = 1,2,3,4...S_y$ and within a time lag $\pm\tau_{lm}^{xy}$ which is defined as

$$\tau_{lm}^{xy} = \min\{t_{l+1}^x - t_l^x, t_l^x - t_{l-1}^x, t_{m+1}^y - t_m^y, t_m^y - t_{m-1}^y\} \tag{A.2}$$

where $S_x$ and $S_y$ are the total number of events (greater than the threshold $\alpha$) in the signals $x(t)$ and $y(t)$, respectively. This definition of the time lag helps to separate independent events, which in turn allows to take into account the fact that different processes are responsible for the generation of events. To count the number of times an event occur in $x(t)$ after it appears in $y(t)$ and vice versa, $C(x|y)$ and $C(y|x)$ are defined as follows:

$$C(x|y) = \sum_{l=1}^{S_x} \sum_{m=1}^{S_y} J_{xy} \tag{A.3}$$

and

$$J_{xy} = \begin{cases} 1, & \text{if } 0 < t_x^l - t_m^y < \tau_{lm}^{xy} \\ \frac{1}{2}, & \text{if } t_x^l = t_m^y \\ 0, & \text{otherwise.} \end{cases} \tag{A.4}$$

$C(y|x)$ is calculated analogously but with exchanged $x$ and $y$. From these quantities we obtain the symmetric measure:

$$Q_{xy} = \frac{C(x|y) + C(y|x)}{\sqrt{(S_x - 2)(S_y - 2)}} \tag{A.5}$$

$Q_{xy}$ is a measure of the strength of event synchronization between $x(t)$ and $y(t)$. It is normalized to $0 < Q_{xy} < 1$, with $Q_{xy} = 1$ for perfect synchronization (coincidence of extreme events) between signals $x(t)$ and $y(t)$.

### A.2.3. Network construction

To construct a rainfall network, each grid cell is considered as a node and links between each pair of nodes are setup based on the statistical relationship between them. The similarity measure used is the ES which gives a $Q$ matrix (Eq.E.4). Applying a certain threshold($\theta$) on the $Q$ matrix (Eq.E.4), we yield an adjacency matrix (rewriting Eq.A.1)

$$A_{i,j} = \begin{cases} 1, & \text{if } Q_{i,j} \geq \theta^{Q}_{i,j}. \\ 0, & \text{otherwise.} \end{cases} \tag{A.6}$$

Here, $\theta^{Q}_{i,j}$= $95th$ percentile is a chosen threshold, and $A_{i,j} = 1$ denotes a link between the $i^{th}$ and $j^{th}$ nodes and 0 denotes otherwise.The adjacency matrix represents the connections in the rainfall network. In this study, we use an undirected network, meaning we do not consider which of the two synchronized events happened first, in order to avoid the possibility of misleading directionalities of event occurrences between nodes that are topographically close to one another.

### A.2.4. Network measures

To analyze and quantify the topological features of complex networks, a large number of network measures have been introduced (Blondel et al., 2008; Malik et al., 2016). We use the within-module degree Z-score ($Z$) and the participation coefficient (P) (Guimerà and Amaral, 2005) to investigate the role of individual nodes within a community. $Z$ identifies hubs and non-hubs within the community. Hubs are nodes with a significantly larger number of links compared to the other nodes in the network. $P$ is a measure of the diversity of the connections between individual nodes and identifies to which extent a node has intra-community or inter-community links.

The within-module degree ($Z_i$ or Z-score) is a within-community version of degree centrality (total number of link of any node) and shows how well a node is connected to other nodes in the same community. It is estimated as (Guimerà and Amaral, 2005)

$$Z_i = \frac{K_i - \overline{K_{s_i}}}{\sigma_{K_{s_i}}} \tag{A.7}$$

where $K_i$ is the total number of links (degree) of node $i$ in the community $s_i$, $\overline{K_{s_i}}$ is the average degree of all nodes in the community $s_i$, and $\sigma_{K_{s_i}}$ is the standard deviation of K in $s_i$. Since two nodes having the same Z-score may play different roles within the community, this measure is often combined with the participation coefficient $P_i$.
The participation coefficient ($P_i$) compares the number of links of node $i$ to nodes in all communities with the number of links within its own community. We define the $P_i$ of node $i$ as (Guimerà and Amaral, 2005)

$$P_i = 1 - \sum_{s_j=1}^{N_M} \left(\frac{K_{is_j}}{K_i}\right)^2 \tag{A.8}$$

where $K_{is_j}$ is the number of links of node $i$ to nodes in community $s_j$, and $k_i$ is the total number of links (degree) of node $i$. $N_M$ represent the number of communities in the network. The participation coefficient of a node is therefore close to one if its links are uniformly

distributed among all the communities, and zero if its entire links are within its own community because in later case $K_{is_j} = K_i$ hence $P_i = 0$.

### A.2.5. Community detection

Complex networks often show subsets of nodes that are densely interconnected. These subsets are called communities. The community structure of a complex network provides insight into the network (Girvan and Newman, 2002). For instance, different communities within a network may have very different properties compared to the averaged properties of the complete network.

There exist several community detection approaches aiming at stratifying the nodes into communities in an optimal way (see Lancichinetti and Fortunato (2009) for an extensive review). The question which community detection algorithm should be used is difficult to answer. However, it has been found that the choice of the community detection algorithm has a small impact on the resultant communities in geophysical data science studies (Halverson and Fleming, 2015). In this study, we use the Louvain method which maximizes the modularity to find the optimal community structure in the network. The optimal community structure is a subdivision of the network into non-overlapping groups of nodes, which maximizes the number of within-group edges and minimizes the number of between-group edges (Blondel et al., 2008; Rubinov and Sporns, 2011).

Modularity is defined, besides a multiplicative constant, as the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random. Positive modularity values suggest the presence of communities. Thus, one can search for community structures by looking for the network divisions that have positive, and preferably large, modularity values (Newman, 2004). Modularity (M) is calculated as:

$$M = \frac{1}{2m} \sum_{i,j} [A_{ij} - \frac{k_i k_j}{2m}] \delta(C_i C_j) \tag{A.9}$$

where $A_{ij}$ represents the number of edges between $i$ and $j$, $k_i = \sum_j A_{ij}$ is the sum of the number of the edges (degree) attached to vertex i, $C_i$ is the community to which vertex $i$ is assigned, the $\delta$-function $\delta(u,v)$ is 1 if $u = v$ and 0 otherwise, and $m = \frac{1}{2}\sum_{ij} A_{ij}$.

Equation (A.9) is solved using the two-step iterative algorithm proposed by Blondel et al. (2008), also known as the Louvain method. The first step consists in optimizing the modularity by permitting only a local modification of communities; in the second step, the communities identified are pooled to assemble a new network of communities. High modularity networks are densely linked within communities but sparsely linked between communities. The algorithm stops when the highest modularity is achieved. The algorithm was implemented using the Brain Connectivity Toolbox (BCT), provided by Rubinov and Sporns (2011), and is available at https://sites.google.com/site/bctnet/.

### A.2.6. Z-P space approach

Following the approach proposed by Guimerà and Amaral (2005), we calculate for each node the participation coefficient $P_i$ and the within-module degree $Z_i$, and plot all nodes onto the Z-P space. Both measures are calculated once the network communities have been determined (Guimerà and Amaral, 2005; Guimerà et al., 2007). Guimerà et al. (2007)

propose to divide the Z-P space into seven classes (R1–R7) which express the different roles of the nodes (Table A.1). In the first step, the nodes are broadly categorized as hubs or non-hubs using the within-module degree (Z). Nodes with $Z \geq 2.5$ are classified as community hubs and nodes with $Z < 2.5$ as non-hubs. At the second level, the hub and non-hub nodes are further characterized using the participation coefficient. Hence, each node is assigned to one of these seven classes.

**Table A.1.:** Definition and interpretation of R classes according to Guimerà et al. (2007), defining the role of each node.

| R-Class | Z | P | Remarks | Characteristics of R class |
|---|---|---|---|---|
| R1 | <2.5 | $\approx 0$ | ultra-peripheral nodes, i.e.,nodes with almost all their links within their community ($P \approx 0$). | representative nodes (almost all intramodular links) |
| R2 | <2.5 | $0 < P \leq 0.625$ | peripheral nodes, i.e., node has at least 60% its links within the community. | node has more intramodular links than intermodular links |
| R3 | <2.5 | $0.625 < P \leq 0.80$ | satellite connectors, i.e., nodes have half of its connection outside the community. | node has more intermodular links than intramodular links |
| R4 | <2.5 | P > 0.80 | kinless nodes, i.e., nodes with a maximum of links (>70%) outside the community. | wrongly assigned nodes |
| R5 | >2.5 | $P \leq 0.30$ | provincial hubs,i.e., hubs with the vast majority of links within their community. | local centers, representative nodes if $P \approx 0$ |
| R6 | >2.5 | $0.30 < P \leq 0.75$ | connector hubs,i.e., hubs with atleast half of its links to other community. | hybrid nodes (connecting two different communities) |
| R7 | >2.5 | P > 0.75 | global hubs,i.e., hubs with links homogeneously distributed among all community. | global connector nodes hence cannot be assigned to the single community. |

Nodes in the classes R1 and R5 with $P \approx 0$ have almost all links within the own community. Since class R5 have provincial hubs (Table A.1) which contain both intracommunity and intercommunity links, the limit on the participation coefficient ($P \approx 0$) helps to identify nodes that have almost all intracommunity links. These nodes with almost all intracommunity links ($P \approx 0$) are local centers in the region and can only be selected as a representative node of the community (Halverson and Fleming, 2015).
Nodes in the classes R2 and R3 are peripheral and satellite connectors respectively (Table A.1). Both the class contains hybrid non-hub nodes which generally connect two different communities. The only difference between R2 and R3 is that R3 nodes have more intercommunity links (outside of its own community).

Similarly, R6 nodes represent the nodes that have many intercommunity links but are hubs. In the given community we interpret them as hybrid hubs which have a maximum connection outside of its own community. Kinless nodes (R4) have the greatest proportion of links outside the community and are interpreted as wrongly assigned nodes in the community. If there exist many R4 nodes in the community a reformation of the communities

or reallocation of such nodes to appropriate community is suggested. The nodes in class R7 maintain homogeneous links with all the communities. We surmise that such nodes may not be clearly associated with a single community hence termed as the global hubs or global connectors (nodes connecting many different climate sub-systems).

The above characterization of nodes is important as it helps in understanding their specific roles in terms of non-hubs, hubs, local centers, hybrid nodes, global hubs. In the context of climate systems, local centers correspond to nodes which are important for local climate phenomena, while bridges correspond to nodes which connect different subsystem of climatology, leading to non-local interaction (teleconnections).

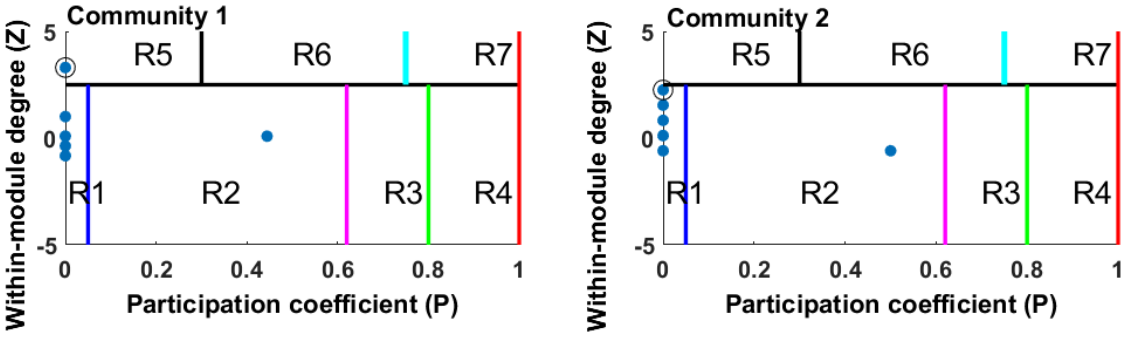



**Figure A.2.:** Nodes of the sample network of Figure A.1 plotted onto the Z-P-space. Nodes 1 and 6 (both encircled) are the representative stations for community 1 and 2, respectively. Nodes 4 and 5 in community 1 and 2, respectively, are the only hybrid nodes and are thus in the R2 class. All other nodes have only intracommunity links and are assigned to the R1 class. Many stations have the same values for Z and P and are thus on top of each other in the R1 class. Nodes 1 and 6 are the local center ($P \approx 0$)and are thus in the R5 and R1 class respectively.

Using the classification of Table A.1, Figure A.2 shows the Z-P space for the sample network of Figure A.1 and the assigned R classes. Node 1 is a hub in community 1, having all of its nodes within the community, and hence can be considered as a representative station. Node 4 of community 1 (non-hub) has intercommunity links and thus falls in the R2 class. For community 2, station 6 is a representative node with all links within the community, and the non-hub node 5 has intercommunity links falling in class R2. There is no kinless node (R4 and R7) in both communities.

If there exists a node fully unsynchronized to the other nodes in the network, i.e. there are no links to other nodes, the proposed Z-P approached will detect this station given its unique characteristcis. This unsynchronized station will lie at the origin of Z-P space and will fall in a community on its own. As an extreme example, one might imagine that in a meteorological sub-region, characterized by fine-scale convective thunderstorms with sparse raingauge coverage, precipitation event synchronization across all raingauges in that sub-region would be poor and each stations would form a separate community.

## A.3 Model application

The method was tested on a gridded rainfall dataset for two reasons: i) the availability and the access to raingauge data is limted, and ii) gridded datasets provide an effective platform to understand the precipitation dynamics. Owing to the assumptions underlying the spatial interpolation, the gridding process used to build the dataset might affect the relationships between nodes. However, these effects can be neglected considering the extent of the study area. The high-resolution ($0.25° \times 0.25°$) daily gridded rainfall data (Pai et al., 2014) was developed by the Indian Meteorological Department (IMD) for the spatial domain of 66.5°E to 100°E and 6.5°N to 38.5°N, covering the mainland region of India. The gridded data was generated from the observed data of 6995 gauging stations across India using spatial interpolation for the period $1901-2013$. Several studies in the past using the same dataset have reported such as downscaling (Sehgal et al., 2016; Lakhanpal et al., 2017) and rainfall variability (Krishnamurthy and Shukla, 2000). This shows that the data are highly accurate and capable of capturing the spatial distribution of rainfall over the country. In this study, out of total 6995 grid stations, 4631 stations were identified for which continuous rainfall data for 63 years (Jan. 1951 to Dec. 2013) was available without any missing values.

The rainfall network is constructed (as explained in section A.2.3) by extracting an event series from 4631 raingauges (Fig.A.3), i.e., by applying a threshold we identify extreme rainfall events in the given time series (Agarwal et al., 2017; Rheinwalt et al., 2015). We define extreme events as precipitation that is greater than the $95th$ percentile at that station. The $95th$ percentile is a good compromise between having a sufficient number of events at each location and a rather high threshold to study heavy precipitation. Next, we compute the $Q$ (Eq.E.4) between each pair of 4631 rainfall grid points. Applying a threshold ($\theta^{Q}_{i,j} = 95th$ percentile) on the $Q$ matrix (Eq.E.4) yields an adjacency matrix (Eq.A.6), representing the connections in the rainfall network. In this study, we use an undirected network, meaning we do not consider which of the two synchronized events happened first, in order to avoid the possibility of misleading directionalities of event occurrences between rain gauges that are topographically close to one another. After formation of the rainfall network, we aimed to obtain a small set of communities representing relevant sub-processes of the rainfall network. In this study, we apply Louvain algorithm (section A.2.5) on the constructed network to unravel the community structure.

The resultant community structure is the rainfall network mapped in Figure A.3.

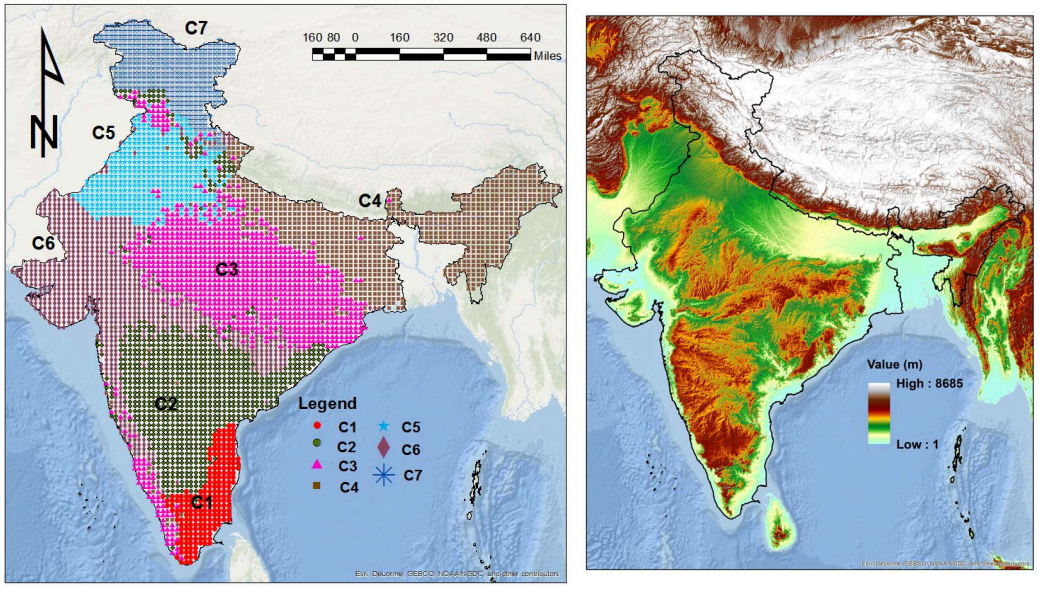



**Figure A.3.:** (a) Community structure of precipitation data in the rainfall network resulting from the Louvain algorithm. (b) Elevation map of the Indian continent.

The obtained community structure (Fig.A.3) shows some similar patterns to those provided by the Indian Institute of Tropical Meteorology Vinnarasi and Dhanya (2016) and Malik et al. (2016). It is also important to emphasize that the formation of the regions using complex networks is based on a cluster of actual connections, rather than on our traditional criteria of geographic proximity, nearest neighbors, regional patterns, and linear correlations.

Table A.2 shows the geographical and statistical interpretation of the resultant community which includes the mean, standard deviation, and coefficient of skewness of the precipitation distribution for each community. Higher mean precipitation shows a greater total amount of precipitation, a larger standard deviation shows a stronger variation of data for the collecting period, and a larger coefficient of skewness indicates more extreme (monthly) precipitation events (Hsu and Li, 2010).

**Table A.2.:** Summary of geographical and statistical analysis for each individual community. Communities formed by maximizing the modularity using Louvain algorithm. Elevation map for India is presented in the Fig.A.3b.

| C. No. | Number of stations | Monthly mean (mm) | Stand. Deviation (mm) | Skewness | Remarks |
|---|---|---|---|---|---|
| 1 | 214 | 79.70 | 98.29 | 2.04 | smallest community, eastern coastline, low elevation region, warm, humid climate regime |
| 2 | 876 | 76.30 | 104.45 | 2.16 | mild elevation, semi-arid climate regime (south) |
| 3 | 1028 | 105.01 | 154.69 | 1.91 | moderate elevation, equatorial grassland (south) semi-arid climate regime |
| 4 | 865 | 150.89 | 178.92 | 1.60 | high elevation, subtropical humid climate regime(Himalayan foothills and northeast) |
| 5 | 433 | 48.26 | 79.39 | 2.71 | moderate elevation, semi-arid climate regime(Central India) |
| 6 | 843 | 75.50 | 127.89 | 2.79 | low elevation, northwest and western coastline, arid and warm, humid climate regime (northwest) |
| 7 | 372 | 66.26 | 85.41 | 2.48 | very high elevation, alpine climate regime |

Considering statistical properties, community 4 (Fig.A.3), which covers almost all of the greenest and most mountainous regions of India (northeastern India), has the highest monthly mean (150.89 mm), the largest variation (178.92 mm) and low skewness (1.6) of precipitation in the region (Table A.2). Meanwhile, community 5 (Fig.A.3), covering dry and lowland areas (northwestern India), shows the lowest monthly mean (48.26 mm) with lower variation. Community 6 (western coastline) shows the greatest skewness along with high variability. One possible reason for the high variability and skewness could be that these regions are near to both coastlines and are low-lying areas with two different climate regimes (arid and humid). Community 3 (southeastern India) shows a high coefficient of skewness (1.91) and second high monthly rainfall (105.01mm) and variability (154.69 mm). All the communities show the positive coefficient of skewness, which indicates precipitation with a long tail toward high values.

Community 7 (mountainous region) shows low monthly precipitation mean, moderate variability and high skewness. In South India, both communities 1 and 2 (Fig.A.3) almost have similar rainfall characteristics but are differentiated by topological (elevation, land, coastline) features.

Further, using a node-to-node connection approach (Guimerà and Amaral, 2005; Guimerà et al., 2007) we explore the microscopic details of each individual station within the community. We fit all raingauges of the rainfall network in the Z-P space (Fig.A.4) according to the estimated network measures (Section A.2.4) of the within-module degree ($Z$) and participation coefficient ($P$).

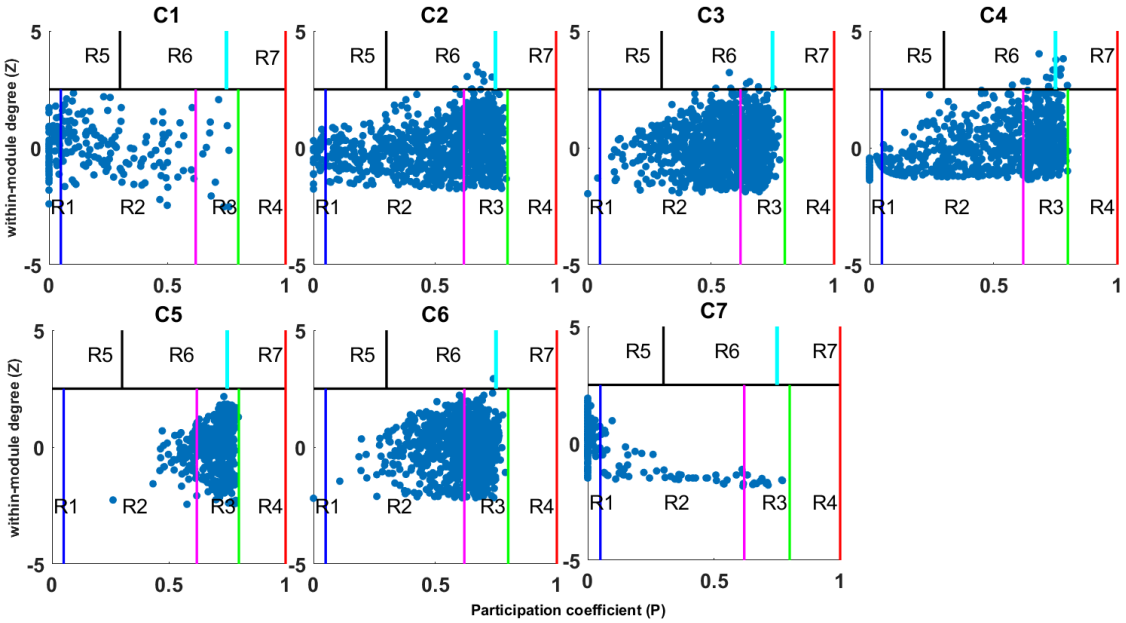



**Figure A.4.:** Role-specific representation of node behavior in the Z-P space (Section A.2.6) plotted for each community (C1 to C7). Within-module degree ($Z$) differentiates between hubs and non-hubs and the participation coefficient ($P$) quantifies the percentage of intra-/inter-community links. Blue colored dots in Z-P space in a particular community represent the raingauge station (node) of that particular community. The significance of each R class is explained in Section A.2.6. Many stations have the same values for $Z$ and $P$ and are thus on top of each other in the different $R$ class.

Figure A.4 shows the Z-P space plot for each community (C1 to C7) separately. Table A.3 shows the percentage of each class of stations in each community. From Fig.A.4 and Table A.3, we find that none of the communities has a kinless node (R4 class node),i.e., no wrongly assigned node. This explains the robustness of the method (Louvain algorithm) used for clustering.

It can be seen that all the communities (C1 to C6) have a dominance of hybrid nodes in their respective community except for community 7, which shows the dominance of nodes with intra-community links. This observation falls along the expected lines, as the Indian sub-continent's precipitation shows the vast variability in topography, climate diversity,etc. The results are quite different from those shown by Agarwal et al., 2017, for German regions. In Germany, the raingauge stations were mostly connected by intra-community links, indicating more homogeneity in the precipitation compared to Indian precipitation.

As explained in Section A.2.6, stations with the almost all ($P \approx 0$) intra-community links can be considered a spatially representative station of the community. We argue that such stations have climatological properties (rainfall time series) that are representative of the other members of their respective communities (Halverson and Fleming 2015).This information has significant importance in big data analysis and uncertainty analysis, as the information from the entire community is available in the form of the representative station.

Further analyzing the Z-P space, we see that the eastern coastline region (C1) to some extent shows good interconnectedness (high number of R1 and R2) and also does not show any hubs (R5 to R7) in the region. This suggests that rainfall in this region is more localized

and does not show any long-range connections. This is in congruence with the general understanding that the eastern coastline region is dominated by the northeastern (NE) monsoonal rainfall while the rest of the country receives rainfall from southwestern (SW) monsoons (Jain et al., 2013).

The mild and moderate-elevation inland regions of India (C2, C3, C5, and C6) show negligible intracommunity links (R1) compared to other high-elevation regions (C4 and C7) and low-elevation regions (C1).These mild and moderate-elevation regions (C2, C3, C5, and C6) are strongly dominated by hybrid stations (R3 and R6) sharing some common dynamics with other regions. For instance, C2 (Southeast) and C3 (Central-East) have very few nodes in the R1 class; the majority of nodes fall in R2 and a significant amount in R3 class stations. This shows that the southeastern and central-eastern regions of the country have short-range and long-range connections. A significant number of R6 class stations reveal that the long-range connections are prevalent over these regions. The ability to detect both short-range and long-range connections is one of the advantages of the complex network approach used in this study, compared to commonly used geostatistical methods which are based on the assumption of a semi-variogram having a decreasing correlation with increasing distance.

Similarly, the western coastline (C6) of India is also dominated equally by R2 and R3 class stations representing short- and long-range connection dynamics in the region. On the contrary, the central-western region (C5) of India is strongly dominated by only R3 class-type stations having a maximum number of links outside the community. This suggests that central-western (C5) regions have no intra-community links to stations. The above observations fall along the expected lines since westerlies enter in India from the West and travel to an entirely different part. Because of a lack of sufficient orographic barriers, we do not see any localized rainfall in this region.

The northeastern region of India (C4) shows a unique kind of pattern, with a significant number of intra-community links, inter-community links, connector hubs and global hubs. This region has a sufficient number of orographic barriers,which helps to accumulate more localized rainfall, represented by short-range connections. Hence, some of the rainfall features in this area are regionally bound and short-range. This region also shows a significant number of inter-community links owing to its long-range connections with the easterlies moisture movement from the C5 regions.

The Himalayan region (C7) shows dominance of R1 class stations, representing a very high degree of interconnectedness in the region. In other words, it suggests that this region receives localized rainfall, having short-range connections. Also, it can be said from the results that this region features a different climatology characterized by seasonal snow and a colder climate than the rest of the regions. Furthermore, it is entirely possible that this region may have connections to regions beyond what is considered in the present study.

From the above analysis, we infer that Z-P space is a useful tool to provide more insight into the qualitative and quantitative connections between the nodes within and outside a community. It also shows the strength of the connections between the communities and is useful in understanding how extreme events in one community affect the other regions. The physical reasoning for the classification of the nodes into seven classes is inline with the general understanding of the precipitation dynamics in India.

**Table A.3.:** Summary of the total number of each type of *R* class stations in the individual community. The significance of each *R* class is described in Section A.2.6.

| C. No. | Explanation of R class | *P*(%)= Percentage of stations in particular R class in each community | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | R1 | R2 | R3 | R4 | R5 | R6 | R7 |
| 1 | Eastern coastline, low-land region having no hubs, mostly dominated by intracommunity links and short-range connections | 33.2 | 61.7 | 5.1 | 0 | 0 | 0 | 0 |
| 2 | Mild-elevation inland region with connector hubs shows the dominance of both intra-community and inter-community links. | 4.3 | 51.6 | 44.1 | 0 | 0 | .9 | 0 |
| 3 | Eastern-central region with moderate elevation shows a lower number of intra-community links to stations. | 0.9 | 59.8 | 39.3 | 0 | 0 | .7 | 0 |
| 4 | Northeastern region of India shows all kinds of connections. Intra-community, inter-community, hubs, non-hubs, global hubs, etc. | 13.0 | 44.7 | 40.7 | 0 | 0 | 1.3 | .5 |
| 5 | No intra-community links, highly dominated by hybrid stations; community shows short-range connections. | 0 | 14.5 | 85.5 | 0 | 0 | 0 | 0 |
| 6 | Negligible intra-community links, dominated by inter-community links and hybrid stations. | .1 | 50.2 | 49.5 | 0 | 0 | .1 | 0 |
| 7 | No hubs, the community has all (ultra-) peripheral nodes that have links within the community, hence well isolated. | 78.5 | 18.3 | 3.2 | 0 | 0 | 0 | 0 |

P(%)=(total stations in any R class of community C/total stations in community C)× 100.

## A.4 Conclusion

This study proposed a novel, complex, network-based approach for quantifying the role of a single (rainfall) station within homogeneous regions, which is of great interest in regionalization studies, estimating missing information, etc. The study used a network information-theoretical approach known as Z-P space for understanding the qualitative and quantitative aspects of the members of a community. The Z-P approach categorizes the members into different classes based on the relative roles they play in the community and their strength of connections within and outside the community. The utility of the method was demonstrated using a synthetic case and then applied to the real-world case of the Indian rainfall network. The entire Indian rainfall network was divided into seven communities, and each community was analyzed using the Z-P approach. The results from the Z-P space approach provided important information such as how the communities are connected within themselves and with others. It was observed that the high-elevation, northern part of India was disconnected from other regions (communities). On the other hand, the southern peninsular region had strong intra-community links as well as inter-community links. It was also observed that the central and eastern parts of the country had many connecter hubs, indicating that these regions have long-range connections with other communities. The stations from the northeastern regions of the country, interestingly, have strong connections

with other communities. The results of the study have significant implication in identifying key node locations in climate systems which play a major role in affecting the climate in the given community.

## A.5 Acknowledgments

This research was funded by the Deutsche Forschungsgemeinschaft (DFG) (GRK 2043/1) within the graduate research training group "Natural risk in a changing world (NatRiskChange)" at the University of Potsdam. The third author acknowledges the research funding from the Humboldt Foundation through Alexander VonHumboldt Fellowship award and Inspire Faculty award, Department of Science and Technology, India for carrying out this research. The authors gratefully thank the Dr. Stephanie Natho (University of Potsdam, Germany) and Roopam Shukla (Department of Energy and Environment, TERI University, New Delhi) for helpful suggestion and reading the paper.

# Appendix B

# Optimal design of hydrometric station networks-based on complex network analysis

*Agarwal, A., Marwan, N., Rathinasamy, M., Ozturk, U., Merz, B. and Kurths, J.: Optimal Design of Hydrometric Station Networks Based on Complex Network Analysis, Hydrol. Earth Syst. Sci. Discuss., 1–21, doi:10.5194/hess-2018-113, 2018.*

Hydrometric networks play a vital role in providing information for decision-making in water resources management. They should be set up optimally to provide as much and as accurate information as possible, and at the same time, be cost-effective. We propose a new measure, based on complex network analysis, to support the design and redesign of hydrometric station networks. The science of complex networks is a relatively young field and has gained significant momentum in the last years in different areas such as brain networks, social networks, technological networks or climate networks. The identification of influential nodes in complex networks is an important field of research. We propose a new node ranking measure, the weighted degree-betweenness, to evaluate the importance of nodes in a network. It is compared to previously proposed measures on synthetic sample networks and then applied to a real-world rain gauge network comprising 1229 stations across Germany to check its applicability in the optimal design of hydrometric networks. The proposed measure is evaluated using the decline rate of network efficiency and the kriging error. The results suggest that it effectively quantifies the importance of rain stations. The new measure is very useful in identifying influential stations which need high attention and expendable stations which can be removed without much loss of information provided by the station network.

## B.1 Introduction

Hydrometric networks monitor a wide range of water quantity and water quality parameters such as precipitation, streamflow, groundwater, or surface water temperature (Keum et al., 2017; Langbein, 1979). Adequate hydrometric monitoring is one of the first and primary tasks towards efficient water resources management. Information from hydrometric stations plays a crucial role in, among other things, flood estimation, water budget analysis, hydraulic design and assessing climate change. Even after the advent of remote sensing

based information, such as precipitation products, in-situ observations are considered as an essential source of information on hydrometeorology.

The basic characteristics of hydrometric networks comprise the number of stations, their locations, observation periods and sampling frequency (Keum et al., 2017). The general understanding is that the higher the number of monitoring stations, the more reliable the quantification of areal average estimates and point estimates at any ungauged location. However, a higher station number increases the cost of installation, operation, and maintenance, but may provide redundant information and, therefore, not increase the information content obtained from the network. Globally, there is a decreasing trend in the number of hydrometric stations in the last decades (Mishra and Coulibaly, 2009). Against the background of shrinking monetary support for hydrometric networks, their optimal design is gaining importance.

The design of hydrometric networks is a well-identified problem in hydrometeorology and has received considerable attention (Mishra and Coulibaly, 2009). For example, Putthividhya and Tanaka (2012) made an effort to design an optimal rain gauge network based on the station redundancy and the homogeneity of the rainfall distribution. Adhikary et al. (2015) proposed a kriging based geostatistical approach for optimizing rainfall networks, and Chacon-Hurtado et al. (2017) provided a generalized procedure for optimal rainfall and streamflow monitoring in the context of rainfall-runoff modelling. Yeh et al. (2006) optimized a rain gauge network applying the entropy method on radar datasets. Several approaches have been developed for optimal network design, such as statistical analysis which include variance and dimension reduction methods (Amorim et al., 2012; Wadoux et al., 2017), spatial interpolation which includes kriging methods (Adhikary et al., 2015) and various interpolation techniques (Kassim and Kottegoda, 1991), information theory-based methods (Taormina et al., 2015; Stosic et al., 2017), optimization techniques such as simulated annealing (Barca et al., 2008; Mishra and Coulibaly, 2009; Pardo-Igúzquiza, 1998; Yoo et al., 2003), physiographic analysis (Laize, 2004), multivariate factor analysis (Hargrove and Hoffman, 2004; Yeh et al., 2006), sampling strategies (Tsintikidis et al., 2002), and user surveys or expert recommendations (Rani and Moreira, 2010). Combinations of methods have also been introduced in the last decade (Chacon-Hurtado et al., 2017; Keum et al., 2017; Mishra and Coulibaly, 2009; Sehgal et al., 2016).

Most of these studies inherently assume that a more optimal network is achieved through expanding the network with supplementary stations. However, a higher number of stations does not necessarily decrease the uncertainty (Stosic et al., 2017; Ozturk et al., 2018c). Hence, a network can be optimized either by eliminating expendable stations from the network to minimize the cost or by expanding the network with the installation of additional stations to reduce the estimation uncertainty. Mishra and Coulibaly (2009) argued that the expendable stations in a network that contribute little or even nothing should be identified and removed, and at the same time, the most valuable or influential stations should be maintained and protected.

In the recent past, the science of complex networks has gained great momentum and have attracted many researchers from different disciplines and application fields, e.g., transportation networks (Bell and Lida, 1997), power grid analysis (Menck et al., 2014; Schultz et al., 2014), streamflow networks (Halverson and Fleming, 2015) and climate networks (Agarwal et al., 2018b; Ozturk et al., 2018a; Paluš, 2018). A complex network is a collection

of nodes interconnected with links in a non-trivial manner. In a functional network, links are set up between each pair of nodes based on how the nodes interact with each other. The investigation of the topology of such complex networks helps in better understanding the mechanisms of highly complex systems with many components (Donges et al., 2009b).

Motivated by the encouraging results obtained by the above mentioned studies on climate networks, we develop a node ranking measure, based on complex network analysis that can be used to identify influential and expendable stations in large hydrometric station networks. Our aim is not to question the credibility of operating stations, but to propose an alternative evaluation procedure for the optimal design and redesign of observational networks.

In section B.2, we introduce the basic concepts of complex networks. The proposed node ranking measure is presented and compared with existing measures in section B.3 using synthetic networks. In section B.4, the new measure is applied to a rain gauge network consisting of 1229 stations across Germany and compared with state-of-the-art methods.

## B.2 Basics of complex networks

### B.2.1. Network construction

A network or a graph is a collection of entities (nodes, vertices) interconnected with lines (links, edges) as shown in Fig.B.1. These entities could be anything, such as humans defining a social network (Arenas et al., 2008), computers constructing a web network (Zlatić et al., 2006), neurons forming brain networks (Bullmore and Sporns, 2012; Pfurtscheller and Lopes da Silva, 1999; Rubinov and Sporns, 2011), streamflow stations creating a hydrological network (Halverson and Fleming, 2015; Sivakumar and Woldemeskel, 2014) or climate stations describing a climate network (Donges et al., 2009a; Gozolchiani et al., 2008; Malik et al., 2012; Boers et al., 2014; Boers et al., 2015; Agarwal et al., 2017).

Formally, a network or graph is defined as an ordered pair $G = \{N, E\}$, containing a set of nodes $N = \{N_1, N_2, \ldots N_N\}$ together with a set $E$ of edges $\{i, j\}$ which are 2-element subsets of N. In this work, we consider undirected and unweighted graph ($G$), where only one edge can exist between a pair of nodes and self-loops of the type $\{i, i\}$ are not allowed. This type of graph can be represented by the symmetric adjacency matrix (Stolbova et al., 2014)

$$A_{i,j} = \begin{cases} 1, & \{i, j\} \in E \\ 0, & \{i, j\} \notin E. \end{cases} \tag{B.1}$$

Figure B.1 is a simple representation of such a network, i.e., one with a set of identical nodes connected by identical links. In general, (large) graphs of real-world entities with irregular topology are called complex networks. The links represent similar evolution or variability at different nodes and can be identified from data using a similarity measure such as correlation (Donges et al., 2011; Jha et al., 2015), synchronization (Conticello et al., 2018; Stolbova et al., 2014; Ozturk et al., 2018a) or mutual information (Paluš, 2018).

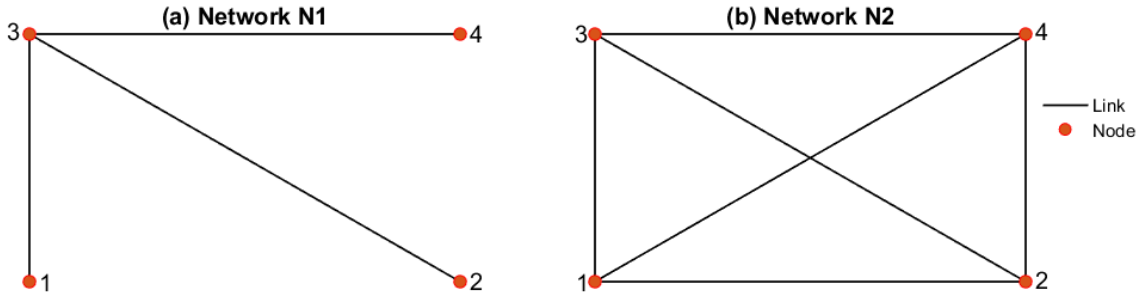



**Figure B.1.:** Topology of two sample networks to explain network structures and measures. (a) Network N1 with four nodes and three links; (b) network N2 with four nodes and six links.

### B.2.2. Node ranking measures

A large number of measures have been defined to characterize the behaviour of complex networks. We focus here on those measures which have been proposed to quantify the importance of nodes in a network: degree $k$, betweenness centrality $B$ (Stolbova et al., 2014), bridgeness $Bri$ (Jensen et al., 2016) and degree and influence of line $DIL$ (Liu et al., 2016).

The degree $k$ of a node in a network counts the number of connections linked to the node directly. For example, the degree of nodes 1, 2 and 4 in network N1 (Fig.B.1a) is 1 and for node 3 is 3. In the network N2 (Fig.B.1b), all nodes have degree 3. The degree can explain the importance of nodes to some extent, but nodes that own the same degree may not play the same role in a network. For instance, a bridging node connecting two important nodes might be very relevant though its degree could be much lower than the value of less important nodes.

The betweenness centrality B is a measure of control that a particular node exerts over the interaction between the remaining nodes. In simple words, $B$ describes the ability of nodes to control the information flow in networks. To calculate betweenness centrality, we consider every pair of nodes and count how many times a third node can interrupt the shortest paths between the selected node pair. In network N1, $B$ of node 3 is 3, i.e., node 3 can disturb three pairs $1-2, 1-4, 2-4$, and for other nodes $B=0$. In the network N2, all nodes have $B=0$ because no node can interrupt the information flow. So node 3 is a critical node in the network N1 but not in the network N2.

Jensen et al. (2016) developed the Bridgeness measure $Bri$ to distinguish local centres, i.e. nodes that are central to a part of the network, from hybrid nodes, i.e. nodes that connect different parts of a network. $Bri$ is a decomposition of betweenness centrality $B$ into a local and a global contribution. Therefore, the $Bri$ value of a node $i$ is always smaller or equal to the corresponding $B$ value and they only differ by the local contribution of the first neighbours. To calculate $Bri$ we consider the shortest path between nodes not in the neighbourhood of node $i$, $N_G(i)$ (Table B.1). For example, in the networks N1 and N2, all nodes have $B=0$, hence $Bri=0$, except node 3 in the network N1 for which all the nodes are in direct neighbourhood hence it also has $Bri=0$.

The degree and influence of line (DIL), introduced by Liu et al. (2016), considers the node degree $k$ and importance of line I to rank the nodes in network. The mathematical equation of DIL measure is presented in table B.1, where the line between node $i$ and $j$ is $e_{ij}$ and its importance is defined as $I_{e_{ij}} = \frac{U}{\lambda}$ where $U = (k_i - p - 1)(k_j - p - 1)$ reflects the

connectivity ability of a line (link), p is the number of triangles having one edge $e_{ij}$ and $\lambda = \frac{p}{2} + 1$ is defined as an alternative index of line $e_{ij}$ . $N_G(i)$ is the set of neighbors of node i (for detailed explanation refer Liu et al. (2016)). DIL equation (Table B.1) suggests that all the nodes having $k_i = 1$, will have $DIL_i = 1$, since the second term of the equation will be zero. Hence, in the network N1 all nodes, except node 3, have $DIL = 1$. Node 3 has $DIL = 3$ equal to its degree, since the second term is zero (all the connected nodes 1, 2 and 4 have $k_j = 1$, hence $I_{e_{ij}} = 0$.

**Table B.1.:** Network measures. $N$:total number of nodes in a network. $k_i$: Degree of node $i$. $B_i$: Betweenness of node $i$, where $\sigma(j,k)$ represents the number of links along the shortest path between $j$ and $k$; while $\sigma_i(j,k)$ is the number of links of the shortest path running through node $i$. In bridgeness, we consider the shortest path between nodes not in the neighbourhood of node $i$ ($N_G(i)$).

| Degree | Betweenness centrality | Bridgeness | DIL |
|---|---|---|---|
| $k_i = \frac{\Sigma_{j=1}^{N} A_{i,j}}{N-1}$ | $B_i = \Sigma_{j \neq k \neq i \notin V}^{N} \frac{\sigma_i(j,k)}{\sigma(j,k)}$ | $Bri_i = \Sigma_{j \notin N_G(i) v k \notin N_G(i)}^{N} \frac{\sigma_i(j,k)}{\sigma(j,k)}$ | $DIL_i = k_i + \Sigma_{j=N_G(i)} I_{e_{ij}} \frac{k_i - 1}{k_i + k_j - 2}$ |

## B.3 Methodology

We propose a new node ranking measure that we call weighted degree-betweenness (WDB). We further compare the efficacy of the proposed measure with the existing node ranking methods using two synthetic networks.

### B.3.1. Weighted degree-betweenness

WDB is a combination of the network measures degree and betweenness centrality. We define WDB of a particular node i as the sum of the betweenness centrality of node i and all directly connected nodes $j$, $j = 1,2,3...n$ in proportion to their contribution to node $i$. Mathematically, the WDB of a node $i$ is given by

$$WDB_i = B_i + I_i \tag{B.2}$$

where $B_i$ is the betweenness centrality of node $i$, and $I_i$ stands for the influence or contribution of the directly connected node $j$, $j = 1,2,3...n$ to node $i$. It is defined for node $i$ as

$$I_i = \sum_{j=1}^{n} \frac{B_j * (k_j - 1)}{(k_i + k_j - 2)} \tag{B.3}$$

where $k_i$ is the degree of node $i$, $k_j$ is the degree of the nodes $j$ which are directly connected to node $i_i$, and $n$ is the total number of directly connected nodes to node $i$.

### B.3.2. Comparison with existing node ranking measures using synthetic networks

In this section, we motivate the development of the new node ranking measure WDB by comparing it to existing measures. Identifying nodes that occupy interesting positions in a real-world network using node ranking helps to extract meaningful information from large datasets with little cost. Usually, the measures degree or betweenness centrality are used for node ranking (C. Gao et al., 2013; Okamoto et al., 2008; Saxena et al., 2016; Agarwal

et al., 2019). However, these measures have certain disadvantages which are explained using a simple network, the undirected and unweighted network $Z = (N, E)$ with 8 nodes and 11 edges shown in Figure B.2. The network measures $k_i$, $B_i$ and $WDB_i$ of each node are given in Table B.2.

Degree is limited as node ranking measure since it cannot distinguish between different roles in the network. For example, nodes 5, 7, and 8 have the same degree ($k_i = 2$), but node 5 serves as bridge node linking the two parts of the network. Information between several nodes in this network can flow through this node only. In a large complex network, such nodes have strategic relevance as most of the information can be accessed quickly just by capturing those nodes. For example, in a social network, the spreading of a disease could be slowed down or hindered by identifying these nodes. In climate networks, an early warning signal could be generated by capturing the flow of information (Hlinka et al., 2017; Donges et al., 2009b).

Betweenness centrality has a higher power in discriminating different roles. For example, nodes 4 and 5 have the highest betweenness centrality $B = 24$ followed by node 6. Their importance for the information flow in the network is obvious, as such high $B$ nodes can be used to control the flow of information in any network. However, betweenness $B$ gives equal scores to local centers (nodes $4, 6$), i.e., nodes of high degree central to a single region, and to global bridges (node 5), which connect different regions. This distinction is important because the roles of these nodes are different. For example, in climate networks, local centers correspond to nodes which are important for local climate phenomena, while bridges correspond to nodes which connect different subsystems of climate, such as monsoon and El Niño/Southern Oscillation, leading to teleconnections (Paluš, 2018).

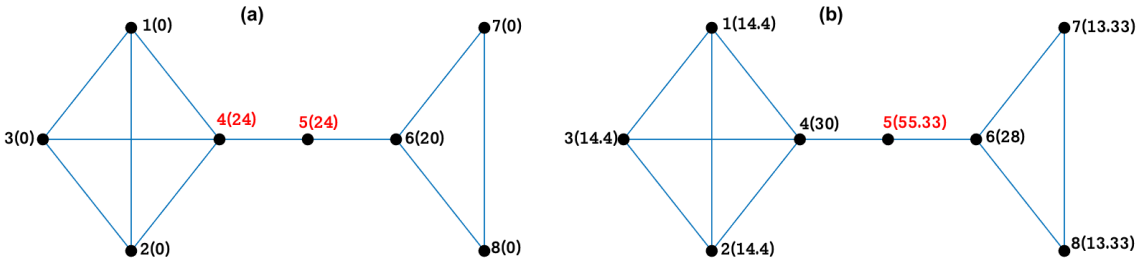



**Figure B.2.:** Synthetic network to explain the betweenness ($B$) and weighted degree-betweenness (WDB) measures, with node number (1 to 8) followed by the betweenness value (a) and WDB value (b) in brackets. Betweenness does not distinguish centers from bridges (marked in red), as it attributes the same value to the local center (node 4) and to the global bridge node (node 5). In contrast, WDB assigns the highest importance to node 5 that plays the role of a global bridge. Further, betweenness does not differentiate between the nodes 1, 2, 3, 7 and 8, while WDB provides a nuanced picture of the influence of all nodes.

The proposed measure WDB has an even higher discrimination power compared to betweenness centrality and effectively ranks the nodes in the network. Node 5 has the highest WDB score and is ranked as the most influential node. This reflects its role as global bridge node, as losing node 5 would disconnect the two parts of the network. WDB is also able to distinguish between the nodes $1, 2, 3$ ($WDB = 14.4$) and the nodes $7, 8$ ($WDB = 13.3$) which is important in case of we need sequentially ranking of nodes.

**Table B.2.:** Network measures for the synthetic network shown in Figure B.2. Ranking of the nodes is based on the proposed WDB measure.

| Node no. | $k_i$ | $B_i$ | $WDB_i$ | Rank | Role of nodes |
|---|---|---|---|---|---|
| 1 | 3 | 0 | 14.4 | 4 | node |
| 2 | 3 | 0 | 14.4 | 4 | node |
| 3 | 3 | 0 | 14.4 | 4 | node |
| 4 | 4 | 24 | 30 | 2 | local center |
| 5 | 2 | 24 | 55.3 | 1 | global bridge |
| 6 | 3 | 20 | 28 | 3 | local center |
| 7 | 2 | 0 | 13.3 | 5 | node |
| 8 | 2 | 0 | 13.3 | 5 | node |

To further evaluate the proposed measure, we compare WDB with other network measures recently published, namely the bridgeness developed by Jensen et al. (2016) and degree and influence of line DIL by Liu et al. (2016). For this comparison, we use the same synthetic network as Jensen et al. (2016) shown in Figure B.3. The corresponding network measure values are given in Table B.3.

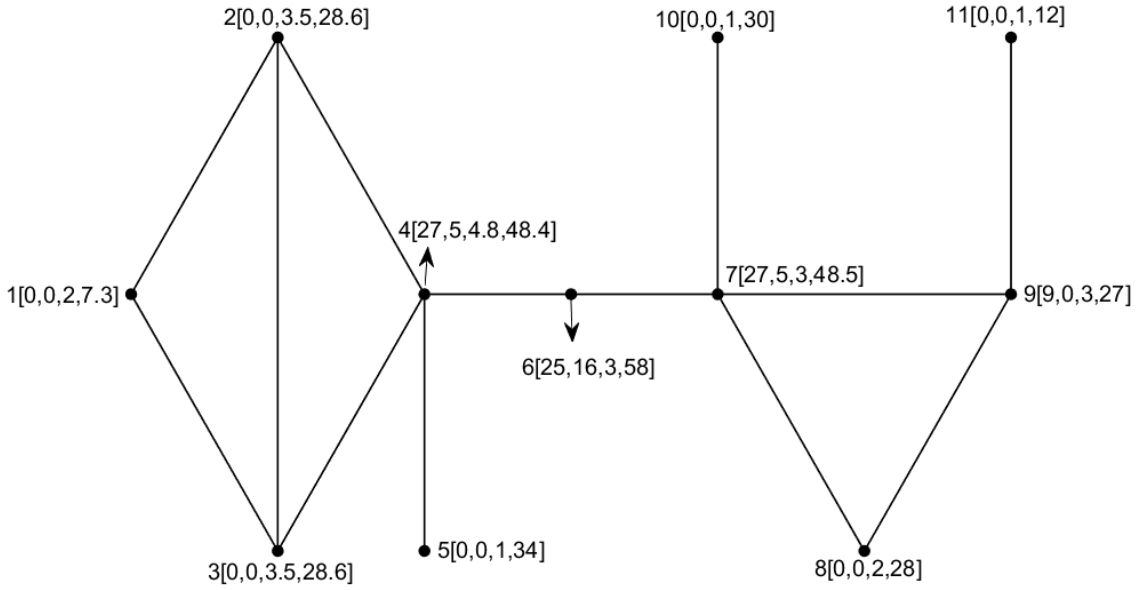



**Figure B.3.:** Synthetic network used to compare the network measures betweenness, bridgeness, and DIL with the proposed measure WDB. Number 1 to 11 are node counts, and values in brackets represent the network measure values in order of $[B, Bri, DIL, and WDB]$.

Fig.B.3 illustrates that betweenness does not distinguish between the local centers (nodes 4, 7) and the global bridge node (node 6). It even assigns a smaller value to the global bridge node. Bridgeness expresses the higher importance of the global bridge node compared to local centers, however, it does not distinguish between all other nodes. Although DIL assigns different values to almost very node, these numbers do not represent the different roles of the nodes and are therefore hardly suitable as node ranking measure. WDB outperforms the existing measures in effectively ranking nodes in the network, such as the global bridge nodes, local centers and dead-end nodes (nodes 5, 10, 11) by assigning highest to lowest

values. For example, WDB is sensitive to rank nodes 4 and 7 for which the bridgeness measure provides equally score.

This comparison of the proposed measure WDB with other measures that have been developed to express the importance of nodes within a network shows that WDB is able to provide a nuanced picture. The resulting node ranking reflects the different roles, such as global bridge, local center, dead-end node, hub (high degree), non-hub (low degree) (shown in Table B.3) of the individual nodes.

**Table B.3.:** Measures for the synthetic network shown in Figure B.3. Ranking of the nodes is based on the proposed WDB measure.

| Node no. | $B_i$ | $Bri_i$ | $DIL_i$ | $WDB_i$ | Rank | Role of the node |
|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 2 | 7·3 | 10 | non-hub node |
| 2 | 0 | 0 | 3·5 | 28.6 | 6 | non-hub node |
| 3 | 0 | 0 | 3·5 | 28.6 | 6 | non-hub node |
| 4 | 27 | 5 | 4.8 | 48.4 | 3 | hub node |
| 5 | 0 | 0 | 3·5 | 28.6 | 4 | dead-end node |
| 6 | 25 | 16 | 3 | 58 | 1 | global bridge |
| 7 | 27 | 5 | 3 | 48.5 | 2 | hub node |
| 8 | 0 | 0 | 2 | 28 | 7 | non-hub node |
| 9 | 9 | 0 | 3 | 27 | 8 | non-hub node |
| 10 | 0 | 0 | 1 | 30 | 5 | dead-end node |
| 11 | 0 | 0 | 1 | 12 | 9 | dead-end node |

### B.3.3. Evaluation of the proposed measure for a rain gauge network

In the context of hydrometric station networks, we hypothesise that higher ranking stations are more influential nodes in the network. Loosing such stations would more strongly reduce the network efficiency, i.e., the flow of information within the network, compared to lower ranking stations. Stations with the lowest ranks in the network are the least influential and are seen as expendable stations. To test this hypothesis, we apply the proposed node ranking measure to a rain gauge network consisting of more than 1000 stations in Germany. The information loss caused by removing stations is quantified via two measures: (a) decline rate of network efficiency, and (b) relative kriging error.

**Rate of network efficiency** The decline rate of network efficiency, as proposed by Liu et al. (2016), quantifies the loss in efficiency with which information flows within a network when nodes are removed from the network. Network efficiency is defined as

$$\eta = \frac{1}{N(N-1)} \sum_{n_i \neq n_j} \eta_{ij} \tag{B.4}$$

where $\eta_i j$ is the efficiency between nodes $n_i$ and $n_j$. $\eta_{ij}$ is inversely related to the shortest path length: $\eta_{ij} = \frac{1}{d_{ij}}$, where $d_{ij}$ is the shortest path between nodes $n_i$ and $n_j$. The average path length $L$ measures the average number of links along the shortest paths between all possible pairs of network nodes. It is a measure of the efficiency of information or mass transport in a network. A network with small $L$ is highly efficient, because two nodes are

likely to be separated by a few links only. The decline rate of network efficiency $\mu$ is defined as

$$\mu = 1 - \frac{\eta_{new}}{\eta_{old}} \tag{B.5}$$

where $\eta_{new}$ is the efficiency of the network after removing nodes, and $\eta_{old}$ is the efficiency of the complete network.

We hypothesise that the network efficiency reduces more strongly if high ranking stations are removed. This implies higher decline rates of efficiency when removing higher ranking stations from the network.

**Relative kriging error** As second measure to evaluate the information loss when stations are removed from the network, we use a kriging based geostatistical approach (Adhikary et al., 2015; Keum et al., 2017). Kriging is an optimal surface interpolation technique assuming that the variance in a sample of observations depends on their distance (Adhikary et al., 2015). It is the best linear unbiased estimator of unknown variable values at unsampled locations in space where no measurements are available, based on the known sampling values from the surrounding areas (Hohn, 1991; Webster and Oliver, 2007). The Ordinary Kriging technique is used in this study for interpolating rainfall data and estimating the kriging error. The kriging estimator is expressed as

$$Z^*(x_o) = \sum_{i=1}^{n} w_i Z(x_i) \tag{B.6}$$

where $Z^*(x_o)$ refers to the estimated value of $Z$ at the desired location $x_0$; $w_i$ represents weights associated with the observation at the location $x_i$ with respect to $x_0$; and n indicates the number of observations within the domain of the search neighborhood of $x_o$ for performing the estimation of $Z^*(x_o)$. Ordinary Kriging is implemented through ArcGISv10.4.1 (Redlands, CA, USA) (ESRI, 2009) and its geostatistical analyst extension.

To evaluate the performance of the WDB measure in identifying influential and expandable stations in a large network, we calculate the increase in the interpolation error of when stations are removed. The relative kriging error before and after removing the stations is denoted as

$$\Re(\%) = \frac{KSE_{new} - KSE_{old}}{KSE_{old}} \times 100 \tag{B.7}$$

where $KSE_{new}$ denotes the standard kriging error after removing stations, and $KSE_{old}$ is the error for the original network. We hypothesize that the relative kriging error is higher when removing high ranking stations. To cover a broad range of rainfall characteristics, the error is calculated for different statistics, i.e. the mean, $90th$, $95th$ and $99th$ percentile rainfall and the number of wet days (precipitation>2.5mm).

## B.4 Application to an extensive rain gauge network

### B.4.1. Rainfall data

To evaluate the proposed measure in the context of the optimal design of hydrometric networks, we apply it to an extensive network of rain stations in Germany and adjacent areas (Figure B.4). The data covers 110 years at daily resolution (1 January 1901 to 31 December 2010). The 1229 rain stations inside Germany (red dots in Fig.B.4) are operated by the German Weather Service. Data processing and quality control were performed according to Oesterle (2001). 211 stations from different sources outside Germany (green dots in Fig.B.4) were included in the analysis to minimize spatial boundary effects in the network construction, however, these stations were excluded from the node ranking analysis.

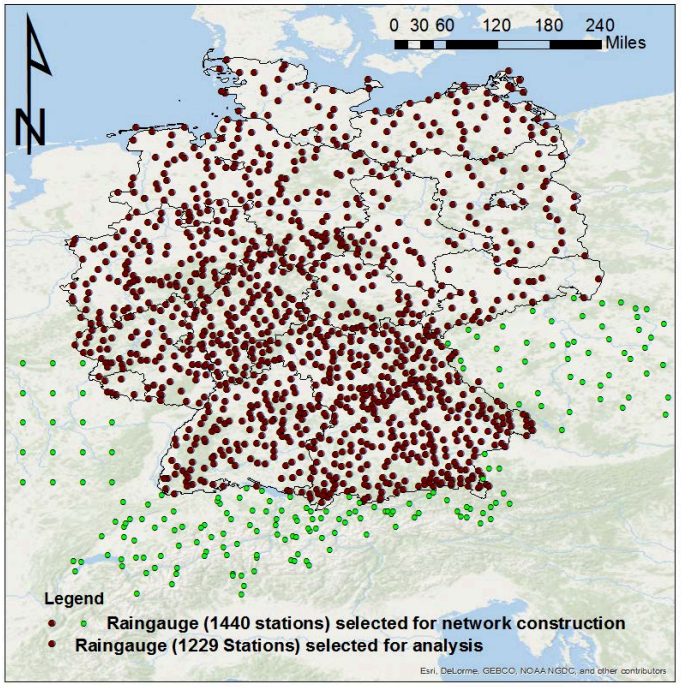



**Figure B.4.:** Location of rain stations in Germany and adjacent areas. Red dots indicate stations lying inside Germany that are used in the analysis. Green dots indicate stations outside of Germany that are used for network construction only to minimize the boundary effect.

### B.4.2. Network construction

We begin the network construction by extracting event time series from the 1440 daily rainfall time series. The event series represent heavy rainfall events, i.e., precipitation exceeding the $95th$ percentile at that station (Rheinwalt et al., 2016). The $95th$ percentile is a good compromise between having a sufficient number of rainfall events at each location and a rather high threshold to study heavy precipitation. All rainfall event series are compared with each other using event synchronization. This results in the similarity matrix $Q$, whereas the entry at index pair (i,j) defines synchronization in the occurrence of heavy rainfall events at station $i$ and station $j$. Applying a certain threshold to the $Q$ matrix

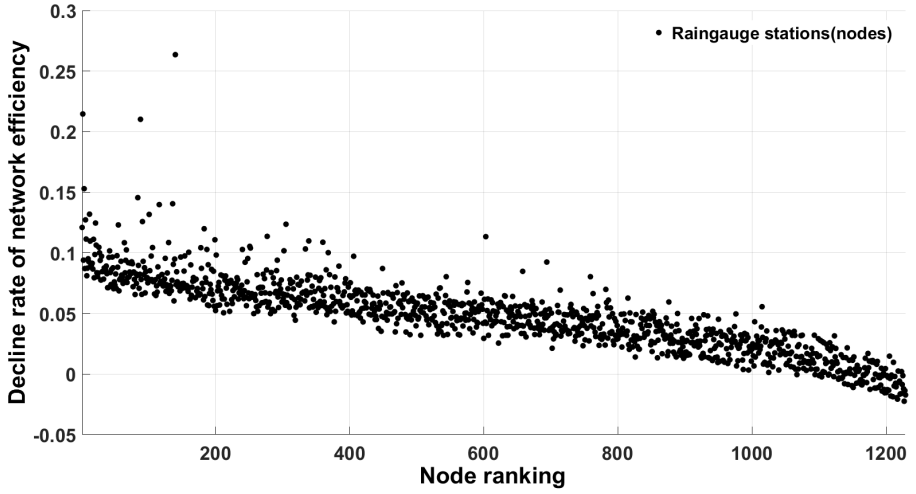



**Figure B.5.:** Decline rate of network efficiency corresponding to the removal of each node in the rainfall network. In each implementation, only one node is removed from the network according to the ranking with replacement (bootstrapping).

yields an adjacency matrix

$$A_{i,j} = \begin{cases} 1, & \text{if } Q_{i,j} \geq \theta^{Q}_{i,j}. \\ 0, & \text{otherwise.} \end{cases} \tag{B.8}$$

Here, $\theta^{Q}_{i,j}$= $95th$ percentile is a chosen threshold, and $A_{i,j} = 1$ denotes a link between the $i^{th}$ and $j^{th}$ nodes and 0 denotes otherwise.The adjacency matrix represents the connections in the rainfall network. Although the constructed network is based on all 1440 stations (to minimize the boundary effect), the subsequent topological analysis is performed only for the 1229 stations lying inside Germany.

## B.4.3. Decline rate of network efficiency

In this section, we evaluate the ranking of stations derived from the proposed WDB measure using the decline rate of network efficiency. The rain gauges are ranked in decreasing order according to their WDB values. Highly ranked rain gauges are interpreted as the most influential stations, and low ranked as expendable stations.

Firstly, we analyze the decline rate of network efficiency $\mu$ when one station is removed from the network. In each trial, we remove only one station (starting with the highest rank). After $n = 1229$ (number of nodes) trials, we investigate the relationship between $\mu$ and the node ranking measured by WDB. We expect an inverse relationship between $\mu$ and WDB: the higher the node ranking, the more important is that node, leading to a higher loss in network efficiency. Fig.B.5 confirms this behavior. $\mu$ is high for high-ranking stations and decays with node ranking. Interestingly, $\mu < 0$ for very low ranking stations, i.e. the network efficiency increases when single, low ranking stations are removed. This is explained by the decrease of the redundancy in the network when such stations are removed. Secondly, we remove successively a larger number of stations, from 1 to 123 stations (10%), considering three cases. In case I, we remove up to the 10% highest ranking stations. This implies that in the first iteration we remove the top-ranked station and in the second iteration we

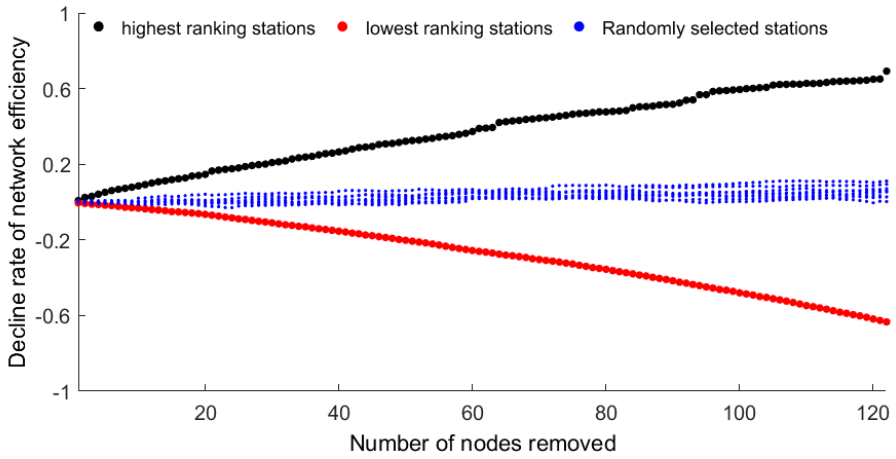



**Figure B.6.:** Decline rate of network efficiency as a function of the number of stations removed from the network. Case I: up to the 10% highest ranking stations are removed (black), case II: up to the 10% lowest ranking stations are removed (red), case III: up to 10% randomly drawn stations are removed (10 trials) (blue).

remove the top two stations and so on. Fig.B.6 shows a clear increase in $\mu$ when more and more influential stations are removed. In case II, up to the 10% lowest ranking stations are successively removed. It can be seen in Fig.B.6 that this affects the network efficiency in a positive way: The efficiency increases when the lowest ranking stations are removed. In case III, up to 10% stations are randomly removed. Case III is repeated ten times to understand the effect of random sampling. In general, $\mu$ increases with removing random stations. However, the effect is much lower (in absolute terms) compared to the effect of removing high or low ranking stations, respectively. The variation in $\mu$ between the ten trials and within one trial is caused by randomness. For example, $\mu$ rises instantaneously when the algorithm picks up a high ranking station.

### B.4.4. Relative kriging error

As the second approach to assess the suitability of WDB for identifying influential and expandable stations, we analyse the change in the kriging error when stations are removed from the network. Similarly to the evaluation using the decline rate of network efficiency in section B.4.3, three cases are investigated: removing the 10% highest ranking stations, removing the 10% lowest ranking stations, and ten trials of removing 10% of the stations randomly. The change in the kriging error is calculated for five characteristics, i.e., mean, $90\%-, 95\%-, 99\%-$percentile, and number of wet days (Table B.4).

Removing the 10% high-ranking stations (case I) leads to positive and high ($\Re > 5\%$) values for all five statistics considered. The kriging error increases substantially when these stations are removed. When the 10% lowest ranking stations (case II) are not considered, the $\Re$ values are small compared to those obtained by removing high ranking stations. The relative errors in estimating the mean, percentile rainfall characteristics ($90th$ and $95th$) and number of wet days at ungauged locations are low ($< 5\%$) for the 10% lowest ranking stations, suggesting that these stations do not contribute much information. For two out of five statistics, i.e., mean and number of wet days, removing the 10% lowest ranking stations actually improves the kriging model. Case III, i.e. removing stations randomly,

shows positive and high ($\Re > 5\%$) values except for the number of wet days because by doing so we remove high ranking nodes as well which lead to higher rates of $\Re(\%)$.

**Table B.4.:** Relative kriging error for the three different cases. The relative kriging error for case III is the average across ten trials. Stars indicate a high relative error $> 5\%$.

| Case | Removal of stations | Relative kriging error $\Re(\%)$ | | | | |
|---|---|---|---|---|---|---|
| | | Mean | 90*th* percentile | 95*th* percentile | 99*th* percentile | Wet days |
| I | 10% highest ranking | 7.2* | 35.9* | 72.3* | 59.1* | 71.1* |
| II | 10% lowest ranking | -5.1 | 4.1 | 3.8 | 11.4* | -1.5 |
| III | 10% randomly selected | 6.6% | 28.5* | 60.8* | 51.9* | 4.8 |

## B.5 Discussion

Building on the young science of complex networks, a novel node ranking measure, the weighted degree-betweenness WDB, is proposed. It is based on degree and betweenness centrality measure of nodes in a network. The comparison of the WDB measure with the existing node ranking measures suggests that it is more informative since it is better able to consider the different role of nodes in a complex network. The WDB measure provides a unique value to each node depending on its importance and influence in the network.

Further, this study proposes to use WDB for supporting the optimal design of large hydrometric networks. It is able to rank the nodes in a large network in relation to their importance for the flow of information, mass or energy. This ranking can be used to identify highly influential and expandable hydrometric stations. For example, removing low ranking stations in the German rain gauge network increases the network efficiency considerably, and may even decrease the error of estimating rainfall at ungauged locations. This is explained by the redundancy in the information that those stations provide, which in turn is attributed to the similarity between the gauges due to the common driving mechanisms or spatial similarity as advocated by Tobler's Law of Geography (Tobler, 1970). The results of our analysis suggests that WDB identifies the expandable nodes correctly as shown by decline rate of efficiency and relative kriging error. On the other hand, awards stations which provide unique information which cannot be generated from other stations in the network. Based on our analysis, we argue that, ranking of all nodes in large networks has the major benefit that the new measure adds to the optimal design of hydrometric networks or redesign of existing hydrometric networks.

The proposed node ranking approach differs from the existing approaches as it considers different aspects of the spatio-temporal relationships in observation networks. This measure is not only useful for optimizing observational networks, but has also potential to support the selection of an optimal number of stations for the prediction in ungauged basins (PUBs) and estimating missing values by identifying influential stations in the region. For example, from a set of N potential stations to be used for PUBs the proposed approach can be applied to select the $M$ influential stations which, when used, reduce the uncertainty (Villarini et al., 2008). Similarly, the proposed method can be applied to gridded satellite data (rainfall, soil moisture), to locate the strategic points where stations should be installed to ensure a highly efficient observation network. For instance, identifying influential grid points in the network of satellite data (rainfall, soil-moisture) will guide where to install monitoring

stations.

An advantage of the proposed method is its capability to differentiate between the different roles played by individual stations. For example, global bridge nodes are able to control the flow of information, energy or mass between different parts of a network. Hence, they are of highest importance. This capability opens new possibilities for its use in complex networks. For instance, in climate networks an early warning signal could be generated by capturing the flow of information at such points (Donges et al., 2009b; Hlinka et al., 2017; Stolbova et al., 2016).

## B.6 Conclusions

This study proposes a novel node ranking measure for identifying the influential and expendable nodes in a complex network. The new network measure weighted degree-betweenness (WDB) combines the existing measures degree and betweenness centrality and considers the neighbourhood of a node. The proposed measure is compared to other measures using synthetic networks. WDB is more sensitive to the different roles of nodes, such as global connecting nodes, hybrid nodes, and local centers, and provides a more informative ranking than the existing node ranking measures.

We propose to use this measure for the optimal design of hydrometric networks. Applying this measure to a network of 1229 rain gauges in Germany allows identifying influential and expendable stations. Two criteria, the decline rate of network efficiency and the kriging error, are used to evaluate the performance of the proposed node ranking measure. The results suggest that the proposed measure is indeed capable of effectively ranking the stations in large hydrometric networks.

We argue that the proposed measure is not only useful for optimizing observational networks, but has the potential to support the selection of an optimal number of stations (by determining influential station of the region) to be used in the prediction in ungauged basins, or to support the estimation of missing values, regionalization, and regional flood frequency analysis. When applied to gridded satellite data, it can be used to locate the strategic points where stations should be installed to ensure a highly efficient network. Furthermore, the new network measure has large potentials in other fields where science complex networks are used, such as in social networks, infrastructure networks, disease spreading networks, and brain networks.

## B.7 Data availability

The precipitation data was provided by the German Weather Service. The data is publicly accessible at https://opendata.dwd.de/. The data was preprocessed by the Potsdam Institute for Climate Impact Research (Conradt et al., 2012).

## B.8 Acknowledgments

This research was funded by Deutsche Forschungsgemeinschaft (DFG) (GRK 2043/1) within the graduate research training group Natural risk in a changing world (NatRiskChange)

at the University of Potsdam urlhttp://www.uni-potsdam.de/natriskchange. Also, we gratefully acknowledge the provision of precipitation data by the German Weather Service.

# Appendix C

# Supporting figures for:wavelet-based multi-scale similarity measure for complex networks

## C.1 Orthographic projection with center at North pole

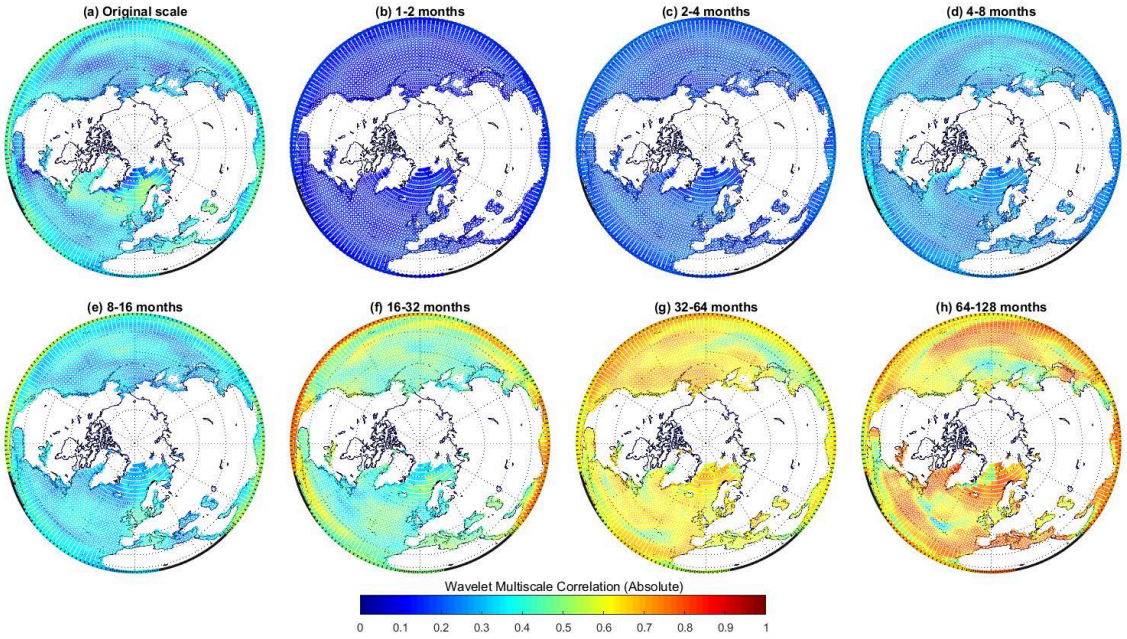



**Figure C.1.:** $95th$ percentile of absolute wavelet multiscale correlation (WMC) values for the SST network constructed on multiscale using the PCC (original scale) and WMC at scales 1-7. Color represents the strength of correlation on each scale.

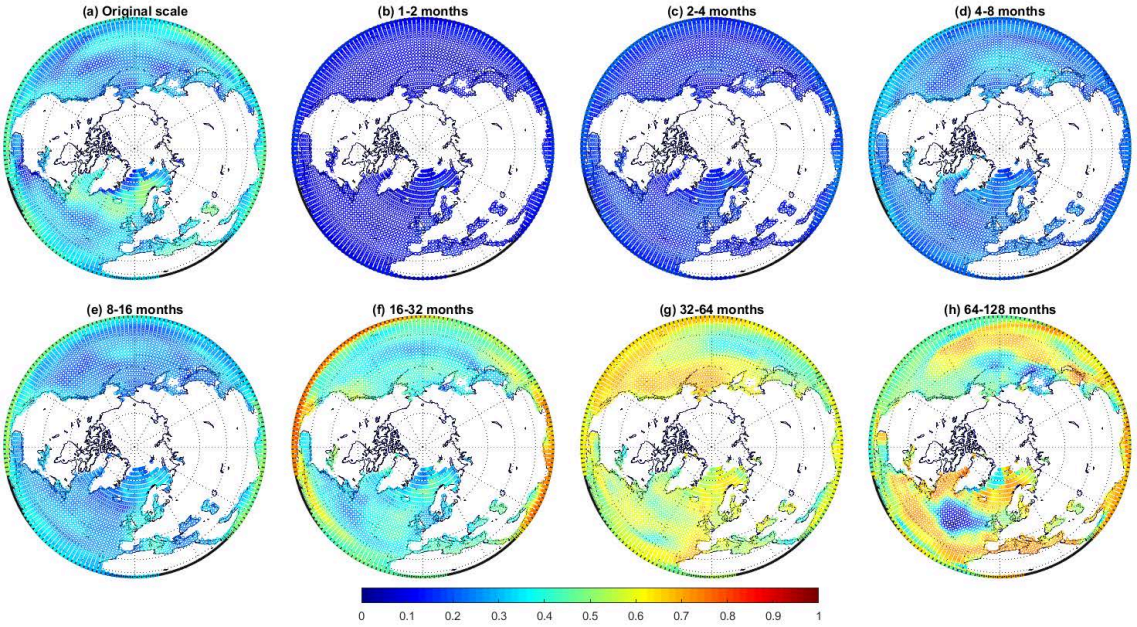



**Figure C.2.:** 95*th* percentile of positive wavelet multiscale correlation (WMC) values for the SST network constructed on multiscale using the PCC (original scale) and WMC at scales 1-7. Color represents the strength of correlation on each scale.

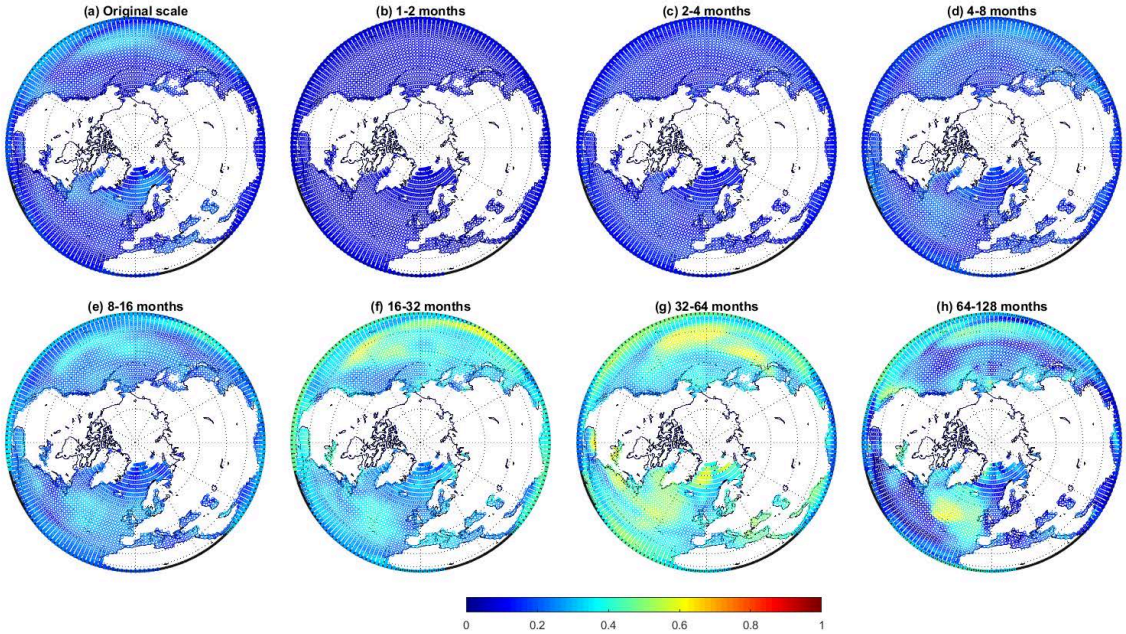



**Figure C.3.:** 95*th* percentile of negative wavelet multiscale correlation (WMC) values for the SST network constructed on multiscale using the PCC (original scale) and WMC at scales 1-7. Color represents the strength of correlation on each scale.

## C.2 Orthographic projection with center at South pole

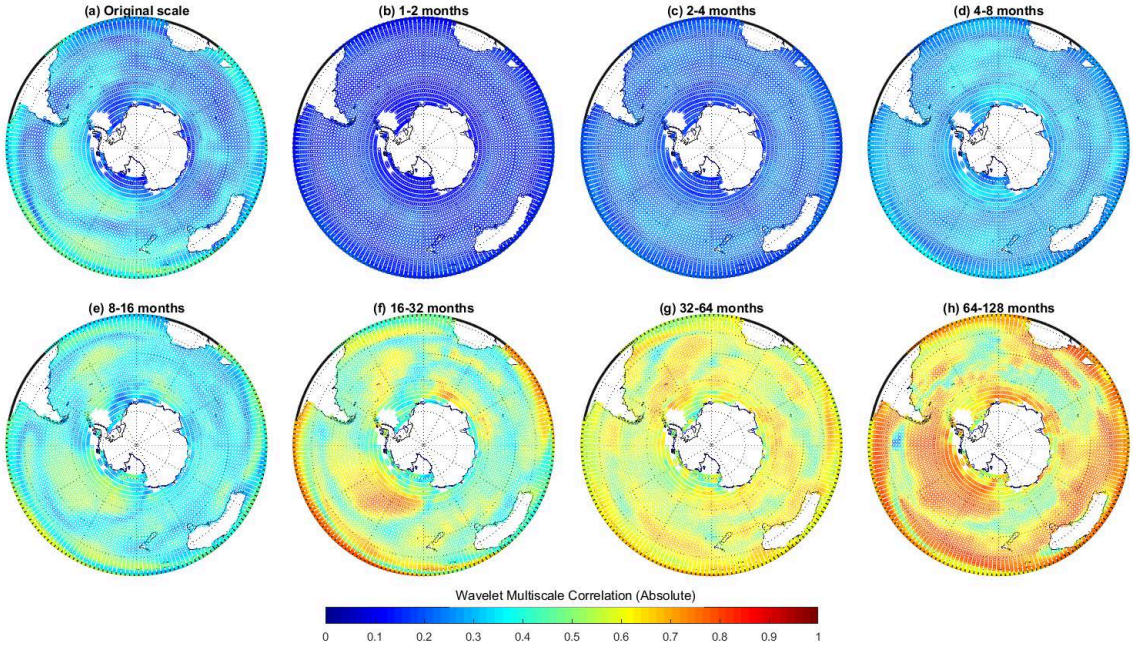



**Figure C.4.:** $95th$ percentile of absolute wavelet multiscale correlation (WMC) values for the SST network constructed on multiscale using the PCC (original scale) and WMC at scales 1-7. Color represents the strength of correlation on each scale.

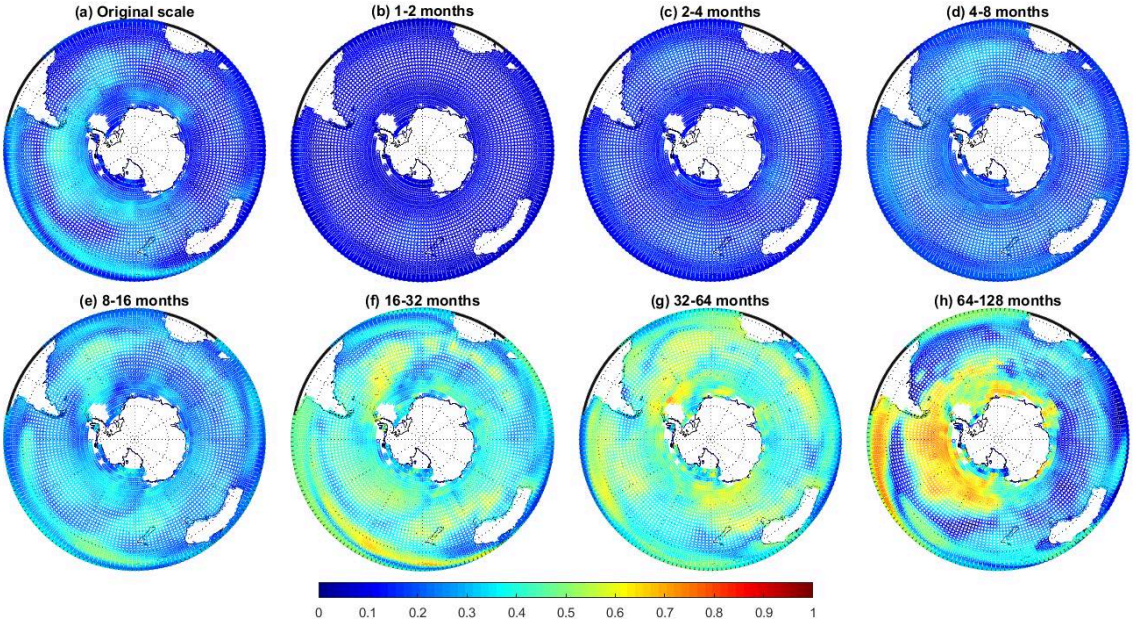



**Figure C.5.:** $95th$ percentile of negative wavelet multiscale correlation (WMC) values for the SST network constructed on multiscale using the PCC (original scale) and WMC at scales 1-7. Color represents the strength of correlation on each scale.

# Appendix D

# Supporting figures for:Short and long-range teleconnections in oceans and atmosphere

## D.1 Wavelet multi-scale correlation

MODWT decomposes the time series into different time scales or frequency components. The wavelet decomposition is realized using the two basis functions known as father wavelet ($\phi(t)$) and mother wavelet ($\psi(t)$) . The general admissibility conditions for $\psi$ to be called a wavelet function are

$$\int_{-\infty}^{\infty} \psi(t)\mathrm{d}t = 0 \tag{D.1}$$

$$\int_{-\infty}^{\infty} \psi(t)^2\mathrm{d}t = 1 \tag{D.2}$$

Any function $f(t)$ can be expressed through these basis functions and their scaled and translated versions are

$$f(t) = \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\Phi_{J,k}(t) + \quad \sum_k d_{J-1,k}\Phi_{J-1,k}(t) + \ldots + \sum_k d_{1,k}\Phi_{1,k}(t) \tag{D.3}$$

where $J$ is the total number of scales to be analyzed, and $k$ is in the range of 1 to $l$ (length of the time series). The coefficients $S_{J,k}$, are the approximation coefficients and $d_{J,k},\ldots,d_{1,k}$ are the wavelet transform coefficients at scales $J$ to 1, while the functions $\phi_{J,k}(t)$ and $\{\Psi_{j,k}|j=1,\ldots,J-1,J\}$ are the basis functions which are obtained through translation and dilation of the father of father ($\phi(t)$) and mother ($\psi(t)$) wavelets function respectively (Giri et al., 2014).
The mother and father wavelet is scaled (or dilated) by a factor $j$ and translated (or shifted) by a factor $k$ to give

$$\phi_{j,k}(t) = 2^{\frac{-j}{2}}\phi(2^{-j}t-k) \tag{D.4}$$

$$\Phi_{j,k}(t) = 2^{\frac{-j}{2}}\Phi(2^{-j}t-k) \tag{D.5}$$

Further, the values of the wavelet transform coefficients at each of the scale and the approximation coefficients at scale $J$ are estimated by:

$$d_{j,k} \approx \int \Phi_{j,k}(t) f(t) dt, \quad j = 1, \ldots, J-1, J \tag{D.6}$$

and

$$s_{j,k} \approx \int \phi_{J,k}(t) f(t) dt \tag{D.7}$$

where the scaling coefficients $s_{J,k}$ capture the smooth trend of the time series at the coarse scale $2^J$, which are also called smooth coefficients; and the wavelet coefficients $d_{j,k}$, also known as detail coefficients can detect deviations from the coarsest scale to the finest scale. The original series $f(t)$ can be reconstructed by summing up the detailed components and the smooth components:

$$f(t) = S_{J,k} + D_{J,k} + D_{J-1,k} + \ldots + D_{1,k} \tag{D.8}$$

where

$$S_{j,k} = \sum_k s_{J,k}\phi_{J,k}(t), D_{J,k} = \sum_k d_{J,k}\Phi_{J,k}(t), \ldots, D_{1,k} = \sum_k d_{1,k}\Phi_{1,k}(t). \tag{D.9}$$

Further, consider two time series $\{X(t)\}$ and $\{Y(t)\}$ with the same length $T$. The wavelet multi-scale correlation (WMC) measure between both time series can be estimated as (Lark et al., 2004; Fernández-Macho, 2012)

$$WMC = \rho_{X,Y}^{l_j} \equiv \frac{COV_{X,Y}^{l_j}}{Var_Y^{l_j} Var_X^{l_j}} \tag{D.10}$$

where

$$Var_X^{l_j} = \frac{1}{T_j} \sum_{t=M_j-1}^{T-1} [d_{j,t}^X]^2 \tag{D.11}$$

$$Var_Y^{l_j} = \frac{1}{T_j} \sum_{t=M_j-1}^{T-1} [d_{j,t}^Y]^2 \tag{D.12}$$

$$COV_{X,Y}^{l_j} = \frac{1}{T_j} \sum_{t=M_j-1}^{T-1} d_{j,t}^X d_{j,t}^Y \tag{D.13}$$

$d_{j,t}^{(.)}$ denotes the MODWT wavelet coefficient of variables $\{X,Y\}$ at scale $l_j$; $T_j' = T - M_j + 1$ stands for the number of coefficients unaffected by the boundary; $M_j = (2^j - 1)(M-1) + 1$ represents the length of the scale $l_j$ wavelet filter, and $M$ is the width of the wavelet filter. The multi-scale correlation measure $\rho_{X,Y}^{l_j}$ denotes the scale-wise correlation between $X$ and $Y$ at different $l_j$ scales. Like the Pearson correlation coefficient (PCC), the value of $\rho_{X,Y}^{l_j}$ ranges between $-1$ and 1. $\rho_{X,Y}^{l_j}$=o implies the variables $X$ and $Y$ are not correlated at scale $l_j$. $\rho_{X,Y}^{l_j} = 1$ and $\rho_{X,Y}^{l_j} = -1$ indicate that the variables are perfectly correlated and anti-correlated, respectively. The values of $\rho_{X,Y}^{1l_j}$.....$\rho_{X,Y}^{l_j}$ indicate the strength of the relation between $X$ and $Y$ at different temporal scale bands.

## D.2 Supplementary figures

We have also plotted the histograms between fraction of links and spatial distances, similar to those presented in the chapter 4. Fraction of links is the ratio between total links and all possible links which is calculated as $\frac{N(N-1)}{2}$, where $N$ is the number of the nodes in the network. All these figures are not the part of main text but compliment the findings presented in the chapter 4.

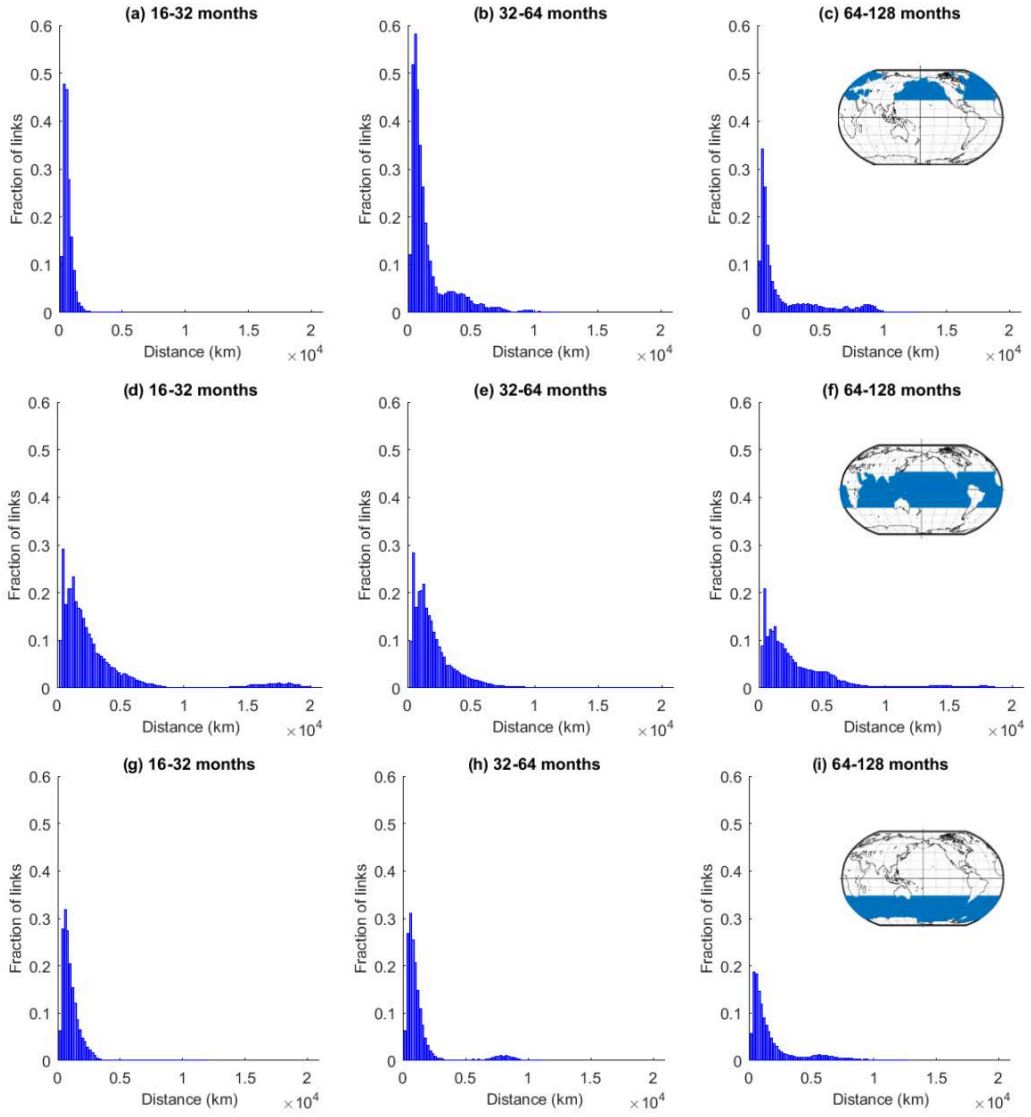



**Figure D.1.:** Fraction of links (actual links/possible links) in SST networks at different timescales within three standard zones, namely NH, TH and SH (within-region links). Short-distance links show near-field spatial correlation, while long-distance links suggest long-range spatial dependencies or teleconnections within the same zone.

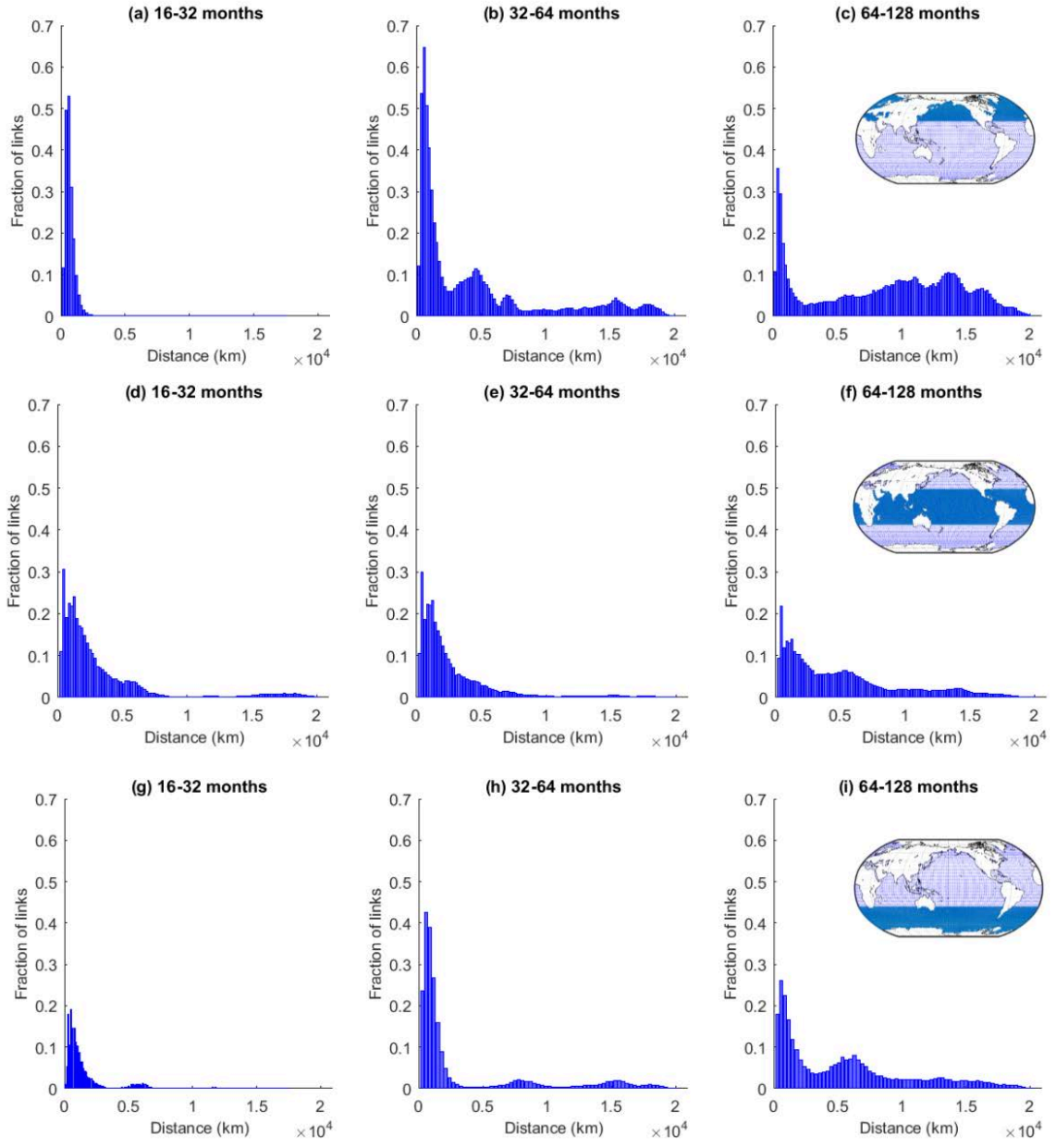



**Figure D.2.:** Fraction of links (actual links/possible links) in SST networks at different timescales constructed for standard zones, namely NH, TH, and SH, considering the interaction with the whole globe (beyond-region links). Short-distance links show near-field correlation, while long-distance links suggest long-range spatial dependencies or teleconnections.

# Appendix E

# Supporting Information for:Unravelling the spatial diversity of Indian rainfall teleconnections

## E.1 Event synchronization

Event synchronization is a promising measure for determining the time synchronization and delay pattern between two event-like signals Quiroga et al. (2002b). We define an event when a value in the signal x(t)(or y(t)) exceeds a threshold (selected by a $\alpha$ percentile). Events in x(t) and y(t) occurring at time $t_l^x$ and $t_m^y$ where l=1,2,3,4...$S_x$, m=1,2,3,4...$S_y$, are considered to be synchronized when they occur within a time lag $\pm\tau_{lm}^{xy}$ which is defined as

$$\tau_{lm}^{xy} = \min\{t_{l+1}^x - t_l^x, t_l^x - t_{l-1}^x, t_{m+1}^y - t_m^y, t_m^y - t_{m-1}^y\} \tag{E.1}$$

where $S_x$ and $S_y$ are the total number of events (greater than the threshold $\alpha$) in the signals x(t) and y(t), respectively. This definition of the time lag helps to separate independent events, which in turn allows to take into account the fact that different processes are responsible for the generation of events. To count the number of times an event occur in x(t) after it appears in y(t) and vice versa, C(x|y) and (C(y|x)) are defined as follows:

$$C(x|y) = \sum_{l=1}^{S_x} \sum_{m=1}^{S_y} J_{xy} \tag{E.2}$$

and

$$J_{xy} = \begin{cases} 1, & \text{if } 0 < t_x^l - t_m^y < \tau_{lm}^{xy} \\ \frac{1}{2}, & \text{if } t_x^l = t_m^y \\ 0, & \text{otherwise.} \end{cases} \tag{E.3}$$

C(y|x) is calculated analogously but with exchanged x and y. From these quantities we obtain the symmetric measure:

$$Q_{xy} = \frac{C(x|y) + C(y|x)}{\sqrt{(S_x - 2)(S_y - 2)}} \tag{E.4}$$

$Q_{xy}$ is a measure of the strength of event synchronization between x(t) and y(t). It is normalized to 0$< Q_{xy} <$ 1, with $Q_{xy}$=1 for perfect synchronization (coincidence of extreme

events) between signals x(t) and y(t).

## E.2 Maximum overlap discrete wavelet transformation

Time series of continuous geophysical variables can be interpreted as the superposition of variations occurring at different scales. Different physical processes drive these patterns, and a partitioning of the variability at different scales can help to isolate and characterize the underlying processes (Sturtevant et al., 2016). Wavelets have been successfully used to characterize the time scales of interactions between hydrometeorological variables (Molini et al., 2010).

The wavelet transform of a signal decomposes it into a set of components with predefined central frequencies and spectral bandwidths. Here we use the maximal overlap discrete wavelet transform(Percival and Walden, 2000) (MODWT), because the orthogonal discrete wavelet transform (DWT) results in a pyramid of wavelet coefficients which does not contain the time synchronization of the events. Further, our experience with DWT suggests that it suffers from 'shift sensitivity' also known as 'shift variance' which is undesirable because DWT coefficients fail to distinguish between input-signal shifts(Rathinasamy and Khosa, 2012). Although MODWT has considerable redundancy but it is shift invariant, and this property renders the MODWT more suited for time series analysis.

MODWT decomposes the time series into different time scales or frequency components. The wavelet decomposition is realized using the two basis functions known as father wavelets and mother wavelet. Any function f(t) can be expressed in these basis functions and their scaled and translated versions as given in Eq.(E.5)

$$f(t) = \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\Phi_{J,k}(t) + \sum_k d_{J-1,k}\Phi_{J-1,k}(t) + \ldots + \sum_k d_{1,k}\Phi_{1,k}(t) \tag{E.5}$$

where $J$ is the total number of scales to be analyzed, and $k$ is in the range of $-1$ to l (length of the time series). The coefficients $S_{J,k}$, are the approximation coefficients and $d_{J,k},\ldots,d_{1,k}$ are the wavelet transform coefficients at scales $J$ to 1, while the functions $\phi_{J,k}(t)$ and $\{\Psi_{j,k}|j=1,\ldots,J-1,J\}$ are the basis functions which are obtained through translation and dilation of the father of father ($\phi(t)$) and mother ($\psi(t)$) wavelets function respectively. The mother and father wavelet is scaled (or dilated) by a factor j and translated (or shifted) by a factor k to give

$$\phi_{j,k}(t) = 2^{\frac{-j}{2}}\phi(2^{-j}t-k) \tag{E.6}$$

$$\Phi_{j,k}(t) = 2^{\frac{-j}{2}}\Phi(2^{-j}t-k) \tag{E.7}$$

Further, the values of the wavelet transform coefficients at each of the scale and the approximation coefficients at scale J are estimated by:

$$d_{j,k} \approx \int \Phi_{j,k}(t)f(t)dt, \quad j=1,\ldots,J-1,J \tag{E.8}$$

and

$$s_{j,k} \approx \int \phi_{J,k}(t)f(t)dt \tag{E.9}$$

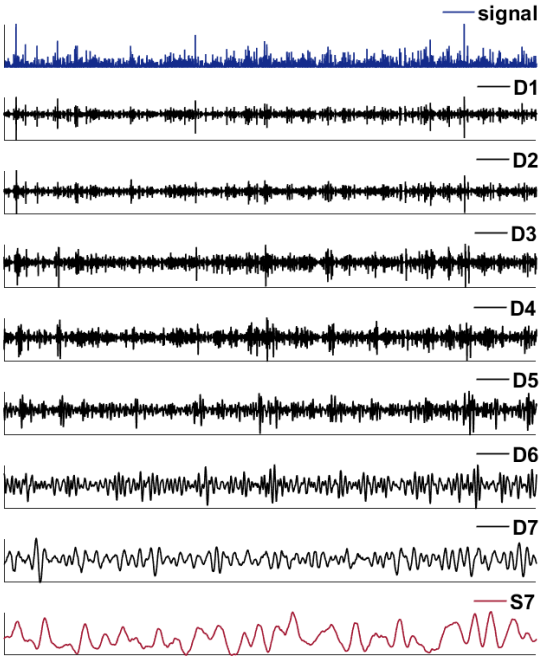



**Figure E.1.:** Scheme of multi-scale decomposition of signals using maximum overlap discrete wavelet transformation (MODWT). The relationship between signal $Y_t$ (blue), detailed component $D_j$ (black), and approximate component $S_j$ (red), is shown.

where the scaling coefficients $s_{J,k}$ capture the smooth trend of the time series at the coarse scale $2^J$, which are also called smooth coefficients; and the wavelet coefficients $d_{j,k}$, also known as detail coefficients can detect deviations from the coarsest scale to the finest scale. The original series $f(t)$ can be reconstructed by summing up the detailed components and the smooth components.

$$f(t) = S_{J,k} + D_{J,k} + D_{J-1,k} + \ldots + D_{1,k} \tag{E.10}$$

where

$$S_{j,k} = \sum_k s_{J,k}\phi_{J,k}(t), D_{J,k} = \sum_k d_{J,k}\Phi_{J,k}(t), \ldots, D_{1,k} = \sum_k d_{1,k}\Phi_{1,k}(t). \tag{E.11}$$

Eq.(E.10) defines a multiresolution analysis (MRA) of f(t); i.e., we express the series f(t) as the sum of a constant vector $S_J$ and $J$ other vectors $D_j$, $j = 1,2,3,...J$, each of which contain a timeseries related to variations in f(t)at a certain scale. We refer to $D_j$ as the $j^{th}$ level wavelet detail. Fig.E.1 shows the MODWT decomposition of a sample signal up to 7 scales resulting in 7 detailed components (D1-D7) and one approximate Component (S7).

Let $Y_t$ represents a time series history of a geophysical process. In order to partition the variability of the process at different scales $j = 1,2,...J$, the signal $Y_t$ is transformed into the wavelet space which provides the required information at different scales. This is obtained by convolving $Y_t$ with a set of low pass (g) and high pass (h) filters. For instance, at each scale j, the MODWT applies a high pass wavelet filter $h_{j,l}$ and a lower pass filter $g_{j,l}$ of length (l) to the time series Y to respectively yield the wavelet coefficients $W_{j,t}$ and $V_{j,t}$ for every point t in the time series (Percival and Walden 2000).

$$W_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} Y_{t-l \bmod 2N} V_{j,t} = \sum_{l=0}^{L_j-1} g_{j,l} Y_{t-l \bmod 2N} \tag{E.12}$$

The $W_{j,t}$ wavelet coefficients distinguish fluctuations in the time series of scale $2^{j-1}$,while the $V_{j,t}$ coefficients provide information about the variations at scale $2^j$ and higher. Let the maximum level of decomposition be $j = J$. This would result in a total 'J+1' series of wavelet coefficients with $W_{j,t}, j = 1,2,3,...J$, and one series of $V_{j,t}$.
Let us now define $D_j$ which represents the time domain reconstruction of $W_j$. It represents the portion of Y attributable to scale j. Let $S_J$ represent the time domain reconstruction of $V_J$. For the maximum level of decomposition, $V_J$ has all of its elements equal to the sample mean of Y.
Therefore, we can write

$$Y = \sum_{j=1}^{J} D_j + S_j \tag{E.13}$$

## E.3 Wavelet Coherence

To compare the results obtained from the MSES, we use wavelet coherence (WC). Wavelet coherence analysis has been used as a robust tool in identifying the relationship between two variables at multiple scales. The wavelet coherence between time series $X_t$ and $Y_t$ was defined by(Torrence and Compo, 1998) as

$$R^2(j,t) = \frac{|\zeta(j^{-1} W_{xy}(j,t))|^2}{\zeta(j^{-1}|W_x(j,t)^2|).\zeta(j^{-1}|W_y(j,t)|^2)} \tag{E.14}$$

where $R^2(j,t)$ takes a value between 0 and 1, $\zeta$ is a smoothing operator and can be written as $\zeta(W) = \zeta_{scale}(\zeta_{time}(W(j,t)))$. $W_{xy}$ represents the cross-wavelet coefficient between X and Y.
$W_x(j,t)$ and $W_y(j,t)$ denote the wavelet coefficients obtained from wavelet transform of X and Y respectively at scale j and time t. The global wavelet coherence at a certain scale j is defined as the time-averaged value of the wavelet coefficients at the scale with the COI. It is estimated by

$$R^2(j) = \frac{1}{n} \sum_{t=t1}^{t2} R^2(j,t) \tag{E.15}$$

Where $n_j$ is the number of points with COI and $n_j = t2 - t1 + 1$.
Global wavelet coherence is a useful measure to examine the common characteristic periodicities between x and y. Grinsted et al. (2004) showed the applicability of WC analysis of the association of precipitation with climate variables. More detailed description of wavelet coherence analysis is can be found in Grinsted et al. (2004).

# Bibliography


Abbasi, A. and Hossain, L. (2013). "Hybrid Centrality Measures for Binary and Weighted Networks". In: *Complex Networks*. Ed. by R. Menezes, A. Evsukoff, and M. C. González. Vol. 424. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–7. ISBN: 978-3-642-30286-2 978-3-642-30287-9. DOI: 10.1007/978-3-642-30287-9_1 (cit. on p. 32).

Abid, M. A., Almazroui, M., Kucharski, F., O'Brien, E., and Yousef, A. E. (2018). "ENSO relationship to summer rainfall variability and its potential predictability over Arabian Peninsula region". In: *npj Climate and Atmospheric Science* 1.1. ISSN: 2397-3722. DOI: 10.1038/s41612-017-0003-7 (cit. on p. 64).

Achard, S. (2006). "A Resilient, Low-Frequency, Small-World Human Brain Functional Network with Highly Connected Association Cortical Hubs". In: *Journal of Neuroscience* 26.1, pp. 63–72. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3874-05.2006 (cit. on pp. 6, 32).

Addison, P. S. (2005). "Wavelet transforms and the ECG: a review". In: *Physiological Measurement* 26.5, R155–R199. ISSN: 0967-3334, 1361-6579. DOI: 10.1088/0967-3334/26/5/R01 (cit. on pp. 15, 32, 33, 40).

Adhikary, S. K., Yilmaz, A. G., and Muttil, N. (2015). "Optimal design of rain gauge network in the Middle Yarra River catchment, Australia". In: *Hydrological Processes* 29.11, pp. 2582–2599. ISSN: 08856087. DOI: 10.1002/hyp.10389 (cit. on pp. 106, 113).

Agarwal, A., Maheswaran, R., Kurths, J., and Khosa, R. (2016a). "Wavelet Spectrum and Self-Organizing Maps-Based Approach for Hydrologic Regionalization -a Case Study in the Western United States". In: *Water Resources Management* 30.12, pp. 4399–4413. ISSN: 0920-4741, 1573-1650. DOI: 10.1007/s11269-016-1428-1 (cit. on p. 26).

Agarwal, A., Maheswaran, R., Sehgal, V., Khosa, R., Sivakumar, B., and Bernhofer, C. (2016b). "Hydrologic regionalization using wavelet-based multiscale entropy method". In: *Journal of Hydrology* 538, pp. 22–32. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2016.03.023 (cit. on pp. 32, 90).

Agarwal, A., Marwan, N., Maheswaran, R., Merz, B., and Kurths, J. (2018a). "Quantifying the roles of single stations within homogeneous regions using complex network analysis". In: *Journal of Hydrology*. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2018.06.050 (cit. on pp. 52, 65–67, 69).

Agarwal, A., Marwan, N., Ozturk, U., and Maheswaran, R. (2019). "Unfolding Community Structure in Rainfall Network of Germany Using Complex Network-Based Approach". In: *Water Resources and Environmental Engineering II*. Springer, pp. 179–193 (cit. on p. 109).

Agarwal, A. (2015). "Hydrologic regionalization using wavelet-based multiscale entropy technique". In: (cit. on p. 5).

Agarwal, A., Marwan, N., Rathinasamy, M., Merz, B., and Kurths, J. (2017). "Multi-scale event synchronization analysis for unravelling climate processes: a wavelet-based

approach". In: *Nonlinear Processes in Geophysics* 24.4, pp. 599–611. ISSN: 1607-7946. DOI: 10.5194/npg-24-599-2017 (cit. on pp. 32, 35, 36, 50, 65, 66, 68, 93, 107).

Agarwal, A., Marwan, N., Rathinasamy, M., Ozturk, U., Merz, B., and Kurths, J. (2018b). "Optimal Design of Hydrometric Station Networks Based on Complex Network Analysis". In: *Hydrology and Earth System Sciences Discussions*, pp. 1–21. ISSN: 1812-2116. DOI: 10.5194/hess-2018-113 (cit. on pp. 32, 106).

Amorim, A. M. T., Gonçalves, A. B., Nunes, L. M., and Sousa, A. J. (2012). "Optimizing the location of weather monitoring stations using estimation uncertainty". In: *International Journal of Climatology* 32.6, pp. 941–952. ISSN: 08998418. DOI: 10.1002/joc.2317 (cit. on p. 106).

Anctil, F. and Coulibaly, P. (2004). "Wavelet analysis of the interannual variability in southern Québec streamflow". In: *Journal of climate* 17.1, pp. 163–173 (cit. on p. 6).

Araghi, A., Mousavi-Baygi, M., Adamowski, J., and Martinez, C. (2017). "Association between three prominent climatic teleconnections and precipitation in Iran using wavelet coherence". In: *International Journal of Climatology* 37.6, pp. 2809–2830. ISSN: 08998418. DOI: 10.1002/joc.4881 (cit. on p. 64).

Arenas, A., Diaz-Guilera, A., Kurths, J., Moreno, Y., and Zhou, C. (2008). "Synchronization in complex networks". In: *Physics reports* 469.3, pp. 93–153 (cit. on p. 107).

Arnhold, J., Grassberger, P., Lehnertz, K., and Elger, C. (1999). "A robust method for detecting interdependences: application to intracranially recorded EEG". In: *Physica D: Nonlinear Phenomena* 134.4, pp. 419–430. ISSN: 01672789. DOI: 10.1016/S0167-2789(99)00140-2 (cit. on pp. 6, 14).

Ashok, K., Guan, Z., and Yamagata, T. (2001). "Impact of the Indian Ocean dipole on the relationship between the Indian monsoon rainfall and ENSO". In: *Geophysical Research Letters* 28.23, pp. 4499–4502 (cit. on pp. 64, 66).

Azad, S. and Rajeevan, M. (2016). "Possible shift in the ENSO-Indian monsoon rainfall relationship under future global warming". en. In: *Scientific Reports* 6.1. ISSN: 2045-2322. DOI: 10.1038/srep20145 (cit. on p. 50).

Bansod, S. D. (2011). "Interannual variability of convective activity over the tropical Indian Ocean during the El Niño/La Niña events". In: *International Journal of Remote Sensing* 32.19, pp. 5565–5582. ISSN: 0143-1161, 1366-5901. DOI: 10.1080/01431161.2010.506896 (cit. on p. 72).

Barca, E., Passarella, G., and Uricchio, V. (2008). "Optimal extension of the rain gauge monitoring network of the Apulian Regional Consortium for Crop Protection". In: *Environmental Monitoring and Assessment* 145.1, pp. 375–386. ISSN: 0167-6369, 1573-2959. DOI: 10.1007/s10661-007-0046-z (cit. on p. 106).

Barrat, A., Barthelemy, M., and Vespignani, A. (2008). "Dynamical processes on complex networks Cambridge University Press". In: *Cambridge/New York* (cit. on p. 14).

Basaras, P., Katsaros, D., and Tassiulas, L. (2013). "Detecting Influential Spreaders in Complex, Dynamic Networks". In: *Computer* 46.4, pp. 24–29. ISSN: 0018-9162. DOI: 10.1109/MC.2013.75 (cit. on p. 32).

Behera, S. K., Krishnan, R., and Yamagata, T. (1999). "Unusual ocean-atmosphere conditions in the tropical Indian Ocean during 1994". In: *Geophysical Research Letters* 26.19, pp. 3001–3004. ISSN: 00948276. DOI: 10.1029/1999GL010434 (cit. on pp. 50, 64).

Bell, M. G. and Lida, Y. (1997). *Transportation Network Analysis: Bell/Transportation Network Analysis*. Chichester, UK: John Wiley & Sons, Ltd. ISBN: 978-1-118-90303-2 978-0-471-96493-3. DOI: 10.1002/9781118903032 (cit. on p. 106).

Benjamini, Y. and Yosef, H. (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." In: *Journal of the Royal Statistical Society.* b 57.1, pp. 289–300 (cit. on p. 52).

Berberan-Santos, M. N., Bodunov, E. N., and Pogliani, L. (1997). "On the barometric formula". In: *American Journal of Physics* 65.5, pp. 404–412. ISSN: 0002-9505, 1943-2909. DOI: 10.1119/1.18555 (cit. on p. 1).

Berndtsson, R. (1988). "Temporal variability in spatial correlation of daily rainfall". In: *Water Resources Research* 24.9, pp. 1511–1517. ISSN: 00431397. DOI: 10.1029/WR024i009p01511 (cit. on pp. 1, 90).

Betzel, R. F. and Bassett, D. S. (2017). "Multi-scale brain networks". In: *NeuroImage* 160, pp. 73–83. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2016.11.006 (cit. on p. 6).

Bharath, R. and Srinivas, V. V. (2015). "Delineation of homogeneous hydrometeorological regions using wavelet-based global fuzzy cluster analysis". In: *International Journal of Climatology* 35.15, pp. 4707–4727. ISSN: 08998418. DOI: 10.1002/joc.4318 (cit. on pp. 64, 65).

Bhatla, R., Singh, A. K., Mandal, B., Ghosh, S., Pandey, S. N., and Sarkar, A. (2016). "Influence of North Atlantic Oscillation on Indian Summer Monsoon Rainfall in Relation to Quasi-Binneal Oscillation". In: *Pure and Applied Geophysics* 173.8, pp. 2959–2970. ISSN: 0033-4553, 1420-9136. DOI: 10.1007/s00024-016-1306-z (cit. on p. 73).

Bigg, G. R., Jickells, T. D., Liss, P. S., and Osborn, T. J. (2003). "The role of the oceans in climate". en. In: *International Journal of Climatology* 23.10, pp. 1127–1159. ISSN: 0899-8418, 1097-0088. DOI: 10.1002/joc.926 (cit. on p. 49).

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008. ISSN: 1742-5468. DOI: 10.1088/1742-5468/2008/10/P10008 (cit. on pp. 94, 95).

Blöschl, G. and Sivapalan, M. (1995). "Scale issues in hydrological modelling: A review". In: *Hydrological Processes* 9.3, pp. 251–290. ISSN: 08856087, 10991085. DOI: 10.1002/hyp.3360090305 (cit. on p. 4).

Boers, N., Donner, R. V., Bookhagen, B., and Kurths, J. (2015). "Complex network analysis helps to identify impacts of the El Niño Southern Oscillation on moisture divergence in South America". In: *Climate Dynamics* 45.3, pp. 619–632. ISSN: 0930-7575, 1432-0894. DOI: 10.1007/s00382-014-2265-7 (cit. on pp. 32, 64, 107).

Boers, N., Rheinwalt, A., Bookhagen, B., Barbosa, H. M. J., Marwan, N., Marengo, J., and Kurths, J. (2014). "The South American rainfall dipole: A complex network analysis of extreme events: BOERS ET AL." In: *Geophysical Research Letters* 41.20, pp. 7397–7405. ISSN: 00948276. DOI: 10.1002/2014GL061829 (cit. on pp. 32, 64, 107).

Bowden, A. J., Burek, C. V., Wilding, R., and Geological Society of London, eds. (2005). *History of palaeobotany: selected essays.* Geological Society special publication 241. OCLC: ocm60397723. London: The Geological Society. 304 pp. ISBN: 978-1-86239-174-1 (cit. on p. 1).

Bullmore, E. and Sporns, O. (2009). "Complex brain networks: graph theoretical analysis of structural and functional systems". In: *Nature Reviews Neuroscience* 10.3, p. 186 (cit. on p. 2).

– (2012). "The economy of brain network organization". In: *Nature Reviews Neuroscience.* ISSN: 1471-003X, 1471-0048. DOI: 10.1038/nrn3214 (cit. on p. 107).

Buschmann, S., Hoffmann, P., Agarwal, A., Marwan, N., and Nocke, T. (in prep.). "GPU-based, interactive Exploration of large spatio-temporal Climate Networks". In: (cit. on p. 58).

Butte, A. J. and Kohane, I. S. (1999). "Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurments". In: WORLD SCIENTIFIC, pp. 418–429. ISBN: 978-981-02-4188-9 978-981-4447-33-1. DOI: 10.1142/9789814447331_0040 (cit. on p. 32).

Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., and Saba, V. (2018). "Observed fingerprint of a weakening Atlantic Ocean overturning circulation". en. In: *Nature* 556.7700, pp. 191–196. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-018-0006-5 (cit. on pp. 60, 82).

Callahan, D., Kennedy, K., and Subhlok, J. (1990). "Analysis of event synchronization in a parallel programming tool". In: *ACM SIGPLAN Notices*. Vol. 25. 3. ACM, pp. 21–30 (cit. on p. 14).

Casagrande, E., Mueller, B., Miralles, D. G., Entekhabi, D., and Molini, A. (2015). "Wavelet correlations to reveal multiscale coupling in geophysical systems". In: *Journal of Geophysical Research: Atmospheres* 120.15, pp. 7555–7572. ISSN: 2169897X. DOI: 10.1002/2015JD023265 (cit. on pp. 3, 5, 32, 50, 64).

Chacon-Hurtado, J. C., Alfonso, L., and Solomatine, D. P. (2017). "Rainfall and streamflow sensor network design: A review of applications, classification, and a proposed framework". In: *Hydrology and Earth System Sciences* 21.6, pp. 3071–3091 (cit. on p. 106).

Chen, D., Cane, M. A., Kaplan, A., Zebiak, S. E., and Huang, D. (2004). "Predictability of El Niño over the past 148 years". In: *Nature* 428.6984, pp. 733–736. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature02439 (cit. on pp. 42, 59).

Conradt, T., Koch, H., Hattermann, F. F., and Wechsung, F. (2012). "Precipitation or evapotranspiration? Bayesian analysis of potential error sources in the simulation of sub-basin discharges in the Czech Elbe River basin". In: *Regional Environmental Change* 12.3, pp. 649–661. ISSN: 1436-3798, 1436-378X. DOI: 10.1007/s10113-012-0280-y (cit. on pp. 29, 118).

Conticello, F., Cioffi, F., Merz, B., and Lall, U. (2018). "An event synchronization method to link heavy rainfall events and large-scale atmospheric circulation features". In: *International Journal of Climatology* 38.3, pp. 1421–1437. ISSN: 08998418. DOI: 10.1002/joc.5255 (cit. on pp. 64, 65, 91, 92, 107).

Darand, M. and Mansouri Daneshvar, M. R. (2014). "Regionalization of Precipitation Regimes in Iran Using Principal Component Analysis and Hierarchical Clustering Analysis". In: *Environmental Processes* 1.4, pp. 517–532. ISSN: 2198-7491, 2198-7505. DOI: 10.1007/s40710-014-0039-1 (cit. on p. 90).

D'Arrigo, R. (2005). "On the variability of ENSO over the past six centuries". In: *Geophysical Research Letters* 32.3. ISSN: 0094-8276. DOI: 10.1029/2004GL022055 (cit. on p. 72).

Deza, J. I., Barreiro, M., and Masoller, C. (2015). "Assessing the direction of climate interactions by means of complex networks and information theoretic tools". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25.3, p. 033105. ISSN: 1054-1500, 1089-7682. DOI: 10.1063/1.4914101 (cit. on p. 32).

Dong, X. (2016). "Influences of the Pacific Decadal Oscillation on the East Asian Summer Monsoon in non-ENSO years: Influences of the Pacific Decadal Oscillation on the East Asian Summer Monsoon". In: *Atmospheric Science Letters* 17.1, pp. 115–120. ISSN: 1530261X. DOI: 10.1002/asl.634 (cit. on pp. 64, 66).

Donges, J. F., Schultz, H. C. H., Marwan, N., Zou, Y., and Kurths, J. (2011). "Investigating the topology of interacting networks: Theory and application to coupled climate

subnetworks". In: *The European Physical Journal B* 84.4, pp. 635–651. ISSN: 1434-6028, 1434-6036. DOI: 10.1140/epjb/e2011-10795-8 (cit. on pp. 92, 107).

Donges, J. F., Zou, Y., Marwan, N., and Kurths, J. (2009a). "Complex networks in climate dynamics: Comparing linear and nonlinear network construction methods". In: *The European Physical Journal Special Topics* 174.1, pp. 157–179. ISSN: 1951-6355, 1951-6401. DOI: 10.1140/epjst/e2009-01098-2 (cit. on pp. 6, 31, 50, 52, 107).

– (2009b). "The backbone of the climate network". In: *EPL (Europhysics Letters)* 87.4, p. 48007. ISSN: 0295-5075, 1286-4854. DOI: 10.1209/0295-5075/87/48007 (cit. on pp. 50, 107, 110, 118).

Donner, R., Barbosa, S., Kurths, J., and Marwan, N. (2009). "Understanding the Earth as a Complex System – recent advances in data analysis and modelling in Earth sciences". In: *The European Physical Journal Special Topics* 174.1, pp. 1–9. ISSN: 1951-6355, 1951-6401. DOI: 10.1140/epjst/e2009-01086-6 (cit. on p. 14).

E, W. (2011). *Principles of multiscale modeling*. OCLC: ocn721888752. Cambridge ; New York: Cambridge University Press. 466 pp. ISBN: 978-1-107-09654-7 (cit. on p. 4).

Eltahir, E. A. B. and Bras, R. L. (1996). "Precipitation recycling". In: *Reviews of Geophysics* 34.3, pp. 367–378. ISSN: 87551209. DOI: 10.1029/96RG01927 (cit. on p. 1).

Eroglu, D., McRobie, F. H., Ozken, I., Stemler, T., Wyrwoll, K.-H., Breitenbach, S. F. M., Marwan, N., and Kurths, J. (2016). "See–saw relationship of the Holocene East Asian–Australian summer monsoon". In: *Nature Communications* 7, p. 12929. ISSN: 2041-1723. DOI: 10.1038/ncomms12929 (cit. on p. 13).

Eustace, J., Wang, X., and Cui, Y. (2015). "Community detection using local neighborhood in complex networks". In: *Physica A: Statistical Mechanics and its Applications* 436, pp. 665–677. ISSN: 03784371. DOI: 10.1016/j.physa.2015.05.044 (cit. on p. 91).

Feliks, Y., Groth, A., Robertson, A. W., and Ghil, M. (2013). "Oscillatory Climate Modes in the Indian Monsoon, North Atlantic, and Tropical Pacific". In: *Journal of Climate* 26.23, pp. 9528–9544. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/JCLI-D-13-00105.1 (cit. on pp. 64, 66, 73).

Feng and Meyers (2003). "Interannual variability in the tropical Indian Ocean: a two-year time-scale of Indian Ocean Dipole". en. In: *Deep Sea Research Part II: Topical Studies in Oceanography* 50.12-13, pp. 2263–2284. ISSN: 09670645. DOI: 10.1016/S0967-0645(03)00056-0 (cit. on p. 59).

Feng, Vasile, R., Segond, M., Gozolchiani, A., Wang, Y., Abel, M., Havlin, S., Bunde, A., and Dijkstra, H. (2016). "ClimateLearn: A machine-learning approach for climate prediction using network measures". In: *Geoscientific Model Development Discussions*, pp. 1–18. ISSN: 1991-962X. DOI: 10.5194/gmd-2015-273 (cit. on p. 63).

Fernández-Macho, J. (2012). "Wavelet multiple correlation and cross-correlation: A multiscale analysis of Eurozone stock markets". In: *Physica A: Statistical Mechanics and its Applications* 391.4, pp. 1097–1104. ISSN: 03784371. DOI: 10.1016/j.physa.2011.11.002 (cit. on pp. 32, 34, 126).

Ferster, B., Subrahmanyam, B., and Macdonald, A. (2018). "Confirmation of ENSO-Southern Ocean Teleconnections Using Satellite-Derived SST". In: *Remote Sensing* 10.2, p. 331. ISSN: 2072-4292. DOI: 10.3390/rs10020331 (cit. on pp. 42, 43, 60).

Fovell, R. G., Tung, W.-w., and American Meteorological Society (2017). *Multiscale convection-coupled systems in the tropics: a tribute to Dr. Michio Yanai*. English. OCLC: 1024071001. ISBN: 978-1-944970-04-8 (cit. on p. 50).

Gao, C., Wei, D., Hu, Y., Mahadevan, S., and Deng, Y. (2013). "A modified evidential methodology of identifying influential nodes in weighted networks". In: *Physica A:*

*Statistical Mechanics and its Applications* 392.21, pp. 5490–5500. ISSN: 03784371. DOI: 10.1016/j.physa.2013.06.059 (cit. on p. 109).

Gaucherel (2002). "Use of wavelet transform for temporal characterisation of remote watersheds". In: *Journal of hydrology* 269.3-4, pp. 101–121 (cit. on p. 6).

Girard, P., Levison, J., Parrott, L., Larocque, M., Ouellet, M.-A., and Green, D. M. (2015). "Modeling cross-scale relationships between climate, hydrology, and individual animals: generating scenarios for stream salamanders". In: *Frontiers in Environmental Science* 3. ISSN: 2296-665X. DOI: 10.3389/fenvs.2015.00051 (cit. on p. 3).

Giri, B. K., Mitra, C., Panigrahi, P. K., and Sekar Iyengar, A. N. (2014). "Multi-scale dynamics of glow discharge plasma through wavelets: Self-similar behavior to neutral turbulence and dissipation". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 24.4, p. 043135. DOI: 10.1063/1.4903332. eprint: https://doi.org/10.1063/1.4903332 (cit. on p. 125).

Girvan and Newman (2002). "Community structure in social and biological networks". In: *Proceedings of the National Academy of Sciences* 99.12, pp. 7821–7826. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.122653799 (cit. on p. 95).

Goldberg, J. A. (2004). "Spike Synchronization in the Cortex-Basal Ganglia Networks of Parkinsonian Primates Reflects Global Dynamics of the Local Field Potentials". In: *Journal of Neuroscience* 24.26, pp. 6003–6010. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.4848-03.2004 (cit. on p. 32).

Goswami, B. and Krishnan, R. (2013). "Opportunities and challenges in monsoon prediction in a changing climate". In: *Climate Dynamics* 41.1, pp. 1–1. ISSN: 0930-7575, 1432-0894. DOI: 10.1007/s00382-013-1835-4 (cit. on p. 63).

Goswami, B., Madhusoodanan, M., Neema, C., and Sengupta, D. (2006). "A physical mechanism for North Atlantic SST influence on the Indian summer monsoon". In: *Geophysical Research Letters* 33.2 (cit. on pp. 64, 66, 73).

Gozolchiani, A., Yamasaki, K., Gazit, O., and Havlin, S. (2008). "Pattern of climate network blinking links follows El Niño events". In: *EPL (Europhysics Letters)* 83.2, p. 28005. ISSN: 0295-5075, 1286-4854. DOI: 10.1209/0295-5075/83/28005 (cit. on pp. 31, 107).

Griggs, D. J. and Noguer, M. (2002). "Climate change 2001: The scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change". en. In: *Weather* 57.8, pp. 267–269. ISSN: 00431656, 14778696. DOI: 10.1256/004316502320517344 (cit. on p. 49).

Grinsted, A., Moore, J. C., and Jevrejeva, S. (2004). "Application of the cross wavelet transform and wavelet coherence to geophysical time series". In: *Nonlinear Processes in Geophysics* 11.5, pp. 561–566. ISSN: 1607-7946. DOI: 10.5194/npg-11-561-2004 (cit. on pp. 68, 132).

Gu, J., Zhu, Y., Guo, L., Jiang, J., Chi, L., Li, W., Wang, Q. A., and Cai, X. (2015). "Recent Progress in Some Active Topics on Complex Networks". In: *Journal of Physics: Conference Series* 604, p. 012007. ISSN: 1742-6596. DOI: 10.1088/1742-6596/604/1/012007 (cit. on p. 50).

Guhathakurta, P., Menon, P., Inkane, P. M., Krishnan, U., and Sable, S. T. (2017). "Trends and variability of meteorological drought over the districts of India using standardized precipitation index". In: *Journal of Earth System Science* 126.8. ISSN: 0253-4126, 0973-774X. DOI: 10.1007/s12040-017-0896-x (cit. on p. 72).

Guimerà, R. and Amaral, L. A. N. (2005). "Cartography of complex networks: modules and universal roles". In: *Journal of Statistical Mechanics: Theory and Experiment* 2005.2, P02001. ISSN: 1742-5468. DOI: 10.1088/1742-5468/2005/02/P02001 (cit. on pp. 91, 94, 95, 100).

Guimerà, R., Sales-Pardo, M., and Amaral, L. A. N. (2007). "Classes of complex networks defined by role-to-role connectivity profiles". In: *Nature Physics* 3.1, pp. 63–69. ISSN: 1745-2473, 1745-2481. DOI: 10.1038/nphys489 (cit. on pp. 95, 96, 100).

Halverson, M. J. and Fleming, S. W. (2015). "Complex network theory, streamflow, and hydrometric monitoring system design". In: *Hydrology and Earth System Sciences* 19.7, pp. 3301–3318. ISSN: 1607-7938. DOI: 10.5194/hess-19-3301-2015 (cit. on pp. 31, 32, 68, 91, 92, 95, 96, 106, 107).

Hannachi, A., Jolliffe, I. T., and Stephenson, D. B. (2007). "Empirical orthogonal functions and related techniques in atmospheric science: A review". In: *International Journal of Climatology* 27.9, pp. 1119–1152. ISSN: 08998418, 10970088. DOI: 10.1002/joc.1499 (cit. on p. 64).

Harenberg, S., Bello, G., Gjeltema, L., Ranshous, S., Harlalka, J., Seay, R., Padmanabhan, K., and Samatova, N. (2014). "Community detection in large-scale networks: a survey and empirical evaluation: Community detection in large-scale networks". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 6.6, pp. 426–439. ISSN: 19395108. DOI: 10.1002/wics.1319 (cit. on p. 67).

Hargrove, W. W. and Hoffman, F. M. (2004). "Potential of Multivariate Quantitative Methods for Delineation and Visualization of Ecoregions". In: *Environmental Management* 34 (S1), S39–S60. ISSN: 0364-152X, 1432-1009. DOI: 10.1007/s00267-003-1084-0 (cit. on p. 106).

Hassan, B. G. and Ping, F. (2012). "Regional Rainfall Frequency Analysis for the Luanhe Basin – by Using L-moments and Cluster Techniques". In: *APCBEE Procedia* 1, pp. 126–135. ISSN: 22126708. DOI: 10.1016/j.apcbee.2012.03.021 (cit. on p. 90).

Hastenrath, S. and Hastenrath, S. (1991). *Climate dynamics of the tropics*. English. OCLC: 968304767. Dordrecht; Boston: Kluwer Academic Publishers. ISBN: 978-94-011-3156-8 (cit. on p. 50).

Hatala, J. A., Detto, M., and Baldocchi, D. D. (2012). "Gross ecosystem photosynthesis causes a diurnal pattern in methane emission from rice: GEP CAUSES DIURNAL CH $_4$ FLUX IN RICE". In: *Geophysical Research Letters* 39.6, n/a–n/a. ISSN: 00948276. DOI: 10.1029/2012GL051303 (cit. on p. 32).

Havlin, S., Kenett, D. Y., Ben-Jacob, E., Bunde, A., Cohen, R., Hermann, H., Kantelhardt, J., Kertész, J., Kirkpatrick, S., Kurths, J., et al. (2012). "Challenges in network science: Applications to infrastructures, climate, social systems and economics". In: *The European Physical Journal Special Topics* 214.1, pp. 273–293 (cit. on p. 2).

Herlau, T., Morup, M., Schmidt, M. N., and Hansen, L. K. (2012). "Modelling dense relational data". In: *Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on*. IEEE, pp. 1–6 (cit. on p. 14).

Hlinka, J., Jajcay, N., Hartman, D., and Paluš, M. (2017). "Smooth information flow in temperature climate network reflects mass transport". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27.3, p. 035811. ISSN: 1054-1500, 1089-7682. DOI: 10.1063/1.4978028 (cit. on pp. 110, 118).

Hohn, M. E. (1991). "An Introduction to Applied Geostatistics". In: *Computers & Geosciences* 17.3, pp. 471–473. ISSN: 00983004. DOI: 10.1016/0098-3004(91)90055-I (cit. on p. 113).

Hoskins, B. J. and Karoly, D. J. (1981). "The steady linear response of a spherical atmosphere to thermal and orographic forcing". In: *Journal of the Atmospheric Sciences* 38.6, pp. 1179–1196 (cit. on p. 60).

Hsu, K.-C. and Li (2010). "Clustering spatial?temporal precipitation data using wavelet transform and self-organizing map neural network". In: *Advances in Water Resources* 33.2,

pp. 190–200. ISSN: 03091708. DOI: 10.1016/j.advwatres.2009.11.005 (cit. on pp. 90, 99).

Hu and Feng (2002). "Interannual Rainfall Variations in the North American Summer Monsoon Region: 1900–98*". In: *Journal of Climate* 15.10, pp. 1189–1202. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/1520-0442(2002)015<1189:IRVITN>2.0.CO;2 (cit. on pp. 1, 72).

Hu and Si (2016). "Technical note: Multiple wavelet coherence for untangling scale-specific and localized multivariate relationships in geosciences". In: *Hydrology and Earth System Sciences* 20.8, pp. 3183–3191. ISSN: 1607-7938. DOI: 10.5194/hess-20-3183-2016 (cit. on pp. 6, 20, 21, 36, 64).

Huang and Shukla (2005). "Ocean–Atmosphere Interactions in the Tropical and Subtropical Atlantic Ocean". In: *Journal of Climate* 18.11, pp. 1652–1672. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/JCLI3368.1 (cit. on p. 50).

Izumo, T., Vialard, J., Lengaigne, M., Boyer Montegut, C. de, Behera, S. K., Luo, J.-J., Cravatte, S., Masson, S., and Yamagata, T. (2010). "Influence of the state of the Indian Ocean Dipole on the following year's El Niño". In: *Nature Geoscience* 3.3, pp. 168–172. ISSN: 1752-0894, 1752-0908. DOI: 10.1038/ngeo760 (cit. on p. 72).

Jain, S. K., Kumar, V., and Saharia, M. (2013). "Analysis of rainfall and temperature trends in northeast India". In: *International Journal of Climatology* 33.4, pp. 968–978. ISSN: 08998418. DOI: 10.1002/joc.3483 (cit. on p. 102).

Jensen, P., Morini, M., Karsai, M., Venturini, T., Vespignani, A., Jacomy, M., Cointet, J.-P., Mercklé, P., and Fleury, E. (2016). "Detecting global bridges in networks". In: *Journal of Complex Networks* 4.3, pp. 319–329. ISSN: 2051-1310, 2051-1329. DOI: 10.1093/comnet/cnv022 (cit. on pp. 108, 111).

Jha, S. K., Zhao, H., Woldemeskel, F. M., and Sivakumar, B. (2015). "Network theory and spatial rainfall connections: An interpretation". In: *Journal of Hydrology* 527, pp. 13–19. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2015.04.035 (cit. on pp. 52, 90–92, 107).

Jiang, X. and Ting, M. (2017). "A Dipole Pattern of Summertime Rainfall across the Indian Subcontinent and the Tibetan Plateau". In: *Journal of Climate* 30.23, pp. 9607–9620. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/JCLI-D-16-0914.1 (cit. on p. 72).

Kakade, S. B. and Dugam, S. S. (2000). "The simultaneous effect of NAO and SO on the monsoon activity over India". In: *Geophysical Research Letters* 27.21, pp. 3501–3504. ISSN: 00948276. DOI: 10.1029/1999GL011201 (cit. on p. 66).

Kassim, A. H. M. and Kottegoda, N. T. (1991). "Rainfall network design through comparative kriging methods". In: *Hydrological Sciences Journal* 36.3, pp. 223–240. ISSN: 0262-6667, 2150-3435. DOI: 10.1080/02626669109492505 (cit. on p. 106).

Keum, J., Kornelsen, K. C., Leach, J. M., and Coulibaly, P. (2017). "Entropy applications to water monitoring network design: a review". In: *Entropy* 19.11, p. 613 (cit. on pp. 105, 106, 113).

Konapala and Mishra (2017). "Review of complex networks application in hydroclimatic extremes with an implementation to characterize spatio-temporal drought propagation in continental USA". In: *Journal of Hydrology* 555, pp. 600–620. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2017.10.033 (cit. on p. 64).

Kreuz, T., Bozanic, N., and Mulansky, M. (2015). "SPIKE-Synchronization: a parameter-free and time-resolved coincidence detector with an intuitive multivariate extension". In: *BMC Neuroscience* 16 (Suppl 1), P170. ISSN: 1471-2202. DOI: 10.1186/1471-2202-16-S1-P170 (cit. on p. 32).

Krishnamurthy and Krishnamurthy (2016). "Teleconnections of Indian monsoon rainfall with AMO and Atlantic tripole". In: *Climate Dynamics* 46.7, pp. 2269–2285. ISSN: 0930-7575, 1432-0894. DOI: 10.1007/s00382-015-2701-3 (cit. on pp. 64, 66).

Krishnamurthy and Shukla (2000). "Intraseasonal and interannual variability of rainfall over India". In: *Journal of Climate* 13.24, pp. 4366–4377 (cit. on p. 98).

Krishnan, R. and Sugi, M. (2003). "Pacific decadal oscillation and variability of the Indian summer monsoon rainfall". In: *Climate Dynamics* 21.3, pp. 233–242. ISSN: 0930-7575, 1432-0894. DOI: 10.1007/s00382-003-0330-8 (cit. on pp. 64, 66, 73).

Krishnan, R. and Swapna, P. (2009). "Significant Influence of the Boreal Summer Monsoon Flow on the Indian Ocean Response during Dipole Events". In: *Journal of Climate* 22.21, pp. 5611–5634. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/2009JCLI2176.1 (cit. on pp. 64, 72).

Kucharski, F., Molteni, F., and Yoo, J. H. (2006). "SST forcing of decadal Indian Monsoon rainfall variability". In: *Geophysical Research Letters* 33.3. ISSN: 0094-8276. DOI: 10.1029/2005GL025371 (cit. on p. 1).

Kucharski, F., Kang, I.-S., Straus, D., and King, M. P. (2010). "Teleconnections in the Atmosphere and Oceans". In: *Bulletin of the American Meteorological Society* 91.3, pp. 381–383. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/2009BAMS2834.1 (cit. on p. 50).

Kumar, K. K., Rajagopalan, B., Hoerling, M., Bates, G., and Cane, M. (2006). "Unraveling the Mystery of Indian Monsoon Failure During El Nino". In: *Science* 314.5796, pp. 115–119. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1131152 (cit. on p. 64).

Labat, D., Ababou, R., and Mangin, A. (2001). "Introduction of Wavelet Analyses to Rainfall/Runoffs Relationship for a Karstic Basin: The Case of Licq-Atherey Karstic System (France)". In: *Groundwater* 39.4, pp. 605–615 (cit. on p. 6).

Lafreniere, M. and Sharp, M. (2003). "Wavelet analysis of inter-annual variability in the runoff regimes of glacial and nival stream catchments, Bow Lake, Alberta". In: *Hydrological Processes* 17.6, pp. 1093–1118 (cit. on p. 6).

Laize, C. L. R. (2004). "Integration of spatial datasets to support the review of hydrometric networks and the identification of representative catchments". In: *Hydrology and Earth System Sciences* 8.6, pp. 1103–1117. ISSN: 1607-7938. DOI: 10.5194/hess-8-1103-2004 (cit. on p. 106).

Lakhanpal, A., Sehgal, V., Maheswaran, R., Khosa, R., and Sridhar, V. (2017). "A non-linear and non-stationary perspective for downscaling mean monthly temperature: a wavelet coupled second order Volterra model". In: *Stochastic Environmental Research and Risk Assessment* 31.9, pp. 2159–2181 (cit. on p. 98).

Lancichinetti, A. and Fortunato, S. (2009). "Community detection algorithms: A comparative analysis". In: *Physical Review E* 80.5. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.80.056117 (cit. on p. 95).

Langbein, W. B. (1979). "Overview of conference on hydrologic data networks". In: *Water Resources Research* 15.6, pp. 1867–1871. ISSN: 00431397. DOI: 10.1029/WR015i006p01867 (cit. on p. 105).

Lark, R. M., Milne, A. E., Addiscott, T. M., Goulding, K. W. T., Webster, C. P., and O'Flaherty, S. (2004). "Scale- and location-dependent correlation of nitrous oxide emissions with soil properties: an analysis using wavelets". In: *European Journal of Soil Science* 55.3, pp. 611–627. ISSN: 1351-0754, 1365-2389. DOI: 10.1111/j.1365-2389.2004.00620.x (cit. on p. 126).

Lark, R. M. and Webster, R. (2001). "Changes in variance and correlation of soil properties with scale and location: analysis using an adapted maximal overlap discrete wavelet

transform". In: *European Journal of Soil Science* 52.4, pp. 547–562. ISSN: 1351-0754, 1365-2389. DOI: 10.1046/j.1365-2389.2001.00420.x (cit. on pp. 32, 34).

Li, Gao, Z., and Zhao, X. (2008). "Multiple scale analysis of complex networks using the empirical mode decomposition method". In: *Physica A: Statistical Mechanics and its Applications* 387.12, pp. 2981–2986. ISSN: 03784371. DOI: 10.1016/j.physa.2008.01.036 (cit. on p. 42).

Liang, Z., Ren, Y., Yan, J., Li, D., Voss, L. J., Sleigh, J. W., and Li, X. (2016). "A comparison of different synchronization measures in electroencephalogram during propofol anesthesia". In: *Journal of clinical monitoring and computing* 30.4, pp. 451–466 (cit. on p. 14).

Liu, J., Xiong, Q., Shi, W., Shi, X., and Wang, K. (2016). "Evaluating the importance of nodes in complex networks". In: *Physica A: Statistical Mechanics and its Applications* 452, pp. 209–219. ISSN: 03784371. DOI: 10.1016/j.physa.2016.02.049 (cit. on pp. 108, 109, 112).

Livina, V. N., Edwards, N. R., Goswami, S., and Lenton, T. M. (2008). "A wavelet-coefficient score for comparison of two-dimensional climatic-data fields". In: *Quarterly Journal of the Royal Meteorological Society* 134.633, pp. 941–955. ISSN: 00359009, 1477870X. DOI: 10.1002/qj.261 (cit. on p. 32).

Looney, D., Hemakom, A., and Mandic, D. P. (2014). "Intrinsic multi-scale analysis: a multivariate empirical mode decomposition framework". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 471.2173, pp. 20140709–20140709. ISSN: 1364-5021, 1471-2946. DOI: 10.1098/rspa.2014.0709 (cit. on p. 32).

Lovejoy, S. (2015). "A voyage through scales, a missing quadrillion and why the climate is not what you expect". In: *Climate Dynamics* 44.11, pp. 3187–3210. ISSN: 0930-7575, 1432-0894. DOI: 10.1007/s00382-014-2324-0 (cit. on p. 4).

Luo, J.-J., Zhang, R., Behera, S. K., Masumoto, Y., Jin, F.-F., Lukas, R., and Yamagata, T. (2010). "Interaction between El Niño and Extreme Indian Ocean Dipole". In: *Journal of Climate* 23.3, pp. 726–742. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/2009JCLI3104.1 (cit. on pp. 42, 59, 72, 81).

Luterbacher, J., Xoplaki, E., Casty, C., Wanner, H., Pauling, A., Küttel, M., Rutishauser, T., Brönnimann, S., Fischer, E., Fleitmann, D., Gonzalez-Rouco, F. J., García-Herrera, R., Barriendos, M., Rodrigo, F., Gonzalez-Hidalgo, J. C., Saz, M. A., Gimeno, L., Ribera, P., Brunet, M., Paeth, H., Rimbu, N., Felis, T., Jacobeit, J., Dünkeloh, A., Zorita, E., Guiot, J., Türkes, M., Alcoforado, M. J., Trigo, R., Wheeler, D., Tett, S., Mann, M. E., Touchan, R., Shindell, D. T., Silenzi, S., Montagna, P., Camuffo, D., Mariotti, A., Nanni, T., Brunetti, M., Maugeri, M., Zerefos, C., Zolt, S. D., Lionello, P., Nunes, M. F., Rath, V., Beltrami, H., Garnier, E., and Ladurie, E. L. R. (2006). "Chapter 1 Mediterranean climate variability over the last centuries: A review". In: *Developments in Earth and Environmental Sciences*. Vol. 4. Elsevier, pp. 27–148. ISBN: 978-0-444-52170-5. DOI: 10.1016/S1571-9197(06)80004-2 (cit. on p. 64).

Malik, N., Bookhagen, B., Marwan, N., and Kurths, J. (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, 1432-0894. DOI: 10.1007/s00382-011-1156-4 (cit. on pp. 14, 18, 23, 32, 64, 91, 107).

Malik, N., Bookhagen, B., and Mucha, P. J. (2016). "Spatiotemporal patterns and trends of Indian monsoonal rainfall extremes". In: *Geophysical Research Letters* 43.4, pp. 1710–1717. ISSN: 00948276. DOI: 10.1002/2016GL067841 (cit. on pp. 1, 94, 99).

Mallat, S. G. (1989). "A theory for multiresolution signal decomposition: the wavelet representation". In: *IEEE transactions on pattern analysis and machine intelligence* 11.7, pp. 674–693 (cit. on pp. 16, 26).

Mamalakis, A., Yu, J.-Y., Randerson, J. T., AghaKouchak, A., and Foufoula-Georgiou, E. (2018). "A new interhemispheric teleconnection increases predictability of winter precipitation in southwestern US". en. In: *Nature Communications* 9.1. ISSN: 2041-1723. DOI: 10.1038/s41467-018-04722-7 (cit. on pp. 50, 60).

Maraun, D. (2005). "Epochs of phase coherence between El Niño/Southern Oscillation and Indian monsoon". In: *Geophysical Research Letters* 32.15. ISSN: 0094-8276. DOI: 10.1029/2005GL023225 (cit. on p. 13).

Marwan, N., Donges, J. F., Zou, Y., Donner, R. V., and Kurths, J. (2009). "Complex network approach for recurrence analysis of time series". In: *Physics Letters A* 373.46, pp. 4246–4254. ISSN: 03759601. DOI: 10.1016/j.physleta.2009.09.042 (cit. on p. 50).

Marwan, N., Feldhoff, J. H., Donner, R. V., Donges, J. F., and Kurths, J. (2014). "Detection of coupling directions with intersystem recurrence networks". In: *IEICE Proceeding Series* 1, pp. 231–234. ISSN: 2188-5079. DOI: 10.15248/proc.1.231 (cit. on p. 32).

Marwan, N., Romano, M. C., Thiel, M., and Kurths, J. (2007). "Recurrence plots for the analysis of complex systems". In: *Physics reports* 438.5-6, pp. 237–329 (cit. on p. 14).

McGregor, S., Timmermann, A., England, M. H., Elison Timm, O., and Wittenberg, A. T. (2013). "Inferred changes in El Niño–Southern Oscillation variance over the past six centuries". In: *Climate of the Past* 9.5, pp. 2269–2284. ISSN: 1814-9332. DOI: 10.5194/cp-9-2269-2013 (cit. on p. 72).

Menck, P. J., Heitzig, J., Kurths, J., and Joachim Schellnhuber, H. (2014). "How dead ends undermine power grid stability". In: *Nature Communications* 5. ISSN: 2041-1723. DOI: 10.1038/ncomms4969 (cit. on p. 106).

Miralles, D. G., Teuling, A. J., Heerwaarden, C. C. van, and Vilà-Guerau de Arellano, J. (2014). "Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation". In: *Nature Geoscience* 7.5, pp. 345–349. ISSN: 1752-0894, 1752-0908. DOI: 10.1038/ngeo2141 (cit. on pp. 1, 2, 32, 64).

Miritello, G., Moro, E., Lara, R., Martinez-Lopez, R., Belchamber, J., Roberts, S. G., and Dunbar, R. I. (2013). "Time as a limited resource: Communication strategy in mobile phone networks". In: *Social Networks* 35.1, pp. 89–95 (cit. on p. 14).

Mishra and Coulibaly (2009). "Developments in hydrometric network design: A review". In: *Reviews of Geophysics* 47.2. ISSN: 8755-1209. DOI: 10.1029/2007RG000243 (cit. on pp. 90, 106).

Mishra, Smoliak, B. V., Lettenmaier, D. P., and Wallace, J. M. (2012). "A prominent pattern of year-to-year variability in Indian Summer Monsoon Rainfall". In: *Proceedings of the National Academy of Sciences* 109.19, pp. 7213–7217. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1119150109 (cit. on p. 72).

Mitra, C., Kittel, T., Choudhary, A., Kurths, J., and Donner, R. V. (2017a). "Recovery time after localized perturbations in complex dynamical networks". In: *New Journal of Physics* 19.10, p. 103004. ISSN: 1367-2630. DOI: 10.1088/1367-2630/aa7fab (cit. on p. 91).

Mitra, C., Kurths, J., and Donner, R. V. (2017b). "Rewiring hierarchical scale-free networks: Influence on synchronizability and topology". In: *EPL (Europhysics Letters)* 119.3, p. 30002 (cit. on p. 6).

Mokhov, I. I., Smirnov, D. A., Nakonechny, P. I., Kozlenko, S. S., and Kurths, J. (2012). "Relationship between El-Niño/Southern Oscillation and the Indian monsoon". In: *Izvestiya, Atmospheric and Oceanic Physics* 48.1, pp. 47–56. ISSN: 0001-4338, 1555-628X. DOI: 10.1134/S0001433812010082 (cit. on pp. 64, 66).

Mokhov, I. I., Smirnov, D. A., Nakonechny, P. I., Kozlenko, S. S., Seleznev, E. P., and Kurths, J. (2011). "Alternating mutual influence of El-Niño/Southern Oscillation and

Indian monsoon: INDIAN MONSOON-ENSO MUTUAL INFLUENCE". In: *Geophysical Research Letters* 38.8, n/a–n/a. ISSN: 00948276. DOI: 10.1029/2010GL045932 (cit. on p. 13).

Molini, A., Katul, G. G., and Porporato, A. (2010). "Causality across rainfall time scales revealed by continuous wavelet transforms". In: *Journal of Geophysical Research* 115 (D14). ISSN: 0148-0227. DOI: 10.1029/2009JD013016 (cit. on pp. 2, 32, 130).

Newman (2003). "The Structure and Function of Complex Networks". In: *SIAM Review* 45.2, pp. 167–256. ISSN: 0036-1445, 1095-7200. DOI: 10.1137/S003614450342480 (cit. on p. 50).

– (2004). "Detecting community structure in networks". In: *The European Physical Journal B - Condensed Matter* 38.2, pp. 321–330. ISSN: 1434-6028, 1434-6036. DOI: 10.1140/epjb/e2004-00124-y (cit. on p. 95).

Newman, Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Di Lorenzo, E., Mantua, N. J., Miller, A. J., Minobe, S., Nakamura, H., Schneider, N., Vimont, D. J., Phillips, A. S., Scott, J. D., and Smith, C. A. (2016). "The Pacific Decadal Oscillation, Revisited". In: *Journal of Climate* 29.12, pp. 4399–4427. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/JCLI-D-15-0508.1 (cit. on p. 42).

Nitzbon, J., Schultz, P., Heitzig, J., Kurths, J., and Hellmann, F. (2017). "Deciphering the imprint of topology on nonlinear dynamical network stability". In: *New Journal of Physics* 19.3, p. 033029 (cit. on p. 6).

Niu, J. (2013). "Precipitation in the Pearl River basin, South China: scaling, regional patterns, and influence of large-scale climate anomalies". In: *Stochastic Environmental Research and Risk Assessment* 27.5, pp. 1253–1268. ISSN: 1436-3240, 1436-3259. DOI: 10.1007/s00477-012-0661-2 (cit. on p. 90).

O'Connor, J. M., Pretorius, P. H., Johnson, K., and King, M. A. (2013). "A method to synchronize signals from multiple patient monitoring devices through a single input channel for inclusion in list-mode acquisitions: Method to synchronize multiple patient monitoring devices signals". In: *Medical Physics* 40.12, p. 122502. ISSN: 00942405. DOI: 10.1118/1.4828844 (cit. on pp. 6, 14).

Oesterle, H. (2001). "Reconstruction of daily global radiation for past years for use in agricultural models". In: *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere* 26.3, pp. 253–256. ISSN: 14641909. DOI: 10.1016/S1464-1909(00)00248-3 (cit. on pp. 23, 29, 114).

Okamoto, K., Chen, W., and Li, X.-Y. (2008). "Ranking of Closeness Centrality for Large-Scale Social Networks". In: *Frontiers in Algorithmics*. Ed. by F. P. Preparata, X. Wu, and J. Yin. Vol. 5059. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 186–195. ISBN: 978-3-540-69310-9 978-3-540-69311-6. DOI: 10.1007/978-3-540-69311-6_21 (cit. on p. 109).

Okin, G. S., Parsons, A. J., Wainwright, J., Herrick, J. E., Bestelmeyer, B. T., Peters, D. C., and Fredrickson, E. L. (2009). "Do Changes in Connectivity Explain Desertification?" In: *BioScience* 59.3, pp. 237–244. ISSN: 1525-3244, 0006-3568. DOI: 10.1525/bio.2009.59.3.8 (cit. on pp. 32, 64).

Özger, M., Mishra, A. K., and Singh, V. P. (2010). "Scaling characteristics of precipitation data in conjunction with wavelet analysis". In: *Journal of Hydrology* 395.3, pp. 279–288. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2010.10.039 (cit. on p. 90).

Ozturk, U., Marwan, N., Korup, O., Saito, H., Agarwal, A., Grossman, M. J., Zaiki, M., and Kurths, J. (2018a). "Complex networks for tracking extreme rainfall during typhoons". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28.7, p. 075301. ISSN: 1054-1500, 1089-7682. DOI: 10.1063/1.5004480 (cit. on pp. 65, 92, 106, 107).

Ozturk, U., Marwan, N., Specht, S., Korup, O., and Jensen, J. (2018b). "A new centennial sea-level record for Antalya, eastern Mediterranean". In: *Journal of Geophysical Research: Oceans* (cit. on p. 49).

Ozturk, U., Wendi, D., Crisologo, I., Riemer, A., Agarwal, A., Vogel, K., López-Tarazón, J. A., and Korup, O. (2018c). "Rare flash floods and debris flows in southern Germany". In: *Science of the Total Environment* 626, pp. 941–952 (cit. on p. 106).

Pai, D., Sridhar, L., Badwaik, M., and Rajeevan, M. (2015). "Analysis of the daily rainfall events over India using a new long period (1901–2010) high resolution (0.25° × 0.25°) gridded rainfall data set". In: *Climate Dynamics* 45.3, pp. 755–776. ISSN: 0930-7575, 1432-0894. DOI: 10.1007/s00382-014-2307-1 (cit. on p. 63).

Pai, D., Sridhar, L., Rajeevan, M., Sreejith, O., satbhai, N., and Mukhopadyay, B. (2014). "Development of a new high spatial resolution (0.25° × 0.25°) Long Period (1901-2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region". In: *Mausam* 65.1, pp. 1–18 (cit. on pp. 66, 98).

Paliwal, A. (2017). "Why India struggle to predict precipitation over its lands". In: *Science* (cit. on p. 63).

Paluš, M. (2014a). "Cross-Scale Interactions and Information Transfer". In: *Entropy* 16.10, pp. 5263–5289. ISSN: 1099-4300. DOI: 10.3390/e16105263 (cit. on pp. 2, 3, 64).

– (2014b). "Multiscale atmospheric dynamics: cross-frequency phase-amplitude coupling in the air temperature". In: *Physical review letters* 112.7, p. 078702 (cit. on p. 79).

– (2018). "Linked by Dynamics: Wavelet-Based Mutual Information Rate as a Connectivity Measure and Scale-Specific Networks". In: *Advances in Nonlinear Geosciences*. Ed. by A. A. Tsonis. Cham: Springer International Publishing, pp. 427–463. ISBN: 978-3-319-58894-0 978-3-319-58895-7. DOI: 10.1007/978-3-319-58895-7_21 (cit. on pp. 6, 32, 50, 64, 92, 106, 107, 110).

Pardo-Igúzquiza, E. (1998). "Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing". In: *Journal of Hydrology* 210.1, pp. 206–220. ISSN: 00221694. DOI: 10.1016/S0022-1694(98)00188-7 (cit. on pp. 90, 106).

Percival, D. B. (2008). "Analysis of Geophysical Time Series Using Discrete Wavelet Transforms: An Overview". In: *Nonlinear Time Series Analysis in the Geosciences*. Ed. by R. V. Donner and S. M. Barbosa. Vol. 112. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 61–79. ISBN: 978-3-540-78937-6 978-3-540-78938-3. DOI: 10.1007/978-3-540-78938-3_4 (cit. on p. 33).

Percival, D. B. and Walden, A. T. (2000). *Wavelet Methods for Time SeriesAnalysis*. Cambridge: Cambridge University Press. ISBN: 978-0-511-84104-0. DOI: 10.1017/CBO9780511841040 (cit. on pp. 41, 70).

Perra, N., Goncalves, B., Pastor-Satorras, R., and Vespignani, A. (2012). "Activity driven modeling of time varying networks". In: *Scientific reports* 2, p. 469 (cit. on p. 14).

Peters, D. P. C., Pielke, R. A., Bestelmeyer, B. T., Allen, C. D., Munson-McGee, S., and Havstad, K. M. (2004). "Cross-scale interactions, nonlinearities, and forecasting catastrophic events". In: *Proceedings of the National Academy of Sciences* 101.42, pp. 15130–15135. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0403822101 (cit. on pp. 2, 64, 68).

Peters, D. P. C., Bestelmeyer, B. T., and Turner, M. G. (2007). "Cross–Scale Interactions and Changing Pattern–Process Relationships: Consequences for System Dynamics". In: *Ecosystems* 10.5, pp. 790–796. ISSN: 1432-9840, 1435-0629. DOI: 10.1007/s10021-007-9055-6 (cit. on pp. 4, 32, 64).

Pfurtscheller, G. and Lopes da Silva, F. (1999). "Event-related EEG/MEG synchronization and desynchronization: basic principles". In: *Clinical Neurophysiology* 110.11, pp. 1842–1857. ISSN: 13882457. DOI: 10.1016/S1388-2457(99)00141-8 (cit. on pp. 6, 14, 92, 107).

Pillai, P. A. and Mohankumar, K. (2010). "Individual and combined influence of El Niño-Southern Oscillation and Indian Ocean Dipole on the Tropospheric Biennial Oscillation". In: *Quarterly Journal of the Royal Meteorological Society*, n/a–n/a. ISSN: 00359009, 1477870X. DOI: 10.1002/qj.579 (cit. on p. 72).

Pillai, P. A., Rao, S. A., Ramu, D. A., Pradhan, M., and George, G. (2018). "Seasonal prediction skill of Indian summer monsoon rainfall in NMME models and monsoon mission CFSv2". In: *International Journal of Climatology*. ISSN: 08998418. DOI: 10.1002/joc.5413 (cit. on p. 64).

Podtaev, S., Morozov, M., and Frick, P. (2008). "Wavelet-based Correlations of Skin Temperature and Blood Flow Oscillations". In: *Cardiovascular Engineering* 8.3, pp. 185–189. ISSN: 1567-8822, 1573-6806. DOI: 10.1007/s10558-008-9055-y (cit. on p. 32).

Polikar, R. (1996). *The wavelet tutorial* (cit. on p. 40).

Putthividhya, A. and Tanaka, K. (2012). "Optimal Rain Gauge Network Design and Spatial Precipitation Mapping based on Geostatistical Analysis from Colocated Elevation and Humidity Data". In: *International Journal of Environmental Science and Development*, pp. 124–129. ISSN: 20100264. DOI: 10.7763/IJESD.2012.V3.201 (cit. on pp. 1, 106).

Quiroga, Arnhold, J., and Grassberger, P. (2000). "Learning driver-response relationships from synchronization patterns". In: *Physical Review E* 61.5, pp. 5142–5148. ISSN: 1063-651X, 1095-3787. DOI: 10.1103/PhysRevE.61.5142 (cit. on pp. 14, 91).

Quiroga, Kraskov, A., Kreuz, T., and Grassberger, P. (2002a). "Performance of different synchronization measures in real data: A case study on electroencephalographic signals". In: *Physical Review E* 65.4. ISSN: 1063-651X, 1095-3787. DOI: 10.1103/PhysRevE.65.041903 (cit. on pp. 14, 17, 67, 91).

Quiroga, Kreuz, and Grassberger (2002b). "Event synchronization: A simple and fast method to measure synchronicity and time delay patterns". In: *Physical Review E* 66.4. ISSN: 1063-651X, 1095-3787. DOI: 10.1103/PhysRevE.66.041904 (cit. on pp. 93, 129).

Quyen, M. L. V., Martinerie, J., Adam, C., and Varela, F. J. (1999). "Nonlinear analyses of interictal EEG map the brain interdependences in human focal epilepsy". In: *Physica D: Nonlinear Phenomena* 127.3-4, pp. 250–266 (cit. on p. 14).

Rahmstorf, S. (2002). "Ocean circulation and climate during the past 120,000 years". In: *Nature* 419.6903, pp. 207–214. ISSN: 0028-0836. DOI: 10.1038/nature01090 (cit. on pp. 60, 82).

Rai, P. and Dimri, A. P. (2017). "Effect of changing tropical easterly jet, low level jet and quasi-biennial oscillation phases on Indian summer monsoon: TEJ, LLJ and QBO phases and Indian summer monsoon". In: *Atmospheric Science Letters* 18.2, pp. 52–59. ISSN: 1530261X. DOI: 10.1002/asl.723 (cit. on p. 72).

Rajeevan, M. and Pai, D. (2007). "On the El Niño-Indian monsoon predictive relationships". In: *Geophysical Research Letters* 34.4. ISSN: 0094-8276. DOI: 10.1029/2006GL028916 (cit. on p. 72).

Ramirez, E., Silva Dias, P. L. da, and Raupp, C. F. M. (2017). "Multiscale Atmosphere–Ocean Interactions and the Low-Frequency Variability in the Equatorial Region". en. In: *Journal of the Atmospheric Sciences* 74.8, pp. 2503–2523. ISSN: 0022-4928, 1520-0469. DOI: 10.1175/JAS-D-15-0325.1 (cit. on p. 50).

Rani, D. and Moreira, M. M. (2010). "Simulation–Optimization Modeling: A Survey and Potential Application in Reservoir Systems Operation". In: *Water Resources Management*

24.6, pp. 1107–1138. ISSN: 0920-4741, 1573-1650. DOI: 10.1007/s11269-009-9488-0 (cit. on p. 106).

Rao, S. (2015). "Indian Ocean Dipole and the monsoon: The Joker in the forecast pack". In: *The IndianEXPRESS* (cit. on p. 64).

Rathinasamy, M., Agarwal, A., Parmar, V., Khosa, R., and Bairwa, A. (2017). "Partial wavelet coherence analysis for understanding the standalone relationship between Indian Precipitation and Teleconnection patterns". In: *arXiv preprint arXiv:1702.06568* (cit. on p. 73).

Rathinasamy, M. and Khosa, R. (2012a). *Multiscale Hydrological Forecasting Using Wavelet Volterra Coupled Models*. OCLC: 864016860. Saarbrücken: LAP LAMBERT Academic Publishing. ISBN: 978-3-659-25613-4 (cit. on p. 4).

– (2012b). "Multiscale nonlinear model for monthly streamflow forecasting: a wavelet-based approach". In: *Journal of Hydroinformatics* 14.2, p. 424. ISSN: 1464-7141. DOI: 10.2166/hydro.2011.130 (cit. on p. 20).

Rathinasamy, M., Khosa, R., Adamowski, J., ch, S., Partheepan, G., Anand, J., and Narsimlu, B. (2014). "Wavelet-based multiscale performance analysis: An approach to assess and improve hydrological models". In: *Water Resources Research* 50.12, pp. 9721–9737. ISSN: 00431397. DOI: 10.1002/2013WR014650 (cit. on pp. 2, 14, 23, 32, 33).

Razavi, T. and Coulibaly (2013). "Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods". In: *Journal of Hydrologic Engineering* 18.8, pp. 958–975. ISSN: 1084-0699, 1943-5584. DOI: 10.1061/(ASCE)HE.1943-5584.0000690 (cit. on p. 90).

Reiner, A., Yekutieli, D., and Benjamini, Y. (2003). "Identifying differentially expressed genes using false discovery rate controlling procedures". en. In: *Bioinformatics* 19.3, pp. 368–375. ISSN: 1367-4803, 1460-2059. DOI: 10.1093/bioinformatics/btf877 (cit. on p. 52).

Rheinwalt, A., Boers, N., Marwan, N., Kurths, J., Hoffmann, P., Gerstengarbe, F.-W., and Werner, P. (2016). "Non-linear time series analysis of precipitation events using regional climate networks for Germany". In: *Climate dynamics* 46.3-4, pp. 1065–1074 (cit. on pp. 14, 18, 19, 50, 93, 114).

Rheinwalt, A., Goswami, B., Boers, N., Heitzig, J., Marwan, N., Krishnan, R., and Kurths, J. (2015). "Teleconnections in Climate Networks: A Network-of-Networks Approach to Investigate the Influence of Sea Surface Temperature Variability on Monsoon Systems". In: *Machine Learning and Data Mining Approaches to Climate Science*. Ed. by V. Lakshmanan, E. Gilleland, A. McGovern, and M. Tingley. Cham: Springer International Publishing, pp. 23–33. ISBN: 978-3-319-17219-4 978-3-319-17220-0. DOI: 10.1007/978-3-319-17220-0_3 (cit. on pp. 14, 32).

Rial, J. (2012). "Synchronization of polar climate variability over the last ice age: in search of simple rules at the heart of climate's complexity". In: *American Journal of Science* 312.4, pp. 417–448 (cit. on p. 13).

Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P., and Van De Ville, D. (2011). "Decoding brain states from fMRI connectivity graphs". In: *NeuroImage* 56.2, pp. 616–626. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2010.05.081 (cit. on p. 32).

Rosenblum, M. G., Pikovsky, A. S., and Kurths, J. (1997). "From phase to lag synchronization in coupled chaotic oscillators". In: *Physical Review Letters* 78.22, p. 4193 (cit. on p. 14).

Rosvall, M. and Bergstrom, C. T. (2007). "An information-theoretic framework for resolving community structure in complex networks". In: *Proceedings of the National Academy of Sciences* 104.18, pp. 7327–7331. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0611034104 (cit. on p. 32).

Rubinov, M. and Sporns, O. (2011). "Weight-conserving characterization of complex functional brain networks". In: *NeuroImage* 56.4, pp. 2068–2079. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2011.03.069 (cit. on pp. 2, 91, 95, 107).

Runge, J., Petoukhov, V., Donges, J. F., Hlinka, J., Jajcay, N., Vejmelka, M., Hartman, D., Marwan, N., Palus, M., and Kurths, J. (2015). "Identifying causal gateways and mediators in complex spatio-temporal systems". In: *Nature Communications* 6, p. 8502. ISSN: 2041-1723. DOI: 10.1038/ncomms9502 (cit. on pp. 6, 50).

Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., and Blöschl, G. (2013). "Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies". In: *Hydrology and Earth System Sciences* 17.7, pp. 2637–2652. ISSN: 1607-7938. DOI: 10.5194/hess-17-2637-2013 (cit. on p. 90).

Saxena, A., Malik, V., and Iyengar, S. R. S. (2016). "Estimating the degree centrality ranking". In: IEEE, pp. 1–2. ISBN: 978-1-4673-9622-6. DOI: 10.1109/COMSNETS.2016.7440022 (cit. on pp. 90, 109).

Scarsoglio, S., Laio, F., and Ridolfi, L. (2013). "Climate Dynamics: A Network-Based Approach for the Analysis of Global Precipitation". en. In: *PLoS ONE* 8.8. Ed. by A. J. Cannon, e71129. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0071129 (cit. on p. 53).

Schiermeier, Q. (2018). "Droughts, heatwaves and floods: How to tell when climate change is to blame". In: *Nature* 560.7716, pp. 20–22. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/d41586-018-05849-9 (cit. on p. 1).

Schiff, S. J., So, P., Chang, T., Burke, R. E., and Sauer, T. (1996). "Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble". In: *Physical Review E* 54.6, pp. 6708–6724. ISSN: 1063-651X, 1095-3787. DOI: 10.1103/PhysRevE.54.6708 (cit. on p. 14).

Schmocker-Fackel, P. and Naef, F. (2010). "Changes in flood frequencies in Switzerland since 1500". In: *Hydrology and Earth System Sciences* 14.8, pp. 1581–1594 (cit. on p. 13).

Schultz, P., Heitzig, J., and Kurths, J. (2014). "A random growth model for power grids and other spatially embedded infrastructure networks". In: *The European Physical Journal Special Topics* 223.12, pp. 2593–2610. ISSN: 1951-6355, 1951-6401. DOI: 10.1140/epjst/e2014-02279-6 (cit. on p. 106).

Schultz, P., Hellmann, F., Heitzig, J., and Kurths, J. (2016). "A Network of Networks Approach to Interconnected Power Grids". In: *arXiv preprint arXiv:1701.06968* (cit. on p. 6).

Sehgal, V., Lakhanpal, A., Maheswaran, R., Khosa, R., and Sridhar, V. (2016). "Application of multi-scale wavelet entropy and multi-resolution Volterra models for climatic downscaling". In: *Journal of Hydrology* (cit. on pp. 98, 106).

Shukla, R., Agarwal, A., Sachdeva, K., Kurths, J., and Joshi, P. (2018). "Climate change perception: an analysis of climate change and risk perceptions among farmer types of Indian Western Himalayas". In: *Climatic Change*, pp. 1–17 (cit. on p. 64).

Shukla, Chakraborty, A., and Joshi, P. (2017). "Vulnerability of agro-ecological zones in India under the earth system climate model scenarios". In: *Mitigation and adaptation strategies for global change* 22.3, pp. 399–425 (cit. on p. 63).

Shukla, Chakraborty, A., Sachdeva, K., and Joshi, P. (2018). "Agriculture in the western Himalayas–an asset turning into a liability". In: *Development in Practice* 28.2, pp. 318–324 (cit. on p. 64).

Shusterman, E. and Feder, M. (1994). "Correlation between the wavelet decomposition components". In: IEEE, pp. 405–408. ISBN: 978-0-7803-2127-4. DOI: 10.1109/TFSA.1994.467329 (cit. on p. 32).

Sivakumar, B. and Woldemeskel, F. M. (2014). "Complex networks for streamflow dynamics". In: *Hydrology and Earth System Sciences* 18.11, pp. 4565–4578. ISSN: 1607-7938. DOI: 10.5194/hess-18-4565-2014 (cit. on pp. 50, 52, 92, 107).

Sivakumar, B., Singh, V. P., Berndtsson, R., and Khan, S. K. (2015). "Catchment Classification Framework in Hydrology: Challenges and Directions". In: *Journal of Hydrologic Engineering* 20.1, A4014002. ISSN: 1084-0699, 1943-5584. DOI: 10.1061/(ASCE)HE.1943-5584.0000837 (cit. on p. 90).

Smith, A., Sampson, C., and Bates, P. (2015). "Regional flood frequency analysis at the global scale". In: *Water Resources Research* 51.1, pp. 539–553. ISSN: 00431397. DOI: 10.1002/2014WR015814 (cit. on p. 90).

Solomon, S., Qin, D., and Manning, M. (2007). *Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge New York (cit. on p. 49).

Stam, C. J. and Reijneveld, J. C. (2007). "Graph theoretical analysis of complex networks in the brain". In: *Nonlinear biomedical physics* 1.1, p. 3 (cit. on p. 2).

Steinbach, M., Tan, P.-N., Kumar, V., Klooster, S., and Potter, C. (2003). "Discovery of climate indices using clustering". In: ACM Press, p. 446. ISBN: 978-1-58113-737-8. DOI: 10.1145/956750.956801 (cit. on p. 2).

Steinhaeuser, K., Chawla, N. V., and Ganguly, A. R. (2010). "An exploration of climate data using complex networks". In: *ACM SIGKDD Explorations Newsletter* 12.1, p. 25. ISSN: 19310145. DOI: 10.1145/1882471.1882476 (cit. on p. 91).

– (2011). "Complex networks as a unified framework for descriptive analysis and predictive modeling in climate science". In: *Statistical Analysis and Data Mining* 4.5, pp. 497–511. ISSN: 19321864. DOI: 10.1002/sam.10100 (cit. on pp. 35, 92).

Steinhaeuser, K., Ganguly, A. R., and Chawla, N. V. (2012). "Multivariate and multiscale dependence in the global climate system revealed through complex networks". In: *Climate Dynamics* 39.3, pp. 889–895. ISSN: 0930-7575, 1432-0894. DOI: 10.1007/s00382-011-1135-9 (cit. on pp. 2, 3, 5, 14, 32, 50, 53, 64).

Stolbova, V., Martin, P., Bookhagen, B., Marwan, N., and Kurths, J. (2014). "Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka". In: *Nonlinear Processes in Geophysics* 21.4, pp. 901–917. ISSN: 1607-7946. DOI: 10.5194/npg-21-901-2014 (cit. on pp. 6, 7, 14, 17, 18, 32, 50–52, 65, 91–93, 107, 108).

Stolbova, V., Surovyatkina, E., Bookhagen, B., and Kurths, J. (2016). "Tipping elements of the Indian monsoon: Prediction of onset and withdrawal". In: *Geophysical Research Letters* 43.8, pp. 3982–3990. ISSN: 00948276. DOI: 10.1002/2016GL068392 (cit. on pp. 64, 118).

Stosic, T., Stosic, B., and Singh, V. P. (2017). "Optimizing streamflow monitoring networks using joint permutation entropy". In: *Journal of Hydrology* 552, pp. 306–312 (cit. on p. 106).

Stuecker, M. F., Timmermann, A., Jin, F.-F., Chikamoto, Y., Zhang, W., Wittenberg, A. T., Widiasih, E., and Zhao, S. (2017). "Revisiting ENSO/Indian Ocean Dipole phase relationships". In: *Geophysical Research Letters* 44.5, pp. 2481–2492. ISSN: 00948276. DOI: 10.1002/2016GL072308 (cit. on pp. 42, 59, 81).

Sturtevant, C., Ruddell, B. L., Knox, S. H., Verfaillie, J., Matthes, J. H., Oikawa, P. Y., and Baldocchi, D. (2016). "Identifying scale-emergent, nonlinear, asynchronous processes of wetland methane exchange". In: *Journal of Geophysical Research: Biogeosciences* 121.1, pp. 188–204. ISSN: 21698953. DOI: 10.1002/2015JG003054 (cit. on pp. 32, 130).

Sun, A. Y., Xia, Y., Caldwell, T. G., and Hao, Z. (2018). "Patterns of precipitation and soil moisture extremes in Texas, US: A complex network analysis". In: *Advances in Water Resources* 112, pp. 203–213. ISSN: 03091708. DOI: 10.1016/j.advwatres.2017.12.019 (cit. on p. 65).

Tan, X., Gan, T. Y., and Shao, D. (2016). "Wavelet analysis of precipitation extremes over Canadian ecoregions and teleconnections to large-scale climate anomalies: Large Precipitation and Climate Anomalies". In: *Journal of Geophysical Research: Atmospheres* 121.24, pp. 14, 469–14, 486. ISSN: 2169897X. DOI: 10.1002/2016JD025533 (cit. on pp. 64, 68).

Taormina, R., Chau, K.-W., and Sivakumar, B. (2015). "Neural network river forecasting through baseflow separation and binary-coded swarm optimization". In: *Journal of Hydrology* 529, pp. 1788–1797. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2015.08.008 (cit. on p. 106).

Tass, P., Rosenblum, M. G., Weule, J., Kurths, J., Pikovsky, A., Volkmann, J., Schnitzler, A., and Freund, H.-J. (1998). "Detection of n : m Phase Locking from Noisy Data: Application to Magnetoencephalography". In: *Physical Review Letters* 81.15, pp. 3291–3294. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.81.3291 (cit. on pp. 14, 17, 18).

Tobler, W. R. (1970). "A Computer Movie Simulating Urban Growth in the Detroit Region". In: *Economic Geography* 46, p. 234. ISSN: 00130095. DOI: 10.2307/143141 (cit. on pp. 55, 90, 117).

Trenberth, K. E. and Shea, D. J. (2005). "Relationships between precipitation and surface temperature". In: *Geophysical Research Letters* 32.14, n/a–n/a. ISSN: 00948276. DOI: 10.1029/2005GL022760 (cit. on p. 1).

Tsintikidis, D., Georgakakos, K. P., Sperfslage, J. A., Smith, D. E., and Carpenter, T. M. (2002). "Precipitation Uncertainty and Raingauge Network Design within Folsom Lake Watershed". In: *Journal of Hydrologic Engineering* 7.2, pp. 175–184. ISSN: 1084-0699, 1943-5584. DOI: 10.1061/(ASCE)1084-0699(2002)7:2(175) (cit. on p. 106).

Tsonis and Roebber (2004). "The architecture of the climate network". In: *Physica A: Statistical Mechanics and its Applications* 333, pp. 497–504. ISSN: 03784371. DOI: 10.1016/j.physa.2003.10.045 (cit. on p. 2).

Tsonis, Swanson, K. L., and Roebber, P. J. (2006). "What Do Networks Have to Do with Climate?" In: *Bulletin of the American Meteorological Society* 87.5, pp. 585–595. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/BAMS-87-5-585 (cit. on pp. 32, 50).

Tsonis, Swanson, K. L., and Wang, G. (2008). "On the Role of Atmospheric Teleconnections in Climate". In: *Journal of Climate* 21.12, pp. 2990–3001. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/2007JCLI1907.1 (cit. on p. 50).

Tsonis, Swanson, K., and Kravtsov, S. (2007). "A new dynamical mechanism for major climate shifts". In: *Geophysical Research Letters* 34.13, n/a–n/a. ISSN: 00948276. DOI: 10.1029/2007GL030288 (cit. on p. 2).

Tsonis, Wang, G., Swanson, L., Rodrigues, A., and Costa, F. (2011). "Community structure and dynamics in climate networks". In: *Climate Dynamics* 37.5, pp. 933–940. ISSN: 0930-7575, 1432-0894. DOI: 10.1007/s00382-010-0874-3 (cit. on p. 91).

Van Brummelen, G. (2013). *Heavenly mathematics: the forgotten art of spherical trigonometry*. Princeton ; Oxford: Princeton University Press. ISBN: 978-0-691-14892-2 (cit. on p. 54).

Varotsos, C. A., Efstathiou, M. N., and Cracknell, A. P. (2013). "On the scaling effect in global surface air temperature anomalies". In: *Atmospheric Chemistry and Physics* 13.10, pp. 5243–5253. ISSN: 1680-7324. DOI: 10.5194/acp-13-5243-2013 (cit. on p. 32).

Vecchi, G. A. and Harrison, D. E. (2002). "Monsoon Breaks and Subseasonal Sea Surface Temperature Variability in the Bay of Bengal*". In: *Journal of Climate* 15.12, pp. 1485–1493. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/1520-0442(2002)015<1485:MBASSS>2.0.CO;2 (cit. on p. 1).

Venugopal, V., Foufoula-Georgiou, E., and Sapozhnikov, V. (1999). "Evidence of dynamic scaling in space-time rainfall". In: *Journal of Geophysical Research: Atmospheres* 104.D24, pp. 31599–31610 (cit. on p. 6).

Villarini, G., Mandapaka, P. V., Krajewski, W. F., and Moore, R. J. (2008). "Rainfall and sampling uncertainties: A rain gauge perspective". In: *Journal of Geophysical Research* 113 (D11). ISSN: 0148-0227. DOI: 10.1029/2007JD009214 (cit. on p. 117).

Vinnarasi, R. and Dhanya, C. (2016). "Changing characteristics of extreme wet and dry spells of Indian monsoon rainfall". In: *Journal of Geophysical Research: Atmospheres* 121.5, pp. 2146–2160 (cit. on p. 99).

Visbeck, M. H., Hurrell, J. W., Polvani, L., and Cullen, H. M. (2001). "The North Atlantic Oscillation: Past, present, and future". In: *Proceedings of the National Academy of Sciences* 98.23, pp. 12876–12877. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.231391598 (cit. on p. 42).

Wadoux, A. M.-C., Brus, D. J., Rico-Ramirez, M. A., and Heuvelink, G. B. (2017). "Sampling design optimisation for rainfall prediction using a non-stationary geostatistical model". In: *Advances in Water Resources* 107, pp. 126–138. ISSN: 03091708. DOI: 10.1016/j.advwatres.2017.06.005 (cit. on p. 106).

Webster, R. and Oliver, M. A. (2007). *Geostatistics for Environmental Scientists*. Statistics in Practice. Chichester, UK: John Wiley & Sons, Ltd. ISBN: 978-0-470-51727-7 978-0-470-02858-2. DOI: 10.1002/9780470517277 (cit. on p. 113).

Werndl, C. (2016). "On Defining Climate and Climate Change". In: *The British Journal for the Philosophy of Science* 67.2, pp. 337–364. ISSN: 0007-0882, 1464-3537. DOI: 10.1093/bjps/axu048 (cit. on p. 1).

Xoplaki, E., González-Rouco, J., Luterbacher, J., and Wanner, H. (2004). "Wet season Mediterranean precipitation variability: influence of large-scale dynamics and trends". In: *Climate Dynamics* 23.1. ISSN: 0930-7575, 1432-0894. DOI: 10.1007/s00382-004-0422-0 (cit. on p. 64).

Yamagata, T., Behera, S. K., Luo, J.-J., Masson, S., Jury, M. R., and Rao, S. A. (2013). "Coupled Ocean-Atmosphere Variability in the Tropical Indian Ocean". In: *Geophysical Monograph Series*. Ed. by C. Wang, S. Xie, and J. Carton. Washington, D. C.: American Geophysical Union, pp. 189–211. ISBN: 978-1-118-66594-7 978-0-87590-412-2. DOI: 10.1029/147GM12 (cit. on pp. 42, 59, 81).

Yamasaki, K., Gozolchiani, A., and Havlin, S. (2008). "Climate Networks around the Globe are Significantly Affected by El Niño". In: *Physical Review Letters* 100.22. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.100.228501 (cit. on pp. 32, 35).

Yan, R. and Gao (2007). "A tour of the tour of the Hilbert-Huang transform: an empirical tool for signal analysis". In: *IEEE Instrumentation & Measurement Magazine* 10.5, pp. 40–45 (cit. on p. 20).

Yang, C., Olson, B., and Si, J. (2011). "A Multiscale Correlation of Wavelet Coefficients Approach to Spike Detection". In: *Neural Computation* 23.1, pp. 215–250. ISSN: 0899-7667, 1530-888X. DOI: 10.1162/NECO_a_00063 (cit. on p. 32).

Yang, Xie, Liu, Ji, and Wang (2015). "Spatial Interpolation of Daily Rainfall Data for Local Climate Impact Assessment over Greater Sydney Region". In: *Advances in Meteorology* 2015, pp. 1–12. ISSN: 1687-9309, 1687-9317. DOI: 10.1155/2015/563629 (cit. on p. 89).

Yeh, M.-S., Lin, Y.-P., and Chang, L.-C. (2006). "Designing an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging and genetic algorithms". In: *Environmental Geology* 50.1, pp. 101–121. ISSN: 0943-0105, 1432-0495. DOI: 10.1007/s00254-006-0190-8 (cit. on p. 106).

Yoo, C., Jung, K.-S., and Ahn, J. H. (2003). "Estimating Characteristics of Rainfall and Their Effects on Sampling Schemes: Case Study for Han River Basin, Korea". In: *Journal of Hydrologic Engineering* 8.3, pp. 145–157. ISSN: 1084-0699, 1943-5584. DOI: 10.1061/(ASCE)1084-0699(2003)8:3(145) (cit. on p. 106).

Zanin, M. (2014). "Complex Networks and Data Mining: Toward a new perspective for the understanding of Complex Systems". In: (cit. on p. 6).

Zhang, R. and Delworth, T. L. (2005). "Simulated Tropical Response to a Substantial Weakening of the Atlantic Thermohaline Circulation". In: *Journal of Climate* 18.12, pp. 1853–1860. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/JCLI3460.1 (cit. on p. 73).

– (2006). "Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes". In: *Geophysical Research Letters* 33.17. ISSN: 0094-8276. DOI: 10.1029/2006GL026267 (cit. on p. 73).

Zhou, Gozolchiani, A., Ashkenazy, Y., and Havlin, S. (2015). "Teleconnection Paths via Climate Network Direct Link Detection". In: *Physical Review Letters* 115.26. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.115.268501 (cit. on p. 50).

Zhou, Zemanová, L., Zamora-López, G., Hilgetag, C. C., and Kurths, J. (2007). "Structure–function relationship in complex brain networks expressed by hierarchical synchronization". In: *New Journal of Physics* 9.6, pp. 178–178. ISSN: 1367-2630. DOI: 10.1088/1367-2630/9/6/178 (cit. on pp. 91, 92).

Zlatić, V., Božičević, M., Štefančić, H., and Domazet, M. (2006). "Wikipedias: Collaborative web-based encyclopedias as complex networks". In: *Physical Review E* 74.1. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.74.016115 (cit. on pp. 92, 107).

Zrinji, Z. and Burn, D. H. (1994). "Flood frequency analysis for ungauged sites using a region of influence approach". In: *Journal of Hydrology* 153.1, pp. 1–21. ISSN: 00221694. DOI: 10.1016/0022-1694(94)90184-8 (cit. on p. 90).

– (1996). "Regional Flood Frequency with Hierarchical Region of Influence". In: *Journal of Water Resources Planning and Management* 122.4, pp. 245–252. ISSN: 0733-9496, 1943-5452. DOI: 10.1061/(ASCE)0733-9496(1996)122:4(245) (cit. on p. 90).

# Selbständigkeitserklärung

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

Potsdam, den 1. Februar 2019 Ankit Agarwal