Institut für Physik der Universität Potsdam,
Potsdam-Institut für Klimafolgenforschung (PIK)
und
DFG Graduiertenkolleg 1364

Extremes In Events And Dynamics

A nonlinear data analysis perspective on the past and present dynamics of the Indian Summer Monsoon

Dissertation
zur Erlangung des akademischen Grades
"doctor rerum naturalium" (Dr. rer. nat.)
in der Wissenschaftsdisziplin "Nonlinear Dynamics"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von
Nishant Malik

Potsdam, im Oktober 2011

Published online at the
Institutional Repository of the University of Potsdam:
URL http://opus.kobv.de/ubp/volltexte/2012/5801/
URN urn:nbn:de:kobv:517-opus-58016
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-58016

To my parents, teachers and colleagues

Abstract

To identify extreme changes in the dynamics of the Indian Summer Monsoon (ISM) in the past, I propose a new approach based on the quantification of fluctuations of a *nonlinear similarity measure*, to identify regimes of distinct dynamical complexity in short time series. I provide an analytical derivation for the relationship of the new measure with the dynamical invariants such as dimension and Lyapunov exponents of the underlying system. A *statistical test* is also developed to estimate the significance of the identified transitions. Our method is justified by uncovering bifurcation structures in several paradigmatic models, providing more complex transitions compared with traditional Lyapunov exponents. In a real world situation, we apply the method to identify millennial-scale dynamical transitions in Pleistocene proxy records of the south Asian summer monsoon system. We infer that many of these transitions are induced by the external forcing of solar insolation and are also affected by internal forcing on Monsoonal dynamics, i.e., the glaciation cycles of the Northern Hemisphere and the onset of the tropical Walker circulation. Although this new method has general applicability, it is particularly useful in analysing short palaeo-climate records.

Rainfall during the ISM over the Indian subcontinent occurs in form of enormously complex spatiotemporal patterns due to the underlying dynamics of atmospheric circulation and varying topography. I present a detailed analysis of summer monsoon rainfall over the Indian peninsular using *Event Synchronization (ES)*, a measure of nonlinear correlation for point processes such as rainfall. First, using *hierarchical clustering* I identify principle regions where the dynamics of monsoonal rainfall is more coherent or homogenous. I also provide a method to reconstruct the time delay patterns of rain events. Moreover, further analysis is carried out employing the tools of *complex network theory*. This study provides valuable insights into the spatial organization, scales, and structure of the 90th and 94th percentile rainfall events during the ISM (June to September). I furthermore analyse the influence of different critical synoptic atmospheric systems and the impact of the steep Himalayan topography on rainfall patterns. The presented method not only helps in visualising the structure of the extreme-event rainfall fields, but also identifies the water vapor pathways and decadal-scale moisture sinks over the region. Furthermore a simple scheme based on complex networks is presented to decipher the spatial intricacies and temporal evolution of monsoonal rainfall patterns over the last six decades. Some supplementary results on the evolution of monsoonal rainfall extremes over the last sixty years are also presented.

Zusammenfassung

Um Extremereignisse in der Dynamik des indischen Sommermonsuns (ISM) in der geologischen Vergangenheit zu identifizieren, schlage ich einen neuartigen Ansatz basierend auf der Quantifikation von Fluktuationen in einem nichtlinearen Ähnlichkeitsmaß vor. Dieser reagiert empfindlich auf Zeitabschnitte mit deutlichen Veränderungen in der dynamischen Komplexität kurzer Zeitreihen. Ein mathematischer Zusammenhang zwischen dem neuen Maß und dynamischen Invarianten des zugrundeliegenden Systems wie fraktalen Dimensionen und Lyapunovexponenten wird analytisch hergeleitet. Weiterhin entwickle ich einen statistischen Test zur Schätzung der Signifikanz der so identifizierten dynamischen Übergänge. Die Stärken der Methode werden durch die Aufdeckung von Bifurkationsstrukturen in paradigmatischen Modellsystemen nachgewiesen, wobei im Vergleich zu den traditionellen Lyapunovexponenten eine Identifikation komplexerer dynamischer Übergänge möglich ist. Wir wenden die neu entwickelte Methode zur Analyse realer Messdaten an, um ausgeprägte dynamische Veränderungen auf Zeitskalen von Jahrtausenden in Klimaproxydaten des südasiatischen Sommermonsunsystems während des Pleistozäns aufzuspüren. Dabei zeigt sich, dass viele dieser Übergänge durch den externen Einfluss der veränderlichen Sonneneinstrahlung, sowie durch dem Klimasystem interne Einflussfaktoren auf das Monsunsystem (Eiszeitzyklen der nördlichen Hemisphäre und Einsatz der tropischen Walkerzirkulation) induziert werden. Trotz seiner Anwendbarkeit auf allgemeine Zeitreihen ist der diskutierte Ansatz besonders zur Untersuchung von kurzen Paläoklimazeitreihen geeignet.

Die während des ISM über dem indischen Subkontinent fallenden Niederschläge treten, bedingt durch die zugrundeliegende Dynamik der atmosphärischen Zirkulation und topographische Einflüsse, in äußerst komplexen, raumzeitlichen Mustern auf. Ich stelle eine detaillierte Analyse der Sommermonsunniederschläge über der indischen Halbinsel vor, die auf Ereignissynchronisation (ES) beruht, einem Maß für die nichtlineare Korrelation von Punktprozessen wie Niederschlagsereignissen. Mit hierarchischen Clusteringalgorithmen identifiziere ich zunächst Regionen mit besonders kohärenten oder homogenen Monsunniederschlägen. Dabei können auch die Zeitverzögerungsmuster von Regenereignissen rekonstruiert werden. Darüber hinaus führe ich weitere Analysen auf Basis der Theorie komplexer Netzwerke durch. Diese Studien ermöglichen wertvolle Einsichten in räumliche Organisation, Skalen und Strukturen von starken Niederschlagsereignissen oberhalb der 90% und 94% Perzentilen während des ISM (Juni bis September). Weiterhin untersuche ich den Einfluss von verschiedenen, kri-

tischen synoptischen Systemen der Atmosphäre sowie der steilen Topographie des Himalayas auf diese Niederschlagsmuster. Die vorgestellte Methode ist nicht nur geeignet, die Struktur extremer Niederschlagsereignisse zu visualisieren, sondern kann darüber hinaus über der Region atmosphärische Transportwege von Wasserdampf und Feuchtigkeitssenken auf dekadischen Skalen identifizieren. Weiterhin wird ein einfaches, auf komplexen Netzwerken basierendes Verfahren zur Entschlüsselung der räumlichen Feinstruktur und Zeitentwicklung von Monsunniederschlagsextremen während der vergangenen 60 Jahre vorgestellt.

List of Publications

This thesis is partially based on the following publications:

1. Malik, N., Zou, Y. Marwan, N., Kurths, J. (2012): *Dynamical regimes and transitions in Plio-Pleistocene Asian monsoon*. **Europhysics Letters** (Accepted)

2. Malik, N., Bookhagen, B. Marwan, N., Kurths, J. (2011): *Analysis of spatial and temporal extreme monsoonal rainfall over South Asia*. **Climate Dynamics**, DOI:10.1007/s00382-011-1156-4

3. Malik, N., Marwan, N., Kurths, J. (2010): *Spatial structures and directionalities in Monsoonal precipitation over South Asia*. **Nonlinear Processes in Geophysics** , 17(5), 371–381. DOI:10.5194/npg-17-371-2010

Acknowledgments

Several people helped me in developing this work at different stages of my PhD. I am most indebted to my supervisor Prof. Dr. Jürgen Kurths, for giving me this tremendous academic opportunity, to carry out my PhD in his group. He was very supportive to me all throughout my PhD years and scientific discussions with him were always extremely enriching and helped me very much in developing my scientific ideas.

Dr. Norbert Marwan has contributed immensely to this work, in form of regular discussions and basic academic help. Working with him was a great learning experience, I can not imagine finishing this work without his constant support. I have benefited enormously from numerous discussions with Dr. Yong Zou on dynamical systems and nonlinear time series analysis. Further, I want to thank Prof. Bodo Bookhagen for several things: firstly for fruitful discussions on monsoonal precipitation, and also for helping me enormously in writing one of the manuscripts and lastly, for providing me the processed TRMM data set. One of the most important help came from my scientific colleagues and fellow PhD students at the Potsdam Institute for Climate Impact Research. My office mate for last three years, Jonathan F. Donges – I am indebted to him for providing congenial academic environment in our office. His expertise in Python programming and complex network theory was of great help as well. Dr. Jobst Heitzig, Kira Rehfeld and Jakob Runge have always been a great support and help. Lately discussions with Bedartha Goswami has also been very interesting. I want to thank Nora and Bedartha for proof reading parts of this thesis. Special thanks to Anna Zakharova, Arghya Mondal and Jojo for being such great friends during my stay in Potsdam.

I would also like to acknowledge the financial support I have got from DFG Graduate School GK -1364. I also got chance to visit north-west Himalayas with the Graduate School, where I got the opportunity to learn and appreciate the work of geologists. I would like to give my special thanks to Prof. Dr. Manfred Strecker for giving me this opportunity and also supporting my several other visits to workshops and conferences to different destinations. Also, I am grateful to different co-ordinators of the graduate school - Dr. Andreas Bergner, Dr. Rasmus Thiede and Dr. Henry Wichura for their support. I would also like to thank students from the graduate school for being good friends.

Nothing would have been possible without the constant support and encouragement of my father Dr. G.R. Malik and my mother Mrs. Usha Malik. They made countless sacrifices in their personal life, so I can pursue my dreams.

Contents

Chapter **1**

Introduction

Equipped with his five senses, man explores the universe around him and calls the adventure Science.†

– Edwin Powell Hubble

Extreme events occur extensively in nature and society. Floods, forest fires, earthquakes, abrupt financial meltdowns, epileptic seizures are all examples of extreme events [*Albeverio et al. (2006)*]. The dynamics of a physical system can also show extreme changes, leading to birth of entirely new features and instabilities in the dynamics of the system. At times these new features and instabilities in the dynamics of the systems could lead to catastrophic consequences in nature and society. One such system where extremes in events and dynamics have extremely high socio-economic impacts is the geophysical system of the Indian Summer Monsoon (ISM). During the ISM, extreme events like floods and droughts are common occurrence with devastating influences on the livelihood of the inhabitants of the region. Also, the ISM has been hypothesised to be one of the tipping elements of the global climate system, which means its dynamics is composed of such instabilities that changes in global climate system could lead to emergence of undesired features in its dynamics, including the total breakdown of the monsoonal circulation itself [*Lenton et al. (2008)*, *Levermann et al. (2009)*].

In this PhD thesis I have attempted to answer several critical questions (with a data oriented analysis of the ISM) related to extremes of the ISM's dynamics in the past and of the rain events during the ISM season. Some of these question are: what factors and forcings has influenced evolution of monsoon over last 3.5 million years ? what are the persistence atmospheric features that are responsible for generation of extreme rain events during the ISM season ? how are the spatial rainfall patterns of the ISM evolving in the warming envi-

† "The Nature of Science, and Other Lectures" by Hubble, Edwin Powell, (Huntington Library, San Marino, CA, 1954)

ronment over the last six decades ?. To answer these questions and to gain deeper insights into the dynamics of the ISM on different time scales I have developed some new tools and methods based on theoretical framework of *nonlinear dynamics* and *complex networks* [*Kantz and Schreiber (2004)*, *Cohen and Havlin (2010)*].

The study of the past dynamics of the ISM is of paramount interest for understanding the present and future behaviours of the ISM dynamics. The most natural theoretical tool for studying the dynamics of the ISM on distinct time scales, is the theoretical framework of *nonlinear dynamics*: as the ISM is fundamentally a nonlinear dynamical system with several feedback mechanisms coupled with local and global dynamical couplings with other subsystems of the global climate system[*Wang (2006)*, *Webster (1987)*]. However, we observe that the tools of nonlinear dynamics applicable to data analysis: popularly known as *nonlinear time series analysis*, still have several limitations when it comes to their application to palaeoclimate time series. I have attempted to overcome these shortcomings and tried to provide some solutions to deal with palaeoclimate data using tools of nonlinear dynamics. I will present a new method based on recurrence properties of dynamical systems, which could be employed to identify extreme dynamical changes or so called the dynamical transitions. Furthermore, this method is used to attempt a reconstruction of the forcings of the underlying system and thus, one can obtain further insights into the basic physics of the system under study.

The mainstream methodological studies on extremes in rainfall during the ISM have mostly concentrated on understanding the statistical properties of such extremes [*May (2004b)*, *May (2004a)*, *W.May (2004)*, *Stephenson et al. (1999)*, *Goswami et al. (2006)*]. In this thesis I have employed entirely new approaches of nonlinear time series analysis coupled with graph theoretic methods to study their spatial and temporal properties. Firstly, I have extended the concept of *Event Synchronization* (ES) to estimate the magnitude of nonlinear correlations that ceases to exist in rainfall events over large spatial areas. Analysis of these correlations helps in improving our understanding about some of the dynamical features of the atmosphere responsible for generation of extreme rainfall events. Using the tools like hierarchal clustering algorithm, I have been able to estimate the regions of coherent rainfall activity over the Indian subcontinent. Where the extreme rainfall events occur in some lag synchronized form. I speculate that this type of synchronized rainfall could lead to large scale floods. Furthermore, I have employed tools from complex network theory to analyse the spatial correlations among 90% and 94% percentile events estimated using the measure of ES. This study reveals several useful insights about the interaction among the atmospheric processes responsible for the generation of such extreme rainfall events.

This thesis is organised as follows: In chapter 2, I provide a brief introduction to the ISM and list the basic climatic terminologies that I have used in later chapters in the thesis. In chapter 3, I have documented the analytical and computational details of the new method based on fluctuation of nonlinear similarities for identifying dynamical transitions. Also, I have provided an example of its application on a Pleistocene record of the ISM. Chapter 4 introduces the concept of Event Synchronization (ES). Further this chapter, describes a method where ES is used in combination with hierarchal clustering to estimate coherent

zones of rainfall over India. In chapter 5, I will present a detailed analysis of extreme rainfall events by employing analysis of ES using complex networks. Appendix A gives the details of the data sets used in the thesis including their sources. Appendix B consist of supplementary results related to the trends in extreme rainfall events in the ISM over the last six decades. Except chapter 2 and appendix B, most of the other material presented in this thesis has either been already published or submitted. The details of the publications are given in the list on page iii.

Chapter **2**

A brief introduction to the Indian Summer Monsoon (ISM)

Come oh verdurous splendor
Allaying heart, slaking thirst
Grace with your cooling presence
Gazing longingly at skies
Awaits you this lovelorn earth

– Rabindranath Tagore*
on his favourite season of 'Monsoon'

In this chapter I intend to provide a brief introduction to the basic physical mechanism of the ISM, its driving factors, dynamics and evolutionary history. Another aim is to give a concise account of the basic terminologies related to the climatological features of the ISM which are frequently used in the later chapters and pertinent to this study. For extensive details about the ISM, the interested reader is referred to summaries and references in *Webster (1987)*, *Webster et al. (1998)*, *Gadgil (2003)*, and *Wang (2006)*.

The origin of the word monsoon (*mausim*) could be traced to 16th century Arab sailors, who used this word to describe seasonal reversal of winds in the Arabian sea. The monsoonal regions of the world are mostly situated between $25°S$ to $35°N$, $30°W$ to $170°E$. Right in the centre of this region lies the Indian subcontinent, one of the most populous parts of the world. The origin of monsoons is in the seasonal reversal of wind directions due to the strong differential heating of land and the surrounding tropical Oceans. This in turn creates a very strong seasonality in rainfall at the annual scale. In particular, the Indian subcontinent receives over 75% of its annual rainfall over the year during the four summer months of June, July, August, and September(JJAS) (see Fig. 2.3 (A)). Most parts of the region receive there quota of monsoonal rainfall during these four summer months. Therefore, the name

* He was a Bengali poet, novelist, musician, painter and playwright who reshaped Bengali literature and music. Awarded the Nobel Prize for Literature in 1913.

summer monsoon. It is also known as Southwest monsoon because of the wind directions during these months. Some regions in the peninsular India receive rainfall during winter (October-November) due to the winter monsoon which is also known as Northeast monsoon season [*Gadgil (2003)*]. Monsoonal precipitation is one of the chief source for the availability of fresh water for most parts of this region, therefore many of the agricultural practices are moulded according to the monsoon season. This brings a direct dependence of agricultural production to the quality and quantity of monsoonal rains [*Webster et al. (1998)*]. It is known that over sixty percent of the population of this region is directly or indirectly involved in agriculture based economic activities. As a consequence, the monsoon is very critical climatic phenomena with large socio-economic impacts for the inhabitants of the Indian subcontinent. Hence, understanding monsoon and its variability on different time scales is of considerable importance in climate sciences and in society [*Webster et al. (1998)*, *Gadgil (2003)*].

2.1 Basic mechanism

Monsoons are among the most prominent and dynamic phenomena of the climate system, manifesting over large parts of the tropics. It has been argued that the basic origin of the ISM lies in the differential heating of the land and the sea during the summer season, which results in setting of a positive moisture advection feedback leading to widespread rainfall over the Indian subcontinent [*Webster (1987)*, *Zickfeld et al. (2005)* and *Levermann et al. (2009)*]. The gradient in sensible heating between land and sea during the peak of summer seasons results in low pressure over the Indian land mass and high pressure over the adjoining seas, which sets in the monsoonal circulation. The release of latent heat in the processes of precipitation over the land provides the feedback for maintaining this temperature gradient and corresponding pressure gradient and thus sustaining monsoonal circulation (see Fig. 2.1). The main source of the moisture flux into the Indian land mass during monsoon season and the associated latent heat has been found to be the southern subtropical Indian Ocean across the equator[*Cadet and Diehl (1984)*].

The strength of monsoonal circulation is critically dependent on the direct heating of the troposphere by the latent released over land from the precipitation of moisture mainly transported from southern subtropical Indian Ocean [*Webster (1987)*, *Clemens et al. (1991)*]. If only the temperature gradient due to sensible heat would be maintaining the monsoonal circulation then, monsoon would have been constant annually. But it shows enormous amount of temporal variability on annual scales. One of the source of this variability is the internal dynamics of monsoon, caused by the fluctuations in storage and transport of the latent heat. The latent heat carried by moisture $\mathscr{L}$ is given by the formula [*Liu et al. (1994)*] :

$$\begin{aligned} \mathscr{L} &= \rho_a C_E L u \Delta q \\ \Delta q &= q_s - q_a = [Q(T_s) - rQ(T_s + \Delta T)] \end{aligned} \tag{2.1}$$

where ρ_a is the air density, C_E is the transfer coefficient, L is the latent heat of vaporisation,

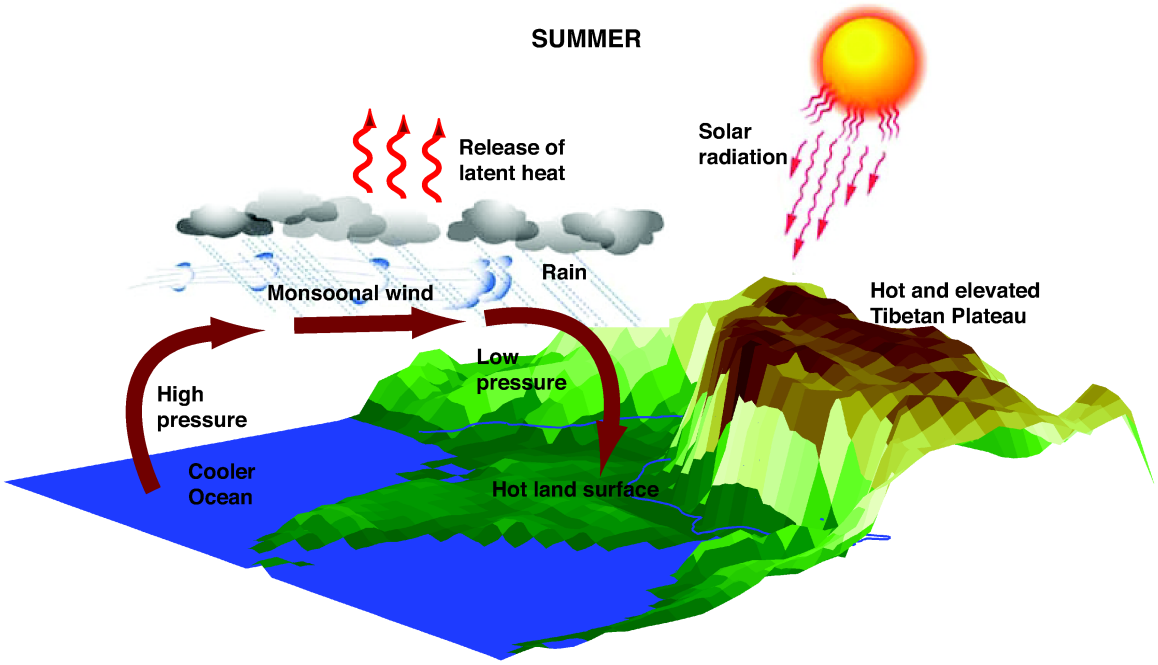

Figure 2.1 Schematic representation of the basic physical mechanism of South Asian summer monsoon : land-sea temperature contrast during peak of summer causes moisture laden winds to blow from the sea to land. The process of rainfall over the land releases further latent heat into the atmosphere which acts as positive feedback for the monsoonal circulation.

u is the wind speed, Δq and ΔT are the sea - air humidity and temperature difference, $Q(T)$ gives the saturation specific humidity as function of temperature T, subscript s represents the value at the sea-air interface, whereas a represents the value at reference point in atmospheric constant flux layer, q is specific humidity, r is relative humidity. If r, u and ΔT are constant then $\mathscr{L}$ is expected to increase with T_s with a rate governed by Clausius-Clapeyron function [*Liu et al. (1994)*]. In realistic situations r, u and ΔT are hardly ever constants. So, in more real scenario we observe $\mathscr{L}$ being positively correlated with u and negatively with q_a [*Liu et al. (1994)*, *Clemens et al. (1996)*], this is the reason for high $\mathscr{L}$ in trade winds. Stronger trade winds, colder conditions in southern subtropical Indian Ocean are associated with higher cross-equatorial moisture transport and enhanced precipitation over the Indian subcontinent [*Cadet and Diehl (1984)*], *Clemens et al. (1996)*].

The transition to the onset or withdrawal of monsoon season could be defined in terms of a change in wind patterns and strength, as the combined effect of sensible and latent heating could result in strong Southwest winds of up to 15 ms^{-1} over the Arabian Sea during the months of June, July and August [*Hastenrath (1987)*]. An alternative approach could be to define the transition to monsoon in terms of rapid, substantial and sustained increase in the precipitation rate or employing more reliable vertically integrated moisture transport (VIMT)

[*Fasullo and Webster (2003)*]. The transition to the monsoon season is usually abrupt, and associated with movement of synoptic scale convective system over land originating from nearby seas and bringing rainfall over large parts of the Indian subcontinent [*Gadgil (2003)*]. The transition from monsoon is equally abrupt too, with weakening south west winds and a emergence of dry, quiescent conditions [*Fasullo and Webster (2003)*].

2.2 Origin and evolution

Initiation of monsoonal circulation has been linked to the uplift of the Tibetan Plateau due to the collision of the Indian and the Asian plate around 50 Ma (megaannum) ago. Geological data sets which can shade a light on the initiation or intensification of monsoon between $40-10$ Myr are still lacking[*Wang et al. (2005)*, *Clemens (2006)*]. Though GCM experiments indicate that a monsoonal circulation similar to present could be supported by solar forcing with the elevation of Tibet-Himalayas set at least half of the present [*Prell and Kutzbach (1992)*]. GCM experiments also indicate that apart from Tibetan plateau uplift the retreat of Paratethys has also played a role in the intensification of monsoonal circulation [*Kutzbach et al. (1993)*]. Several other factors are also expected to influence the origin and later intensification of monsoon such as patterns of land-sea distribution [*Ramstein et al. (1997)*], emergence of geological features like closure of oceanic gateways [*Cane and Molnar (2001)*]. Long term records of *G. bulloides* from the ODP (Ocean Drilling Program) site 722 indicate monsoon related upwelling in the Arabian Sea significantly increased around 8.5 Ma [*Kroon and Ganssen (1989)*, *Prell and Kutzbach (1992)*]. Also, δ^{13}C data of pedogenic carbonate from northern Pakistan in the Himalayan foreland indicate rapid ecological transition from C3 dominated vegetation to C4 dominated vegetation between $7.0-7.4$ Ma [*Quade et al. (1989)*]. Indicating that intensification of monsoon to modern levels could have taken place around this period.

Several distinct forcings and mechanisms have operated on the evolution of present day monsoon. Fig 2.2 provides a flow diagram representation of the three dominant forcings on monsoon. A large variety of monsoon related palaeo proxy records are available from the Pleistocene - Peliocene period (up to 5.0 Ma BP) (see [*Wang et al. (2005)*, *Clemens (2006)*]). Dust flux records from ODP sites belonging to this period has been extensively studied and analysed [*deMenocal (1995)*, *deMenocal (2004)*, *Larrasoaña et al. (2003)*, *Trauth et al. (2009)*, *Clemens et al. (1996)*, *Clemens et al. (1991)*]. These studies have accounted the variability of monsoon over Pleistocene - Pliocene period to two main forcings viz. changes in solar insolation induced by variation of orbital parameters of the Earth and the global ice volume. Astrodynamical calculation make it possible to reconstruct the orbital cycles and associated changes in solar forcings [*Berger and Loutre (1991)*, *Laskar et al. (2004)*, *Laskar (1999)*, *Laskar et al. (1993)*]. Using advanced approaches from time series analysis one could attempt to reconstruct the response of monsoon to these forcings [*Malik et al. (2011b)*]. This is one of the aims of my thesis too and chapter 3 will deal with theoretical challenges involved in reconstruction of such forcings from the palaeo records.

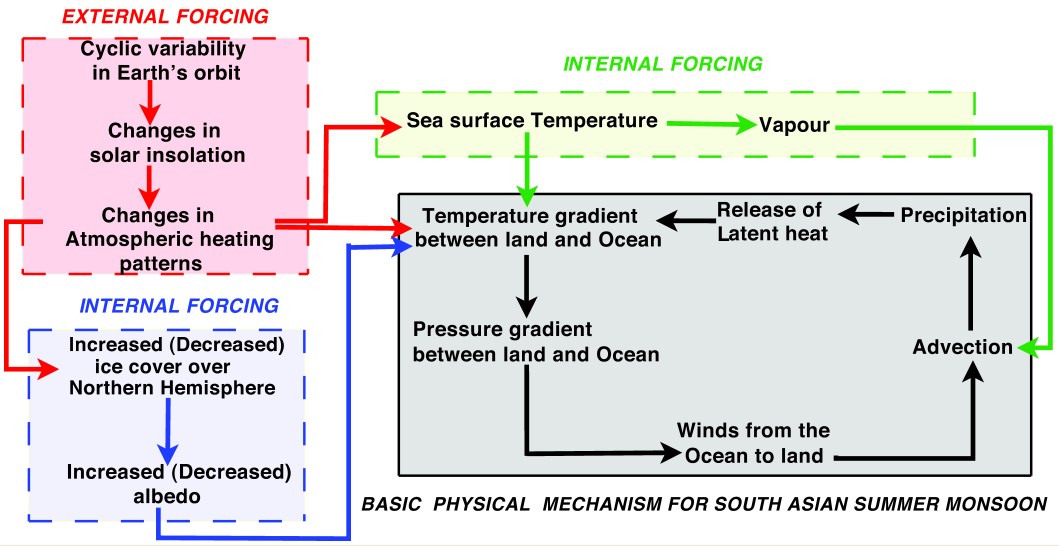

Figure 2.2 Three dominant forcing on the South Asian monsoonal system: Solar variability (red), Northern Hemisphere (NH) ice cover (blue) and Sea Surface Temperature (green). The Solar variability also effects the NH ice cover and SST. The basic physical mechanism of monsoon is shown inside the black box.

At the millennial time scales (10^3 years) there is no known source of variability in solar insolation and during these periods monsoon has mostly responded to internal dynamics of the climate system. Most of the variability in palaeo monsoon records found on these time scales could be accounted to the Dansgaard-Oeschger (DO) events [*Dansgaard et al. (1993)*, *Wang et al. (2001)*, *Schulz et al. (1998)*, *Altabet et al. (2002)*]. Though the exact physical mechanism of the origin of DO events in Greenland ice records and their transmission to the monsoon is still debated [*Rahmstorf (2003)*, *Ganopolski and Rahmstorf (2001)*, *Cane and Clement (1999)*], but abrupt transitions corresponding to DO events have been observed on palaeo records on millennial time scales [*Wang et al. (2001)*, *Schulz et al. (1998)*, *Altabet et al. (2002)*]. Palaeomonsoon records on centennial scales (10^2 years) show some periodicities in the range of 84−102 years and 210 years. It has been argued that these cycles correspond to 88 year and 205 year solar variability [*Agnihotri et al. (2002)*, *Neff et al. (2001)*]. Which may not be plausible due to direct influence of changes in solar irradiance as these are quite small (1.5 Wm^{-1}) [*Wang et al. (2005)*, *Clemens (2006)*], but could have been possible through a indirect influence of changes in the UV part of the solar spectrum in the ozone production [*Beer et al. (2000)*]. Records from the Holocene also show periodicities of $3-6$ years corresponding to ENSO cycles [*Berger and von Rad (2002)*,*Charles et al. (1997)*]. All these points indicate that the ISM is extremely interwound component of global climate system.

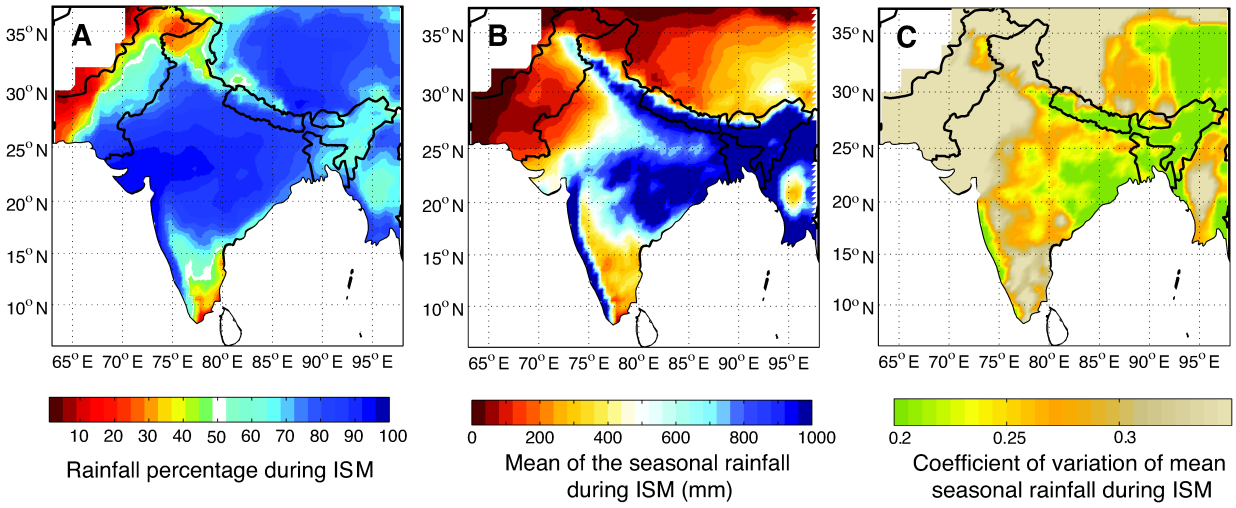

Figure 2.3 Spatial variability's in the ISM. (A) The percentage of total annual rainfall that occurs during the ISM season. A major part of south Asia receives above 70% rainfall of their annual rainfall budget during the ISM. (B) The net amount of average rainfall that occurs during the ISM. Note the higher rainfall amounts along west coast of peninsular India and central India. (C)The coefficient of variation of mean annual rainfall. The data set used in generating above figure is described in Appendix A Sec. A.1.1

2.3 Spatial and temporal variability

The most intricate feature of the ISM is its spatio-temporal variability. In this section I will discuss the spatial and temporal variability of the ISM and its causes on shorter time scales ranging from seasonal to annual scales. Also, the presented thesis deals with this issue in quite detail, so here I will attempt to provide an introduction to climatological features which are responsible for these particular variabilities. The ISM accounts for a large part of the annual rainfall budget over most of the Indian subcontinent. For instance, the Ganges Plain and central parts of the Himalayas receive $\sim$ 80% of their annual rainfall budget during the ISM [*Bookhagen and Burbank (2010)*] (see Fig. 2.3). This, in turn leads to high societal, economic and scientific significance of this climatic phenomenon [*Webster et al. (1998)*, *Gadgil (2003)*]. That is why one of the important scientific aim in studying the ISM has always been to improve forecasts for the ISM at different spatial and temporal scales. The main challenge in this aim has been to better understand the mechanisms behind these strong spatio-temporal variabilities of the ISM on multiple spatial and temporal scales.

The spatial variability and patterns of monsoon over South Asia are governed by both dynamics of monsoonal circulations and orography. The influence of orography on monsoon is most visible in regions of Western Ghats and Himalayas. Also, monsoonal rainfall is not localised at a particular point in space rather it covers a certain region in form of many events spread over large regions. To quantify the seasonal spatial variability of the ISM I calculate the coefficient of variation V_j at grid point j (Fig. 2.3 (C)), which is defined as $V_j = \sigma_j/\mu_j$, where

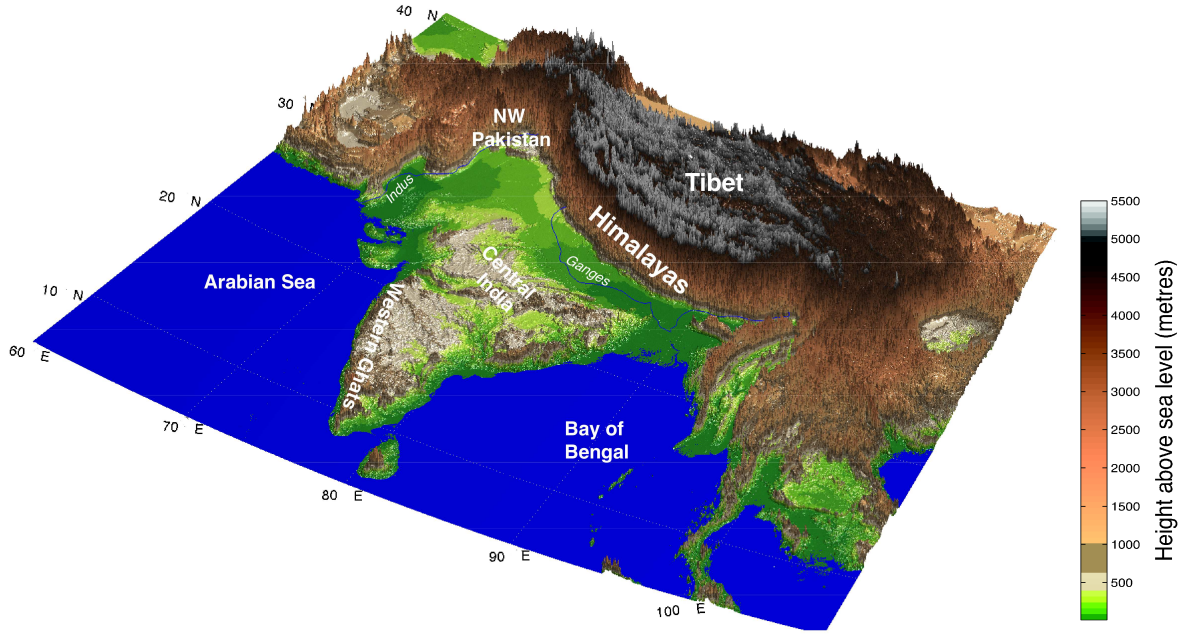

Figure 2.4 Topographic map (based on ETOPO1 data provided by NOAA, see description in Appendix A Sec. A.3] of the Indian peninsular and the Himalaya. The Himalayas form a high topographic barrier in the north resulting in orographic rainfall. Along the west coast of India are the western Ghats forming an additional orographic barrier. The blue lines show the two important rivers draining the Himalaya, the Indus in the west and the Tsangpo-Brahamputra-Ganges system in the central and eastern Himalayas.

μ_j is the mean seasonal rainfall during the ISM at grid point j (shown in Fig. 2.3 (B)) and σ_j is the standard deviation of mean seasonal rainfall during the ISM at grid point j. In Fig. 2.3 (C) we observe that most regions have its value above 0.25, indicating that the mean seasonal rainfall during the ISM deviates at least by 25% each season. Most regions with lower amount of seasonal rainfall during the ISM (see Fig. 2.3 (B)) also show higher variabilities (Fig. 2.3 (C)) in the range of over 0.3 i.e., above 30% each season. These are quite large deviations, and several factors are responsible for them. Next I will briefly synthesise some of the main reasons for this spatial variability during the ISM.

Though orographic forcing is almost constant on the shorter time scales, but it can considerably modulate the rainfall-generating processes [*Roe (2005)*]. This could lead to the emergence of complex spatial patterns of rainfall. The the ISM is influenced by large topographic barriers, such as the Himalaya and western Ghats (Fig.2.4). In the case of the Himalaya and adjacent Tibetan Plateau, the topography also alters the pathways of water-vapour transport. It has been previously documented that the steep topography of the Himalaya significantly influences spatiotemporal rainfall distribution [*Bookhagen and Burbank (2006)*, *Bookhagen and Burbank (2010)*, *Bookhagen (2010)*]. The large elevated Tibetan plateau region is considered

to heavily influence the energy budget of this region, also leading to teleconnections with different synoptic-scale weather phenomena [*Yanai and Wu (2006)*].

Rainfall during the ISM is a result of the interaction of several weather systems extending from meso-scales to synoptic and super synoptic scales. One of the most dominant synoptic system is the monsoonal trough, during the ISM season it extends from the Bay of Bengal in the east to western India and Pakistan in the west [*Ding and Sikka (2006)*]. Within the monsoon trough rain producing synoptic systems like monsoon depression and lows can be formed [*Sikka (1977)*]. Occasionally, monsoonal depressions formed over Bay of Bengal can turn into cyclonic storms and can bring heavy rainfall while moving over land while quickly weakening in the process. There is another class of synoptic systems in monsoon called the mid-tropheric cyclones [*Keshavamurthy (1973)*]. Which can bring both intense convective and non-convective rainfall [*Ding and Sikka (2006)*]. There are mesoscale vortices that are formed along the west of peninsular India, in combination with super synoptic offshore troughs they can bring very heavy rainfall (10 cm day^{-1} to even 50 cm day^{-1}) [*Ding and Sikka (2006)*]. In the north west of the subcontinent the interaction between westerly trough with monsoon trough leads to profound variation in the spatial dynamics of the monsoon [*Malik et al. (2011a)*].

The onset of the ISM occurs on average at June 1st at the southern tip of the Indian peninsular [*Webster (1987)*]. While the onset of the monsoon is relatively stable through time, the intra-seasonal variation is high [*Fasullo and Webster (2003)*, *Webster (1987)*]. An important feature of the ISM dynamics is the existence of two well pronounced rainfall regimes, known as the "Active Phase" and "Break Phase". (e.g., *Webster et al. (1998)* and *Waliser (2006)*, *Ding and Sikka (2006)*). These phases form a crucial part of the internal dynamics of the ISM and its inter-seasonal oscillations (ISO). During an active period heavy rainfall appears over most

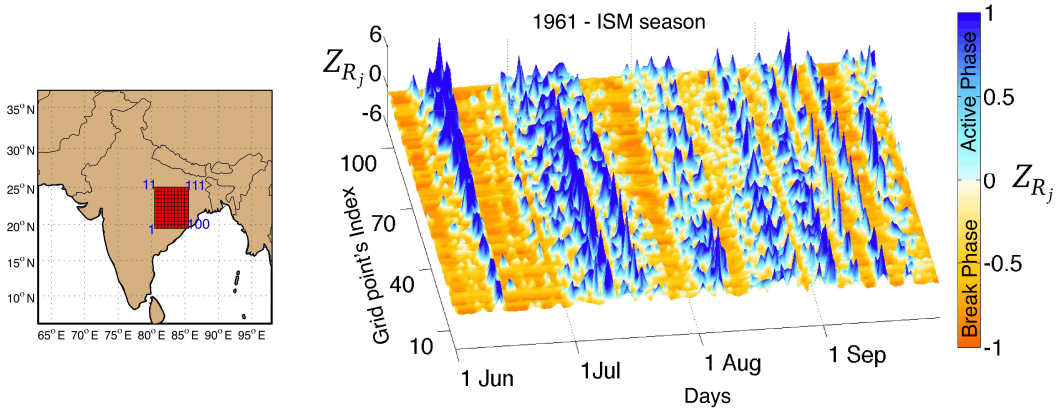

Figure 2.5 Illustration of active and break phases of monsoon during one of the season (1961) from the data set described in Appendix A Sec. A.1.1. The matrix in red colour shows the 111 grid points used for generating the figure on the right. We observe spike of rainfall activity over the whole region (shades of blue) and then quiescent state with very little rainfall activity (shades of orange). Z_{R_j} is the z-score of daily rainfall at grid point j

parts of India; in a break period the phase reverses and rainfall is recorded only along the Himalayan foothills and south-east peninsular India [*Singh et al. (1992)*, *Krishnmurthy and Shukla (2000)*]. During its active phase the monsoon could evolve on large spatial scales within a short period of time and hence large parts of the land mass could receive rainfall simultaneously or within short time delay of a few days [*Malik et al. (2010)*]. We can see this in Fig. 2.5 where the whole region shown in the red in the map almost simultaneously starts to receive heavy rainfall. Active periods of monsoon seem to be formed as a cumulative effect of several events scattered over large spatial scales (100 to 1000 km), which may also include massive rain events resulting in 400 mm of rainfall at a single day [*May (2004b)*, *Stephenson et al. (1999)*]. Such massive events usually outcome of formation of large spatial scale weather systems like monsoon depressions or midtroposheric cyclones within the monsoonal trough during its active phase. The origin of these active and break phases has been hypothesised to be due to alterations in heat sources and sinks as the rainfall activity moves from central India to north west of the Indian subcontinent [*Bhide et al. (1997)*].

The ISM rainfall is a result of inherently large spatial scale processes in the atmosphere and there are several other sources of interseasonal variation of monsoon apart from the active and break phases which are mainly due to internal dynamics of monsoon. Some of these sources correspond to large spatial scale atmospheric oscillations, which are even larger than the size of south Asia [*Waliser (2006)*]. Most notable among them is the Madden-Julian Oscillation(MJO), a $30-60$ day oscillation in the tropics [*Madden and Julian (1971)*]. A mode of $10-20$ day variability has also been observed in the ISM, which is also known as quasibiweekly mode (QBM) and hypothesised to be a Rossby wave [*Chatterjee and Goswami (2004)*]. The interannual variability of the ISM is a cumulative manifesting of the above seasonal variations and influences of other climatic phenomena on the globe via teleconnections.

2.4 Teleconnections of the ISM

The mechanism of the ISM's interaction with other large-scale climatological features is extremely complex and topic of many classical studies in climate [*Walker (1924)*, *Walker (1928)*]. The ISM system is understood to be an active component of the global climate system in the tropics and inherently coupled with oceanic variability around the region and also on the global scale [*Webster (2006)*, *Webster et al. (1998)*].

The ISM region is among the largest concentration of latent heating on the Earth during boreal summer and has the potential of interacting with other modes of variability in the global climate system [*Fasullo and Webster (2002)*, *Rodwell and Hoskins (2001)*]. One of such known interactions and which has been focus of extensive scientific debate is the link between the ENSO (El Niño Southern Oscillation) and the ISM [*Walker (1924)*, *Walker (1928)*, *Goswami (1998)*, *Torrence and Webster (1999)*, *Kumar et al. (1999)*,*Krishnamurthy and Goswami (2000)*,*Fasullo and Webster (2002)*, *Kumar et al. (2006)*]. According to a set of theories Pacific ENSO SST causes alterations in Walker circulation subsidence, which in turn influences the

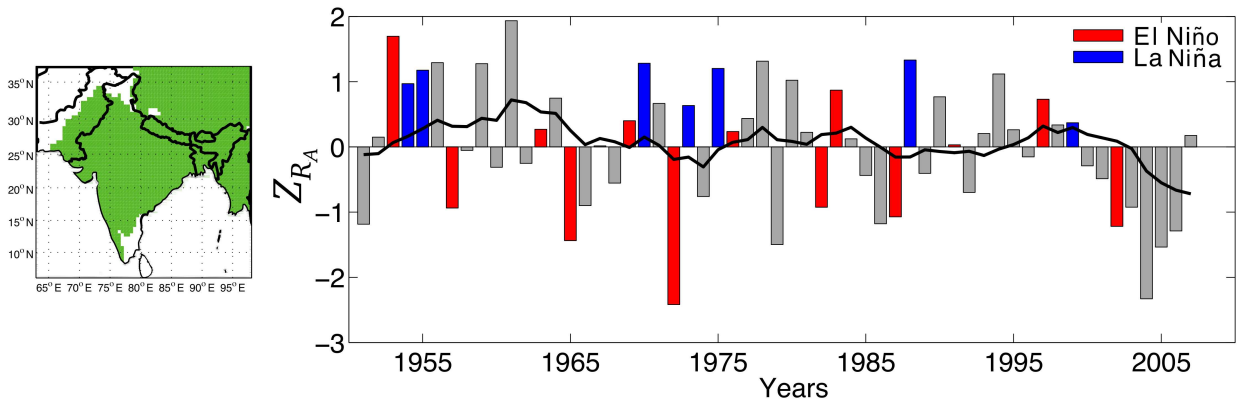

Figure 2.6 The green region in the map on the left shows the area which receive above 50% of annual rainfall budget during the ISM. The curve on right shows the area averaged z-score of annual rainfall, Z_{R_A} of the region shown in green colour in the map. The black line is the 10 year running mean of Z_{R_A}. Data set used is Appendix A Sec. A.1.1.

ISM [*Soman and Slingo (1997)*, *Dai and Wigley (2000)*], resulting in lessening of moisture transport into south Asia during El Niño years and enhancement of moisture transport into south Asia during La Niña years [*Kumar et al. (2006)*]. Interestingly this relationship has been observed to be dynamic in time i.e., it has not been constant over time rather it has changed in strength from decade to decade [*Torrence and Webster (1999)*, *Kumar et al. (1999)*, *Mokhov et al. (2011)*]. Fig 2.6 shows the annual variation in the monsoonal rainfall with highlighted El Niño and La Niña years and we can see that El Niño has not always led to a poor rainfall over south Asia. It has also been speculated that this relationship is bi-directional, i.e. ISM influences ENSO too, via the variation in atmospheric heating on the scale of south Asia due to fluctuations in latent heat associated with monsoonal rainfall [*Mokhov et al. (2011)*]. Apart ENSO, studies have also revealed that the ISM rainfall is inversely correlated with NAO (North Atlantic Oscillation) [*Dugam et al. (1997)*].

SST variability of the regional ocean could have influenced the ISM rainfall due to variation it could cause in the supply of moisture [*Webster (2006)*]. A statistically significant correlation exists between the Arabian sea SST and the ISM rainfall [*Clark et al. (2000)*]. Similarly substantial positive correlation has been shown to exist between the ISM and the IOD(Indian Ocean Dipole). It has been noted that 73% of the positive IOD events during 1958-1997 coincident with the positive anomalies of the following the ISM rainfall and 67% of the negative IOD events correspond with negative anomalies of the ISM rainfall [*Ashok et al. (2001)*].

2.5 Extremes in the ISM events and dynamics and their impacts

Monsoon rains are almost life line for billions of inhabitants of south Asia. Extreme events during monsoon occur often with devastating impacts on life and property of the people

in the region. Two of the most undesirable extremes during a monsoon season are floods and droughts. Floods are usually caused by short spells of very heavy rainfall. Whereas droughts are the aggregate effect of extended spells of no to very little rainfall activity. Both floods and droughts are disastrous for the agriculture production, the adverse impacts of these extreme events are even more magnified because of the heavy dependence of south Asian economy on agriculture [*Gadgil and Kumar (2006)*]. Famine in the past in parts of south Asia has corresponded to droughts and floods during monsoon. Hence, a systematic scientific study of extreme events, their causes, impacts and identification of most vulnerable regions and communities has high societal and economic significance [*Webster et al. (1998)* and *Gadgil (2003)*].

Extremes in dynamics is more related to long term behaviour of monsoon. Abrupt climate changes on Earth due to volcanic eruptions and glacial cycles has influenced monsoon in the past on different time scales [*Gupta et al. (2003)*, *Schulz et al. (1998)*, *Rahmstorf (2003)*]. It has also been hypothesised such abrupt transitions in the dynamics of monsoon was responsible for the collapse of Indus valley civilisation. It is of paramount interest to understand the extremes in the dynamics of the ISM over last millions of years of its evolution so to better comprehend the variabilities, instabilities and complexities in the dynamics of monsoon. Monsoon has been identified as one of the tipping element of the global climate system and it is an important question of present day climate science to understand how stable monsoon and its dynamics against anthropogenic climate change [*Lenton et al. (2008)*, *Levermann et al. (2009)*, *Goswami et al. (2006)*].

Chapter **3**

Fluctuation of similarity to detect dynamical transition in a short time series

I have been interested in phenomena involving complexity, diversity and evolution since I was a young boy.

– Murray Gell-Mann

3.1 Introduction

Detecting points of changes of dynamical complexity is of considerable interest in a wide variety of fields, such as finance (detecting stock market crashes) [*Mantegna and Stanley (2000)*], life sciences (identifying pathological activities of heart attacks or epileptic seizures [*Lehnertz and Elger (1998)*], palaeoclimate (uncovering tipping points in the climate system [*Scheffer (2009)*] or physics (studying the response of interacting many-body system to an external perturbation [*Chakrabarti and Acharyya (1999)*]. Dynamical transitions could be due to : different dynamics when the control parameter passes bifurcation thresholds; internal slow drifting of some parameter with dynamics qualitatively unchanged; the system experiences abrupt changes under external forceful influences; parameter variation of the external forcing; or noise effects as well. One of the geophysical analogies to such situations is the solar forcing which acts as the main external driver for the evolution of Earth's climate on different time scales. The variation of parameters might not always lead to extreme changes in the dynamics but could also lead to some subtle changes which are very hard to detect either by mere visual inspection or by linear time series analysis methods.

Among many geophysical processes, a highly practical problem is to identify major dynamical transitions that have occurred in the Asian monsoonal system and to obtain clues about the processes and events that causes these transitions. This is important because the monsoons are among the most critical climatic phenomena manifesting in the tropics with far reaching societal impacts. Their study at palaeo time scales is of enormous interest for the understanding of the global climate system and its evolution. Here, I study the dynamical evolution of

the Asian monsoonal system at millennial time scales and find that major dynamical shifts in it are driven by the changes in the solar insolation and the Northern Hemisphere (NH) glaciation cycles [*Clemens et al. (1996)*].

Since many geophysical processes are nonstationary and only short time series are available due to technical measurement restrictions, I propose here a rather simple yet robust method to detect dynamical transitions in short time series. This task becomes even more challenging when mutually interacting and indistinguishable processes are responsible for the dynamical transitions. Most existing nonlinear methods proposed in the literature in dealing with nonstationary time series are mainly based on recurrences [*Schreiber (1997), Rieke et al. (2002), Marwan et al. (2009)*] or recurrence time statistics [*Gao (1999)*] of the underlying system. These methods are often either context dependent (e.g., epileptic seizure detection in EEG data [*Rieke et al. (2002)*] or require very long time series [*Gao (1999)*], which are practically, often unlikely and unavailable. In this chapter, I propose a new method to identify dynamical transitions in short, univariate time series. A measure of similarity is introduced to distinguish two points of similar dynamics from two time points of different dynamics. Our method hinges on the concept of nonlinear interdependency which was proposed for extracting coupling direction for bivariate time series [*Arnhold et al. (1999)*]. Instead of measuring the interdependence between two different time series, I quantify the *fluctuation of similarity* between states at two time points, which takes into account the information contained in the entire single time series.

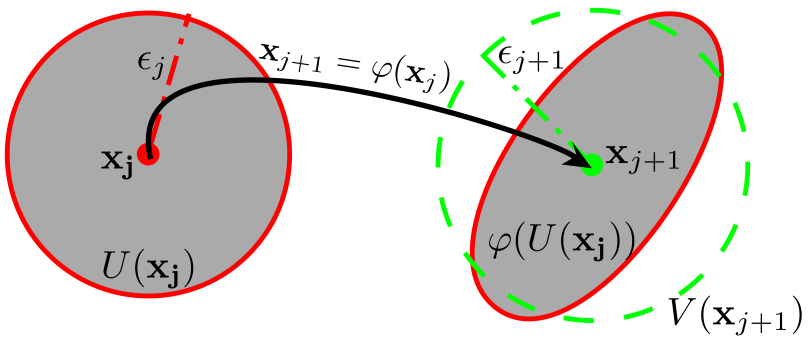

Figure 3.1 Schematic representation of the ε_j ball neighbourhood $U(\mathbf{x}_j)$ of the time point $\mathbf{x}_j$ corresponding to k-close neighbours of $\mathbf{x}_j$ and its deformation into an ellipsoid due to the application of mapping φ on it in a case of $\mathbf{x}_{j+1} = \varphi(\mathbf{x}_j)$ (see grey region within the red boundary). The neighbourhood of $\mathbf{x}_{j+1}$ corresponding to its k-nearest neighbours is usually different and it is shown with the green boundary $V(\mathbf{x}_{j+1})$. Any expansion in ball $U(\mathbf{x}_j)$ by inclusion of more points will lead to increase in the size $\varphi(U(\mathbf{x}_j)$ by the same multiplicative factor.

3.2 Method

The concept of *similarity* here is defined in terms of the *conditional closeness* of one point to the neighbourhood of another point. Let $\mathbf{x}_j$ represent the j-th vector of a delay embedded time series in phase space with length N. The embedding dimension m and time delay L are suggested respectively by fixed nearest neighbours and mutual information (cf. [*Kantz and Schreiber (2004)*]. In the phase space I denote the neighbourhood of any point $\mathbf{x}_j$ as $U(\mathbf{x}_j)$ containing k close-neighbours, namely $U(\mathbf{x}_j) = \{\mathbf{x}_l : \|\mathbf{x}_l - \mathbf{x}_j\| < \varepsilon_j\}$, where $l \in \{l_1, l_2, \cdots, l_k\}$, and $\|\cdot\|$ is a norm. The value of ε_j is set in way that we have only a fixed number of k close-neighbours for point $\mathbf{x}_j$. The criterion for setting k is that it should be substantially smaller than N. In all the examples and application below I have taken 5% of points as k close-neighbours. The point-wise *closeness* of $\mathbf{x}_j$ to its k neighbours is obtained as the mean distance $d(\mathbf{x}_j) = \frac{1}{k}\sum_{l=l_1}^{l_k} \|\mathbf{x}_j - \mathbf{x}_l\|$. After some time evolution τ at time $j+\tau$, the k close-neighbourhood of $\mathbf{x}_{j+\tau}$ is generally different and is given by $V(\mathbf{x}_{j+\tau}) = \{\mathbf{x}_i : \|\mathbf{x}_i - \mathbf{x}_{j+\tau}\| < \varepsilon_{j+\tau}\}$, where $i \in \{i_1, i_2, \cdots, i_k\}$. Fig. 4.2 show a schematic representation of these neighbourhoods and their evolution for two consecutive time points. The *conditional closeness* of $\mathbf{x}_{j+\tau}$ to the neighbourhood of $\mathbf{x}_j$ is calculated in a similar way by the mean distance as $d(\mathbf{x}_{j+\tau}|\mathbf{x}_j) = \frac{1}{k}\sum_{l=l_1}^{l_k} \|\mathbf{x}_{j+\tau} - \mathbf{x}_l\|$. Thus, the similarity of $\mathbf{x}_j$ to $\mathbf{x}_{j+\tau}$ is defined by the ratio of *closeness* and *conditional closeness* as

$$S_{j|j+\tau} = \frac{d(\mathbf{x}_j)}{d(\mathbf{x}_{j+\tau}|\mathbf{x}_j)}. \tag{3.1}$$

In a full analogy, $S_{j+\tau|j}$ characterising the similarity of $\mathbf{x}_{j+\tau}$ conditioned to $\mathbf{x}_j$ can be calculated, which often yields $S_{j|j+\tau} \neq S_{j+\tau|j}$ since $d(\mathbf{x}_{j+\tau}|\mathbf{x}_j) \neq d(\mathbf{x}_j|\mathbf{x}_{j+\tau})$. Larger values of $S_{j|j+\tau}$ indicate higher similarities in the signal. $S_{j|j+\tau}$ is local measure corresponding to each time point j and to improve its usefulness in practical applications, below I describe a strategy to extract a global feature of $S_{j|j+\tau}$ over a window of finite number of time points.

I set $\tau = 1$, i.e., I compare only two consecutive time points. If no dynamical transition occurs at time point j then determinism holds between time point j and $j+1$, i.e., $\mathbf{x}_{j+1} = \varphi(\mathbf{x}_j)$, where φ is a smooth mapping and $S_{j|j+1} \to$ const. If a dynamical transition occurs at point j then determinism will break down and there exits no such φ, which yields large fluctuation in the magnitude of $S_{j|j+1}$. I measure such large fluctuations in the values of $S_{j|j+1}$ by defining the new parameter *fluctuation of similarity* σ_S as the variance of $S_{j|j+1}$,

$$\sigma_S^2 = \left\langle (S_{j|j+1} - \mu_S)^2 \right\rangle, \tag{3.2}$$

where $\mu_S = \left\langle S_{j|j+1} \right\rangle$, and $\langle\cdot\rangle$ denote an average over N points. Therefore, a point of sharp change in σ_S, combined with a statistical significance test of this change will indicate a dynamical transition.

3.3 Analytical framework

Before, applying the above method in numerical examples, in the first part of this section I will obtain the first order approximation for the variance σ_S of the newly introduced measure of dynamical similarity $S_{j|j+\tau}$ (Eq. 3.1), in terms of the dimension of the attractor. In the latter half of this section I will also provide a discussion on the connections between the local Lyapunov exponents and $S_{j|j+\tau}$.

3.3.1. First order approximation

The method presented above relies on estimating the dynamical similarity $S_{j|j+\tau}$ (Eq. 3.1), between two points $\mathbf{x}_j$ and $\mathbf{x}_{j+\tau}$ separated by a time lag τ in a properly embedded phase space. Let assume that there exists a smooth mapping φ such that

$$\mathbf{x}_{j+1} = \varphi(\mathbf{x}_j). \tag{3.3}$$

If there is no such φ then I say that a dynamical transition occurs at j. To identify these transition points I have set $\tau = 1$, which amounts to comparing dynamical similarity of two consecutive time points in the embedded space, namely we only need to calculate $S_{j|j+1}$ for the intended application. To get detailed insights into the properties of $S_{j|j+1}$ I make use of scaling laws that are expected to exist for $d(\mathbf{x}_j)$ (the mean distance of point $\mathbf{x}_j$ to its k closest neighbours) and $d(\mathbf{x}_{j+1}|\mathbf{x}_j)$ (the mean distance of $\mathbf{x}_{j+1}$ to the k closest neighbours of $\mathbf{x}_j$), in case Eq. (3.3) is true [*Arnhold et al. (1999)*]. Suppose that a vector $\mathbf{x}_j$ in phase space has k nearest neighbours then for $k \ll N$, we will have the following scaling law (for further extensive analytical details cf. [*Arnhold et al. (1999), Pettis et al. (1979), Parker and Chua (1989)*]:

$$d(\mathbf{x}_j) = \overline{d(\mathbf{x}_j)}(k/N)^{\alpha_j}, \tag{3.4}$$

where N is the length of the time series. The coefficient $\overline{d(\mathbf{x}_j)}$ is the mean density of the whole point cloud around $\mathbf{x_j}$, i.e., $\overline{d(\mathbf{x}_j)} = \frac{1}{N}\sum_{k=1}^{N} \|\mathbf{x}_j - \mathbf{x}_k\|$. In a short time series $\alpha_j = D_p + \delta_j$, where D_p is the effective dimension of the attractor and it is equal to the reciprocal of the nearest neighbour dimension [*Arnhold et al. (1999), Pettis et al. (1979), Parker and Chua (1989)*]. δ_j are fluctuations due to the shortness of the time series and numerical errors. For a stochastic time series $D_p = m$ where m is the embedding dimension. When N tends to infinity, α_j converges to D_p. Our attempt here will be to provide the relationship between the dynamical invariant D_p and the fluctuation of similarity σ_S under the constraint that we are only considering short time series, i.e., for finite value of N. As the *conditional closeness* between $\mathbf{x}_j$ and $\mathbf{x}_{j+1}$, namely $d(\mathbf{x}_{j+1}|\mathbf{x}_j)$ has a similar geometric formulation as the distance $d(\mathbf{x}_j)$, I expect $d(\mathbf{x}_{j+1}|\mathbf{x}_j)$ to scale with the ratio k/N, which reads

$$d(\mathbf{x}_{j+1}|\mathbf{x}_j) = \overline{d(\mathbf{x}_{j+1}|\mathbf{x}_j)}(k/N)^{\beta_j}. \tag{3.5}$$

Since the definition of the similarity between two consecutive time points is $S_{j|j+1} = \frac{d(\mathbf{x}_j)}{d(\mathbf{x}_{j+1}|\mathbf{x}_j)}$, I write the scaling law for similarity by taking into account Eqs. (3.4, 3.5) in the following form

$$S_{j|j+1} = \frac{\overline{d(\mathbf{x}_j)}}{\overline{d(\mathbf{x}_{j+1}|\mathbf{x}_j)}}(k/N)^{\gamma_j}, \tag{3.6}$$

where $\gamma_j = \alpha_j - \beta_j$. The dynamical similarities between two consecutive time points $\mathbf{x}_j$ and $\mathbf{x}_{j+1}$ will be determined by the relationship between the exponents α_j and β_j. If no abrupt transition has occurred at the time point j then determinism should exist between time points j and $j+1$, i.e., a mapping of the kind φ exists and Eq. (3.3) holds. Then I expect $\gamma_j \approx 0$, i.e., $\beta_j \approx \alpha_j$ if N is infinity. In other words $d(\mathbf{x}_j)$ and $d(\mathbf{x}_{j+1}|\mathbf{x}_j)$ are expected to scale by the same exponent. I provide an intuitive explanation of this in the sketch in Fig. 4.2.

A more rigourous proof for this follows on the same lines as a proof for the invariance of point wise dimension (e.g., see *Ott (2002)*. A numerical exposition of the above scalings for the Rössler system and the Logistic map is provided in the next section (Figs. 3.3, 3.4). I also expect in Eq. (3.6) $\overline{d(\mathbf{x}_j)}/\overline{d(\mathbf{x}_{j+1}|\mathbf{x}_j)} \to$ const. if $N \to \infty$. Hence, I could also say that $S_{j|j+1} \to$ const. for $N \to \infty$. For rigourous mathematical expression for $\overline{d(\mathbf{x}_j)}/\overline{d(\mathbf{x}_{j+1}|\mathbf{x}_j)}$ c.f. [*Pettis et al. (1979), Parker and Chua (1989)*]. Here the important point to note is that all the scalings are only asymptotically valid. In a practical case of time series of finite length we will often observe fluctuating deviations of the exponents of the scalings. Next I attempt to study the influence of these fluctuations on the method and find an approximate expression for σ_S, the measure used for identifying transition.

For convenience, I define a variable r_j as

$$r_j^{\gamma_j} = \frac{\overline{d(\mathbf{x}_j)}}{\overline{d(\mathbf{x}_{j+1}|\mathbf{x}_j)}}. \tag{3.7}$$

Then Eq. (3.6) will be written as

$$S_{j|j+1} = (r_j k/N)^{\gamma_j}. \tag{3.8}$$

Next let define a ratio $c_j = \beta_j/\alpha_j$, therefore, I have $c_j \to 1$ for the case when determinism holds between two consecutive time points, i.e., Eq. (3.3) is valid. This could be observed numerically too and is shown in Figs. (3.3, 3.4). Further, I have $\gamma_j = (1 - c_j)\alpha_j$. Defining, $\Delta_j = 1 - c_j$, so for the case when determinism holds then $\Delta_j \to 0$ for N tending to infinity. As mentioned above that $\alpha_j = D_p + \delta_j$, I will have $\gamma_j = \Delta_j D_p + \Delta_j \delta_j$. Then I can rewrite Eq. (3.8) as :

$$S_{j|j+1} = (r_j k/N)^{\Delta_j D_p + \Delta_j \delta_j}. \tag{3.9}$$

As Δ_j and δ_j are small terms therefore, I can neglect their product. Then I log transform Eq. (3.9) to finally get :

$$S_{j|j+1} = \exp(D_p \ln(r_j k/N)\Delta_j). \tag{3.10}$$

Expanding the right hand side of Eq. (3.10) in terms of exponential series and neglecting the

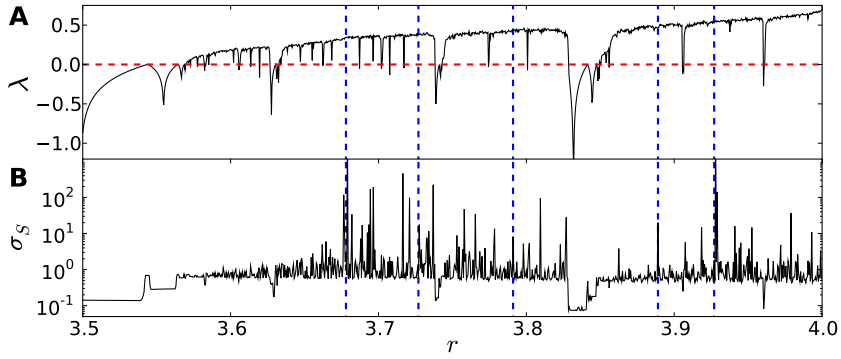

Figure 3.2 Analysis of the Logistic map. (A) Lyapunov exponent λ, (B) σ_S. Several vertical dashed lines ($r = 3.678, 3.727, 3.791, 3.889, 3.927$) highlight chaos-to-chaos transitions with laminar phases, which agree well with the results provided by the super-track function [*Oblow (1988), Marwan et al. (2009)*].

higher order terms in Δ, then I get

$$S_{j|j+1} = 1 + D_p \ln(r_j k/N)\Delta_j.$$

Writing $\Delta'_j = \Delta_j(\frac{\ln r_j}{\ln(k/N)} + 1)$, then the above expression is

$$S_{j|j+1} = 1 + D_p \ln(k/N)\Delta'_j.$$

Therefore, the mean of $S_{j|j+1}$ taken over a window of size n is,

$$\mu_S = \left\langle S_{j|j+1} \right\rangle = 1 + D_p \ln(k/N) \left\langle \Delta'_j \right\rangle.$$

Hence I can write,

$$\sigma_S^2 = (D_p \ln(k/N))^2 \sigma^2_{\Delta'_j}. \tag{3.11}$$

Equation (3.11) shows the explicit dependence of σ_S on the effective dimension D_p of the attractor, since the term $\ln(k/N)$ is kept constant over the whole length of the time series. *So, changes in the structure of the attractor will lead to changes in the value of* σ_S. The fluctuation term $\sigma^2_{\Delta'_j}$ converges for a large enough window size, which are shown numerically in the next section (Fig. 3.5). The reason for this quick convergence is the law of large numbers as the two constituent of Δ'_j i.e., r_j and Δ_j are themselves expected to converge to constant values for large N.

A dynamical transition will lead to the breakdown of determinism between consecutive time points, i.e., Eq. (3.3) will not be valid anymore. This simply means that $\gamma_j \nrightarrow 0$ i.e., β_j does not tend to α_j, which in turn produce a very large fluctutation in the values of $S_{j|j+1}$. Then, these fluctuations are captured by σ_S. The statistically most significant fluctuations would correspond to dynamical transitions, and are identified by means of statistical signifi-

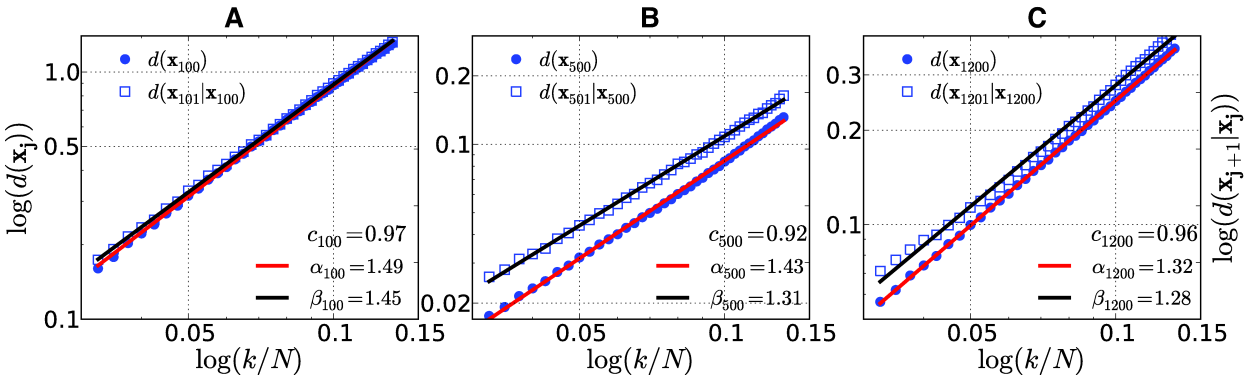

Figure 3.3 Scaling laws in Rössler system for three different randomly chosen time points taken over a short time series of length $N = 1500$. See Eq. (3.4) for red lines and Eq. (3.5) for blue lines. Note that $c_j \approx 1$ as there are no dynamical transitions.

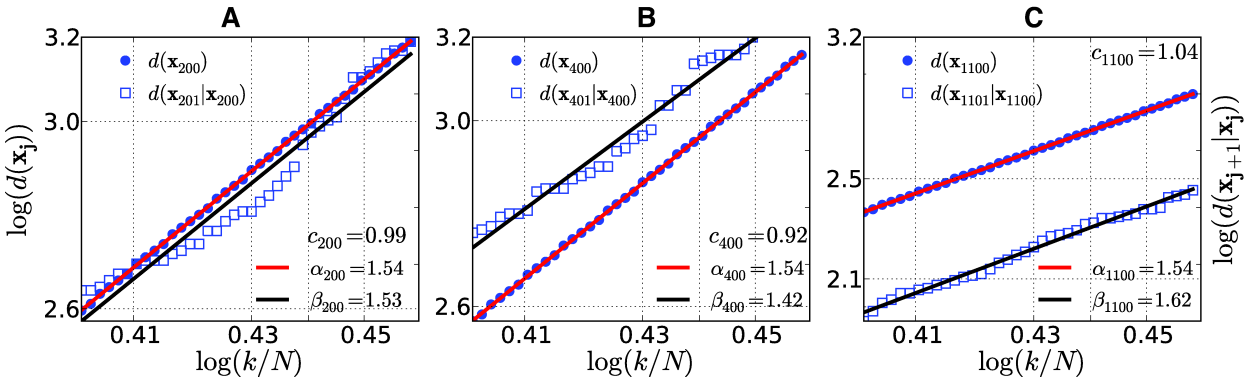

Figure 3.4 Scaling laws in the Logistic map for three different randomly chosen time points over a short time series of length $N = 1500$. See Eq. (3.4) for red lines and Eq. (3.5) for blue lines. Note that $c_j \approx 1$ as there are no dynamical transitions.

cance tests as described later in the following text. I will continue this discussion about the analytical properties of $S_{j|j+1}$ and its mean and variance in the Section 3.3.3.

As a numerical example in support of Eq. 3.11, I choose to analyse dynamical transitions in the Logistic map :

$$x_{i+1} = r x_i (1 - x_i). \tag{3.12}$$

with a varying bifurcation parameter r from 3.5 to 4 with step size 0.0005. I consider time series of 1000 points at each value of r (after removing 500 transient points). As shown in Fig. 3.2, all bifurcations identified by the similarity measure have been confirmed by the Lyapunov exponents λ. Furthermore, σ_S indicates even more subtle transitions (i.e., chaos-to-chaos with laminar phases), which are not uncovered by λ. This is because σ_S depends on the structure of the attractor via D_p (See Eq. 3.11).

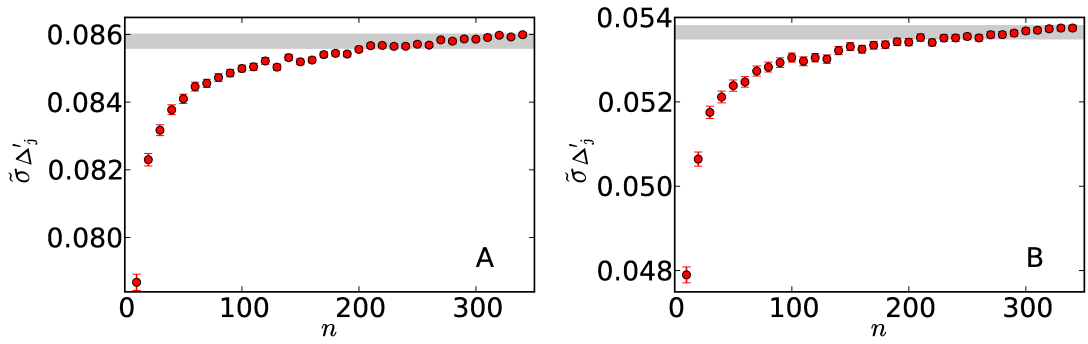

Figure 3.5 Convergence of $\sigma_{\Delta'_j}$ into small range of width 0.0004. $\tilde{\sigma}_{\Delta'_j}$ is the median over over 10,000 bootstrap realisations of $\sigma_{\Delta'_j}$ and the error bar gives the corresponding standard error in the estimation of median. (A) Rössler system, (B) Logistic map (in both cases, the total length of time series is $N = 1500$.)

3.3.2. Numerical examples of scaling laws

In this section I provide a numerical exposition of the previously stated scaling laws. I employe the Rössler system and the Logistic map for the purpose. The Rössler system I have employed is given by the set of equations

$$\begin{aligned} \dot{x} &= -y - z, \\ \dot{y} &= x + ay, \\ \dot{z} &= 0.3x - 4.5z + xz. \end{aligned} \tag{3.13}$$

I have set the value of parameter $a = 0.39$ corresponding to screw type chaos. Then I take $N = 1500$ values from the x variable and embed it with time delay coordinates as described above. The resulting scaling relationships are shown in Fig. 3.3 for three randomly chosen points. Here I have not introduced any dynamical transition in the time series, so determinism holds for consecutive time points. We observe for all the three points that when determinism holds, $c_j \approx 1$ as expected in discussion above on scaling laws.

Further I employ the Logistic map Eq. 3.12 with $r = 4.0$. Again, I use short times of $N = 1500$ points. In Fig. 3.4 I have shown the corresponding scaling laws for this case. I have kept the parameter r constant over the time series which would mean that determinism exists between the consecutive time points. Also, in this case we again observe $c_j \approx 1$ as expected.

In Sec. 3.3 I argued that $\sigma_{\Delta'_j}$ converges to a constant when taking large enough window size n. In Fig. 3.5 I show this convergence of $\sigma_{\Delta'_j}$. In Fig. 3.5(A) I have considered the Rössler system time series as described above. The median of $\sigma_{\Delta'_j}$ i.e., $\tilde{\sigma}_{\Delta'_j}$ is calculated for window size n taking 10,000 realisation of $\sigma_{\Delta'_j}$ by boot strapping. The value of $\tilde{\sigma}_{\Delta'_j}$ is small and quickly converges to a region of small width of 0.0004. A similar conclusion was reached for the Logistic map in Fig. 3.5(B). This also supports the usefulness of the windowing technique I have used in this work.

3.3.3. Relationship with Lyapunov spectrum

Lyapunov exponents λ_i are the most extensively used measures for a quantitative characterisation of dynamics. Several dynamical invariants are conjectured in terms of them such as Lyapunov dimension. However, a reliable numerical method to estimate λ_i from short time series remains a challenging problem [*Wolf et al. (1985)*], frequently encountered in the palaeoclimate research. The main objective of this section is to understand the new measure $S_{j|j+1}$, its mean μ_S and variance σ_S, in terms of these well known dynamical measures of Lyapunov exponents.

Suppose $\lambda_1^j, \lambda_2^j, \ldots, \lambda_m^j$ are the eigenvalues of the Jacobian matrix $\mathbf{D}\varphi(\mathbf{x}_j)$. If $\mathbf{x}_{j+1} = \varphi(\mathbf{x}_j)$ then the deformation of the infinitesimal $\varepsilon(k)$ ball neighbourhood of $\mathbf{x}_j$ in any direction i will be multiple of $\exp(\lambda_i^j)$ (see Fig. 3.6). Defining $\Lambda_i^j = \exp(\lambda_i^j)$, where Λ_i^j are called the *Lyapunov numbers*. The *local Lyapunov exponents*, λ_i are given by

$$\lambda_i = \frac{1}{N}\sum_{j=1}^{N} \lambda_i^j. \tag{3.14}$$

The *global Lyapunov exponent* L_i corresponding to the i-th direction is the asymptotic value of λ_i

$$L_i = \lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} \lambda_i^j. \tag{3.15}$$

If the distance metric used for calculation of $d(\mathbf{x}_j)$ is the Euclidean then a simple geometrical consideration yields

$$d(\mathbf{x}_{j+1}|\mathbf{x}_j) = \left(\frac{1}{N}\sum_{i=1}^{m} \Lambda_i^{j\,2}\right)^{\frac{1}{2}} d(\mathbf{x}_j),$$

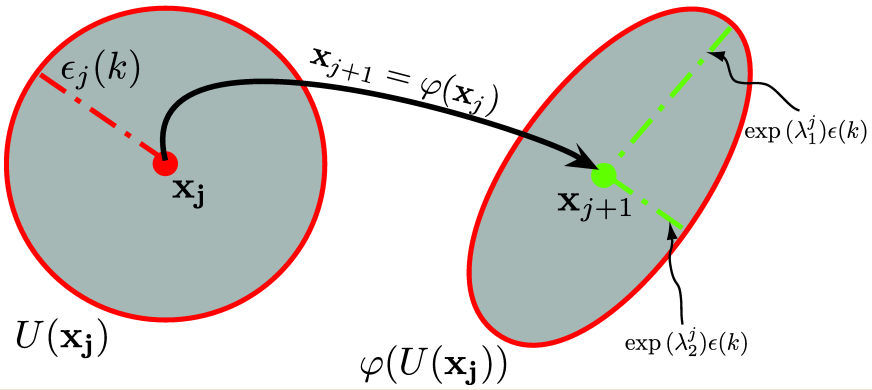

Figure 3.6 Evolution of the $\varepsilon(k)$ neighbourhood of the time point $\mathbf{x}_j$ and into an ellipsoid by the application of the smooth mapping φ such that $\mathbf{x}_{j+1} = \varphi(\mathbf{x}_j)$. The expansion or contraction in any direction i is a multiple of $\exp(\lambda_i^j)$ where λ_i^j are the eigenvalues of $\mathbf{D}\varphi$ at j.

which directly leads to

$$S_{j|j+1} = \left(\frac{1}{N} \sum_{i=1}^{m} \Lambda_i^{j2} \right)^{-\frac{1}{2}}. \tag{3.16}$$

Hence, $S_{j|j+1}$ measures the total deformation of the $\varepsilon(k)$ ball neighbourhood of point $\mathbf{x}_j$ when a mapping φ is applied on it. These deformations should be substantially smaller if there is no dynamical transition because a dynamical transition by definition means that there does not exist φ, i.e., $\mathbf{x}_{j+1} \neq \varphi(\mathbf{x}_j)$. *This leads to large fluctuations in the magnitude of* $S_{j|j+1}$, *which will be captured by* σ_S. Though, a point to note is that the global Lyapunov exponent L_i are not defined for a time series with a transition, due to the constraints imposed by ergodicity.

The expansion in i-th direction, $\lambda_i^j > 0$ i.e., $\Lambda_i^j > 1$ (respectively if contraction $\lambda_i^j < 0$, i.e. $\Lambda_i^j < 1$) has the most dominant contribution to $S_{j|j+1}$ in Eq. (3.16) since it comes from the largest positive eigenvalue corresponding to the expansion. Therefore, I conclude that the mean of $S_{j|j+1}$, i.e., μ_S over small n points will be structurally the same as the sum of the largest local Lyapunov exponents, while μ_S over large n will be structurally the same as the sum of largest global Lyapunov exponents. This is shown numerically for the Logistic map in Fig. 3.7. We know the sum of the largest Lyapunov exponents is conjectured to be related with dynamical invariants such as Lypaunov dimension and topological entropy [*Ott (2002)*]. Numerical examples show that μ_S is not well suited for detecting dynamical transitions in the series because it is less sensitive in quantifying and capturing large fluctuations in $S_{j|j+1}$. It will be better to use σ_S for the purpose. More so, σ_S is better suited to quantify variation of parameters of a dynamical system and corresponding changes in the dynamics because any variation in the parameters of dynamical system will lead to variation in the Lyapunov spectrum too.

To get further insight into the properties of σ_S I take a look at the distribution of S (subscript dropped for simplicity) over window size of n vectors, $P(S,n)$. We know that λ_i^j are in a

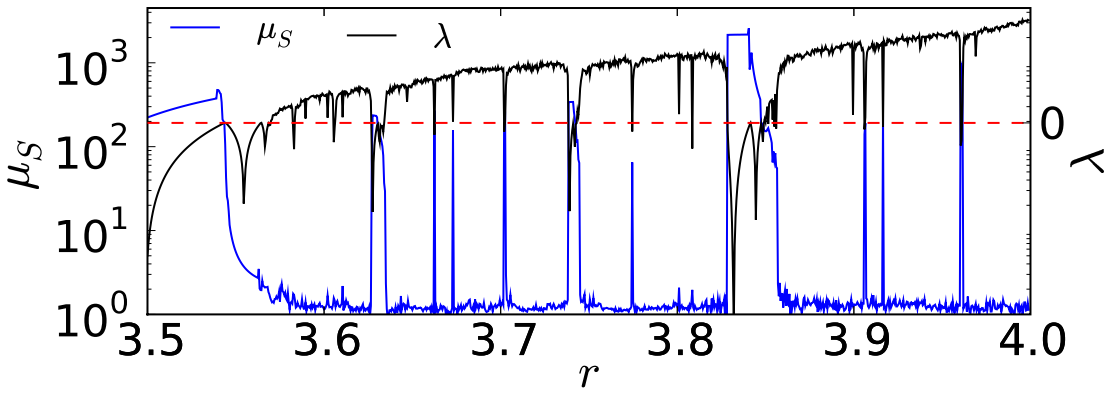

Figure 3.7 One-to-one correspondence between μ_S (blue line) and the Lyapunov exponent λ (black line) for the logistic map. The parameters used for this figure are the same as used for Fig. 3.2

sense random numbers for chaotic systems. The expression of S according to Eq. (3.16) consists of summation over $\Lambda_i^j = \exp(\lambda_i^j)$, therefore, the central limit theorem implies that S follows a Gaussian distribution at least asymptotically. Following [*Ott (2002)*] I reason that asymptotically, $P(S,n)$ has the following general analytical form,

$$P(S,n) \sim \frac{1}{\sqrt{2\pi n \Phi''(S)}} \exp(-n\Phi(S)), \tag{3.17}$$

where $\Phi(S)$ is a convex quadratic function with minimum zero, occurring at $S = \mu_S$ i.e., $\Phi(\mu_S) = 0$ also, $\Phi'(\mu_S) = 0$, $\Phi''(\mu_S) > 0$. Expanding $\Phi(S)$ around μ_S and neglecting higher order terms, I rewrite Eq. (3.17) as,

$$P(S,n) \sim \frac{1}{\sqrt{2\pi n \Phi''(S)}} \exp(-n\Phi''(S)\frac{(S-\mu_S)^2}{2}), \tag{3.18}$$

which gives the familiar looking form of a Gaussian distribution. Then $\sigma_S = (n\Phi''(S))^{-1/2}$. A similar expression for distribution and variance could also be written for local Lyapunov exponent [*Ott (2002)*], where $\Phi(\lambda_i)$ is known as the spectrum of local Lyapunov exponents and can be used for characterising the dynamics of the system [*Grassberger et al. (1988), Sepúlveda et al. (1989), Prasad and Ramaswamy (1999)*]. So, as an analogy I propose that $\Phi(S)$ can also be used to characterise the dynamics. The distribution of S for different types of dynamics may follow a Gaussian distribution of the type $P(S,n)$ asymptotically but each type of dynamics must correspond to unique μ_S and σ_S. This is because of the fact that each type of dynamics has a unique φ and hence unique eigenvalues and the corresponding deformations and values of S should also be unique. In future research I intend to develop a method based on estimation of $\Phi(S)$ to classify distinct dynamics.

3.4 Numerical examples for dynamical transitions

In this section I apply the similarity measures to more realistic situations where I have a short time series (of length N) obtained from a dynamical system whose parameters vary in a certain way with time.

3.4.1. Significance test

In all of the numerical studies below, I use the following procedure: I first calculate $S_{j|j+\tau}$, based on which the variation of σ_S is computed with sliding windows of size k. The temporal profile of σ_S is therefore used to disclose possible hidden transitions of complexity changes. I apply a bootstrapping method [*Palus (2008)*] to test the null hypothesis that the similarity measure represented by $S_{j|j+\tau}$ is uniform for the original process. I create the null-model by randomly drawing k values with replacement from the series of $S_{j|j+\tau}$, where k is the

window size. Repeating this procedure many times (e.g., 10,000 times throughout the paper), I construct an ensemble of σ_S^i. The 0.05 and 0.95 percent quantiles of this ensemble could be interpreted as 90% confidence bounds. The time points where the values cross this bound are less likely to happen, hence indicating large transitions occurred at these time points.

3.4.2. Transition due to nonlinear and abrupt changes in parameters

As a paradigmatic example, I use the Rössler model (Eq. 3.14). The system shows two distinct chaotic attractors when increasing the parameter a from 0.32 to 0.39, respectively corresponding to the spiral and screw type chaos. The transition between both chaotic regimes is due to the formation of a homoclinic orbit (Eq. 3.14). In the first case I implement a transient variation of the control parameter a by $a(t) = 0.32 + 0.07|\sin(\frac{\pi}{600}t)|$ at each integration step, where t is the step size for the fourth order Runge-Kutta integrator ($t = 0.001$). The sampling time between two consecutive points is $\delta t = 0.2$ and I use 6000 points of the x-component (Eq. 3.14) (Fig. 3.8). In the case of nonstationarity (i.e., due to time varying parameter a), overembedding techniques are necessary and widely used, as it was suggested that $m > 2D + p$, where D is the dimension of the attractor and p is the number of varying parameters [*Hegger et al. (2000)*]. Therefore I choose the embedding parameters $m = 10$ and $L = 15$. I calculate σ_S by sliding window of size $k = 300$ and 90% overlapping between two windows based on the similarity series of $S_{j|j+\tau}$. σ_S varies significantly from lower to higher values and vice-versa (highlighted by grey bands in Fig. 3.8A), which indicate slow changes in the complexity of dynamics between the two distinct chaotic states.

In the next case, the parameter a switches between 0.32 and 0.39 after 1500 sampled points, yielding abrupt transitions from one dynamical complexity to the other. The other parameters for calculating σ_S are the same as in the previous case. σ_S convincingly captures the abrupt transitions that were introduced (Fig. 3.8B). It is important to emphasise that this method works analogously in situations where I introduced many missing values in the time series to mimic the case I have in the proxy palaeo records below.

I have successfully tested the method on several other numerical models (e.g., Baker's map where the problem is to identify nonstationarity due to slow parameter drift [*Schreiber (1997)*] and the Lorenz system). At this point I presume that a sharp point of inflection in σ_S indicates a change in the complexity of the system, as the fluctuation level of similarities drastically changes at this point. Next I will apply this procedure to an observational time series to identify possible transition points.

3.4.3. Identifying nonstationarity

One of the challenging problem could be identifying a continuous drift in the dynamics of a time series. For this purpose I use the generalised Baker's map [*Farmer et al. (1983)*] and

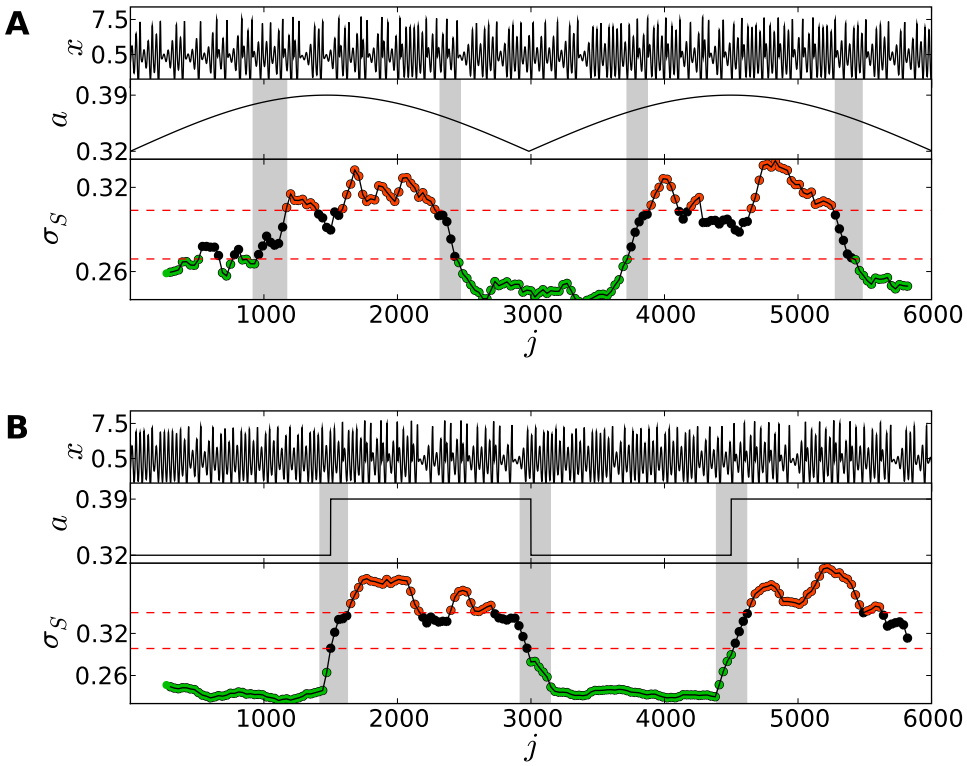

Figure 3.8 Transition between spiral and screw type chaos in the Rössler system (Eq. 3.14). (A) slow nonlinear approaching to transition points. (B) abrupt transitions. Time series of the x-component are shown in the upper panels of both (A, B), middle panels show the variation of a, lower panels are σ_S. Horizontal dashed lines in the lower panels of both (A, B) are 90% confidence bounds. The values of σ_S above and below the band are coloured in orange and green respectively and black if lies inside the band. This is done to the highlight the regimes of two distinct dynamical complexity. Note that crossover of the significance levels happen only around the points of dynamical transition (highlighted by vertical grey bands).

generated a time series following the same procedure as described in Ref. [*Schreiber (1997)*].

$$\begin{aligned} \text{if } v_i \leq \alpha: \; u_{i+1} &= \beta u_i, \;\; v_{i+1} = v_i/\alpha, \\ \text{if } v_i > \alpha: \; u_{i+1} &= 0.5 + \beta u_i, \;\; v_{i+1} = \frac{(v_i - \alpha)}{(1-\alpha)}. \end{aligned} \tag{3.19}$$

The Lyapunov exponents for the above set of equations are

$$\lambda_1 = \alpha \ln \frac{1}{\alpha} + (1-\alpha) \ln \frac{1}{1-\alpha}, \tag{3.20}$$

$$\lambda_2 = \ln \beta. \tag{3.21}$$

I create a drift in the parameter β as done in [*Schreiber (1997)*], namely, by generating a time series of length 15,000 by varying β in each iteration by $\beta = i/15{,}000$ and fixing the value

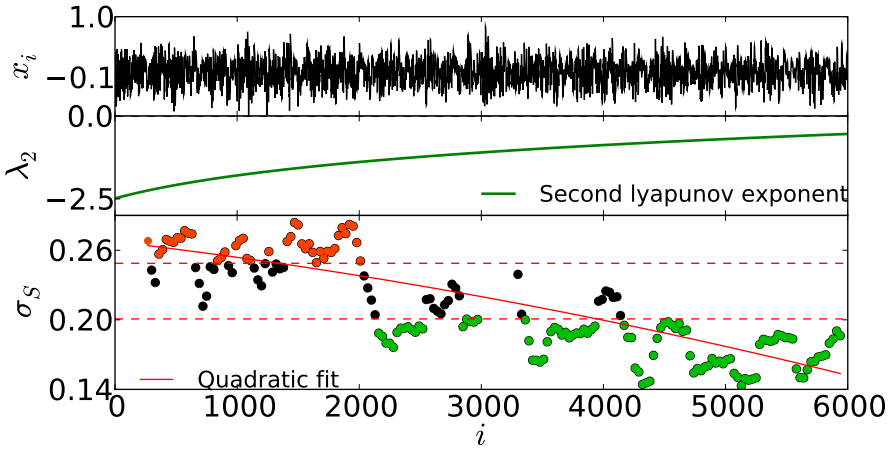

Figure 3.9 Identifying drift in the dynamics of the Baker's map with continuously changing parameter β: According to Eq. (3.20), λ_1 is constant; λ_2 is equal to $\ln \beta$, so it varies continuously in a nonlinear way. We observe σ_S changing from significantly higher values to significantly lower values as the time progress. A quadratic fit describes this evolution (red curve), indicating the nonlinear change of parameter of the system.

of $\alpha = 0.4$. This creates a nonstationary time series with drift in dynamics while the maximal Lyapunov exponent λ_1 is constant given $\alpha = 0.4$ (Eq. 3.20). The trend from the time series is removed by taking

$$x_i = \frac{w_i - \langle w \rangle_k}{\sqrt{\langle (w_i - \langle w \rangle_k)^2 \rangle_k}},$$

where $w_i = u_i + v_i$ and I took $k = 50$. I consider only a short section of the time series by taking points from $i = 1000$ to $i = 7000$, which means I have considered only 6000 time points. However, in the original work of [*Schreiber (1997)*], 40,000 data points were used. I use embedding parameter $m = 5$ and $L = 2$. The result is show in the Fig. 3.9, where I observe that values of σ_S goes from significantly higher values to significantly lower with the evolution of the parameter, which is representative of the drift in dynamics, that has taken over the time points.

3.4.4. Transitions involving strange non-chaotic attractors

Strange non-chaotic attractors (SNA) are observed in quasi-periodically driven dissipative dynamical system [*Yalçinkaya and Lai (1997), Lai (1996)*]. Here I apply the presented method for identifying dynamical transitions to and from SNA. I will attempt to identify transitions in

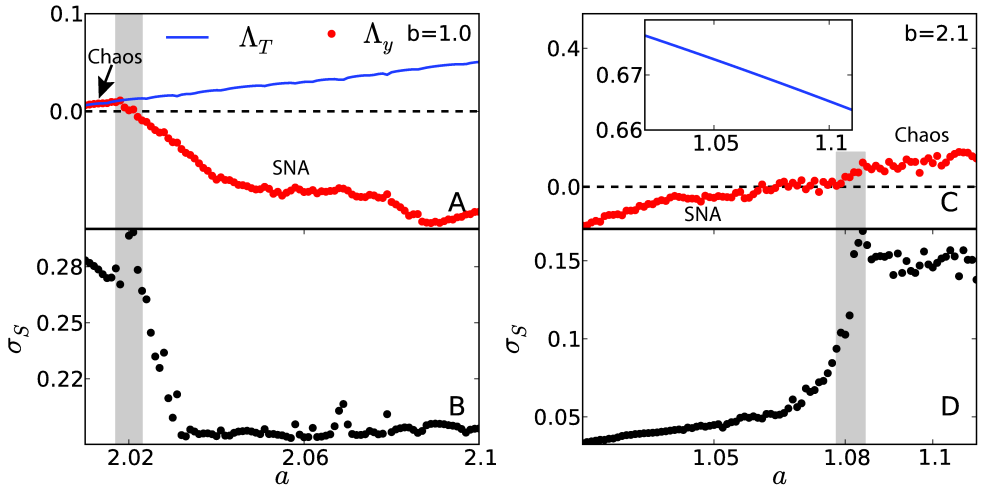

Figure 3.10 Transitions to SNA via intermittency.(A) Lyapunov exponents; (B) σ_S shows large fluctuations right before the SNA appears. Respectively, (C, D) show a transition from SNA to chaos.

coupled map of the form:

$$\begin{aligned} x_{i+1} &= (x_i + 2\pi\omega) \mod (2\pi), \\ y_{i+1} &= \frac{1}{2\pi}(a\cos(x_i) + b)\sin(2\pi y_i). \end{aligned} \tag{3.22}$$

The two types of Lyapunov exponent viz, largest transverse Lyapunov exponent Λ_T and largest Lyapunov exponent Λ_y of the subsystem y, are given by the following set of equations [*Yalçinkaya and Lai (1997), Lai (1996)*]

$$\begin{aligned} \Lambda_T &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \ln|a\cos(x_j) + b|, \\ \Lambda_y &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \ln|a\cos(x_j) + b\cos(2\pi y_j)|. \end{aligned} \tag{3.23}$$

It is known that in the case of $\Lambda_T > 0$ and $\Lambda_y < 0$ we have SNA while in the case of $\Lambda_T > 0$ and $\Lambda_y > 0$ we have chaotic regime. In Fig. 3.10 (A, B) the grey band represents the transition to SNA from chaos. The transition occurs via on-off intermittency, i.e., close to the critical value we observe intermittency. The values of Λ_y fluctuate and finally become negative.

I generate short time series of length $N = 2000$ at 100 different values of a separated by 0.002. Then I calculate σ_S using embedding parameters $m = 5$ and $L = 2$. In Fig. 3.10 I have plotted σ_S with the Λ_T and Λ_y. As Λ_y starts to increase and fluctuate around the zero value, σ_S also increases representing the change of dynamics from SNA to chaos (Fig.3.10 (C, D)).

3.5 Application to Pleistocene record of south Asian Monsoon

One of the most important monsoonal system on the Earth is the south Asian summer monsoon which affects the life of billions of inhabitants of the region. Its origin lies in the temperature gradient between the land and the sea that ceases to exist during the peak of summer, which in turn leads to a pressure difference between the land and the sea resulting in strong moisture laden winds blowing from the high pressure areas of the sea to the low pressure areas of the land (See Fig. 3.11A). The lithogenic grain size (LGS) in marine sediments is an ideal proxy for the *wind* intensity (carrying capacity) over geological time scales which can be used for studying the variability of the monsoon in the past. One of the longest records for the Asian monsoon is a LGS record from the Ocean Drilling Program (ODP) site 722B extending for the last 3.41 million years (Ma) before present [*Clemens et al. (1996)*]. It has a sampling rate of 0.001 Ma but with a large number of missing values in the series (see Fig. 3.11B, also see A.4.2 and *Clemens et al. (1996)* for data source).

To get over the problem of missing values, I replace the missing values with flags (e.g., large number 9999) which act as place holders. The time series was normalised to zero mean and unit variance before placing the flags. As LGS is a proxy of a physical process with high dimensions and large parameter space, I embed the time series in a high dimensional space by taking embedding parameters $m = 30$ and $L = 5$, i.e., a vector in embedded space spans 0.15Ma. Such a high dimensional space (in a range of $15 \sim 30$) was suggested to minimise the possible effects of both nonstationarity and noise present in time series [*Hegger et al. (2000)*]. I used a window size of $n = 100$ (0.1 Ma) with 90% of overlapping of $S_{j|j+1}$. I present a 90% confidence bound as estimated by bootstrapping. The values of σ_S above and below the significance band separate out two distinct dynamical regimes. The time points where crossing of significance band takes place are the points where dynamical transitions have occurred. In the parlance of palaeoclimatic terminology, the values of σ_S for the LGS data set could be interpreted as patterns of palaeoclimatic variability of the south Asian monsoonal system on the orbital time scales. The transitions identified by σ_S are highlighted by grey bands in Fig. 3.11D and summarised in Tab. 5.1, which also shows a comparison to the results reported by linear time series methods in the literature on similar data sets.

Band's index	1	2	3	4	5	6	7	8	9	10
σ_S on LGS	0.14 – 0.27	0.9 – 1.0	1.1 – 1.15	1.23 – 1.28	1.67 – 1.75	1.86 – 1.91	2.04 – 2.1	2.3 – 2.39	2.52 – 2.67	2.72 – 2.79
Ref. 1 on LGS	–	–	–	~ 1.2	~ 1.7	–	–	–	~ 2.6	–
Ref. 2 on Dust flux	–	~ 1.0	–	–	~ 1.7	–	–	–	–	~ 2.8

Table 3.1 Comparison of transitions identified by σ_S and by other approaches reported for similar data sets. Time unit is Ma. Ref. 1 is [*Clemens et al. (1996)*] and Ref. 2 is [*deMenocal (1995)*]

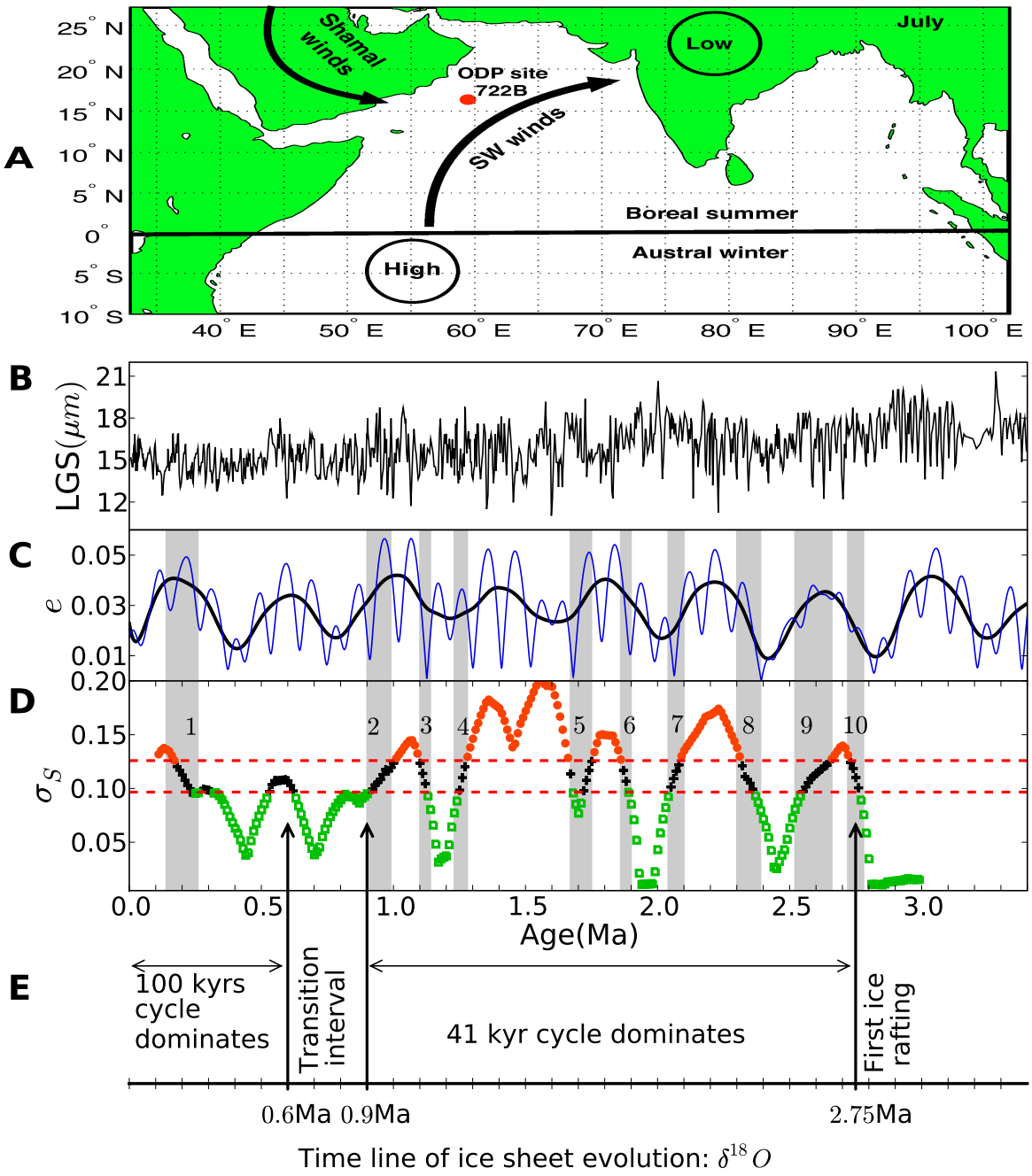

Figure 3.11 (A) Map of winds in July from high pressure in the Indian Ocean to low pressure over the land. The bullet dot denotes the location of the ODP site 722B ($16°37'$N,$59°48'$E, 2028 meters below the sea level). (B) LGS time series. (C) The blue line is the orbital eccentricity e and the dark bold line is the associated 400 kyr component. (D) σ_S. The 90% confidence bound is represented by horizontal dashed lines. Orange (dots) and green (open squares) separate the regimes of two different dynamical complexity. (E) Time line of the NH glaciation as reconstructed from δ^{18}O record [*Raymo (1994)*]. Vertical arrows indicate the important events in the ice history, which are adapted from [*Ruddiam (2008)*]. See the text for details of bands in (C, D).

The main external forcing at the orbital time scales (10^6 years) on the south Asian monsoonal system is the change in the solar insolation induced by the cyclic variability of orbital parameters of Earth. Internally, the evolution of monsoon has also been effected by changes in the Northern Hemisphere (NH) ice sheets and the sea surface temperatures [*Clemens et al. (1991), Pausata et al. (2011)*]. In Fig. 2.2 I give a flow diagram for the above major forcings, highlighting the mechanism of their influences on the monsoonal system. The dominant factor in the astrodynamical calculations of the insolation received by the earth is $e\sin(\beta)$, where e is the eccentricity and β is the precession [*Laskar et al. (2004)*]. The high number of missing values and low temporal resolution (> 0.001Ma) of LGS data set do not allow to study the influences of the higher frequency variation of solar insolation on monsoon. Therefore, I have concentrated on comparing the dynamical changes and transitions that have occurred in the monsoon over the time scales of cyclicity in eccentricity of Earth's orbit (400 kyr and 100 kyr). In the original analysis of [*Clemens et al. (1996)*] (besides LGS data), the authors concentrated on coherence between the monsoonal variability and the insolation forcing associated with the obliquity and precession bands. The relations to the eccentricity have not been discussed seriously, since their methods generally could not yield interpretable results, which they hence argued the direct insolation forcing associated with eccentricity is negligible for understanding monsoon strength. As a matter of fact, the orbital eccentricity e characterises the lowest frequency variation of the solar cycle, modulating the precession on two different time scales, 100 kyr and 400 kyr, which have been observed in many palaeoclimate data sets, e.g., dust flux records [*Trauth et al. (2009), deMenocal (1995)*]. It has been noted in *deMenocal (2004)* that the highest palaeoclimatic variability in ODP records from the region occurs during the periods of maximum orbital eccentricity e. I show e by a thin blue line in Fig. 3.11C together with its 400 kyr component as a dark bold line. Because of the length and resolution of the available LGS data, I compare the transitions captured by σ_S to the variabilities of both the 400 kyr cycle and glaciation of the NH (Fig. 3.11C, E, respectively). The δ^{18}O is one of the best proxies reflecting global surface *temperature*, serving as an indicator of the global ice volume variability [*Raymo (1994)*] (Fig. 3.11E). Based on the information provided by δ^{18}O record, I have marked the important events in the glaciation time line of NH by vertical arrows in Fig. 3.11D, E. δ^{18}O data used in the analysis is from the reference *Tiedemann et al. (1994)* (also see A.4.3).

From Fig. 3.11D and Tab. 5.1, we find that most of the transitions identified essentially show the basic response mechanism of the monsoonal system to the external solar forcing, with some exceptions. The response behaviour is captured by the measure σ_S, which reveals a strong relationship with the 400 kyr component of the eccentricity e. Higher values of e lead to higher values of σ_S and lower values of e imply lower values of σ_S, indicating σ_S varying between two different phases of complexity. In Fig. 3.12 I have provided a scatter plot between the 400 kyr component of the eccentricity e (e^{400}) and σ_S. We can see a kind of threshold type behaviour of monsoon with respect to changes in e^{400}. No point of high complexity (dots) is observed for $e^{400} < 0.023$. Similarly no point of low complexity (open squares) is observed for $e^{400} > 0.037$. Which hints towards that higher complexity in dynamics of monsoon tends

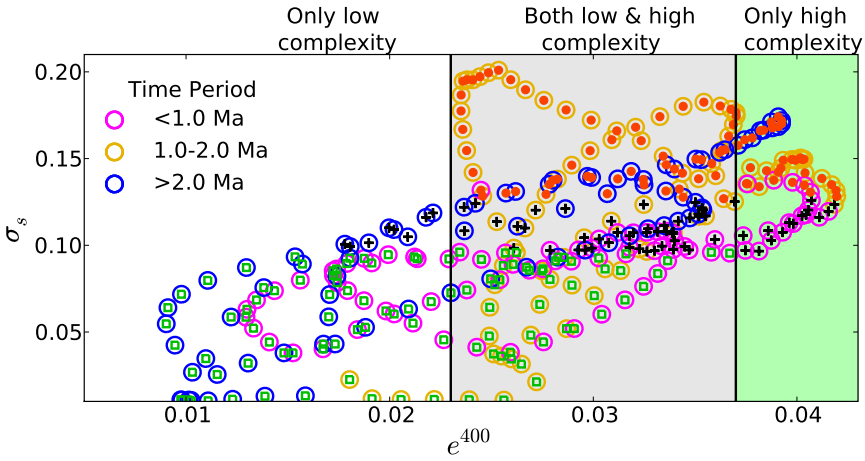

Figure 3.12 Scatter plot between σ_S and 400 kyr cycle of e. Green (open squares), black (plus) and orange (dots) points are as same as in Fig. 3.11 and represent the low complexity dynamics, the statistically insignificant values and the high complexity dynamics respectively. The outline colour of each point shows its time period. See the text for details.

to occur during the high insolation periods. In the range $0.023 < e^{400} < 0.037$, it is seen that both kind of dynamics are possible and in this range I think that the higher frequency variation of solar insolation and internal feedbacks of monsoonal system must be important in determining the dynamical regimes.

We can observe in Fig. 3.11 D that mostly low complexity points (open squares) occurred before 1.0 Ma. The response to solar insolation is subdued in this time slice as reflected by small variation of σ_S compared to other parts of its variation in the series. This is the period of glaciation in NH and most parts of the Earth were cooler than modern state [*Ruddiam (2008), Raymo (1994), Ravelo (2006)*]. This indicates towards the possible important role played by NH ice sheet evolution on the evolution of modern summer monsoon over south Asia. Furthermore, we notice two other important transitions which are denoted by the bands-4 (~ 1.2) and 10 ($\sim$2.75). The former was supported by the transition of the ice cycles [*Clemens et al. (1996)*], while the latter one corresponds to time point when the first ice sheet appeared in the NH. After 1.2 Ma relatively large deviations were observed in σ_S. This was the period of development of strong Walker circulation and after 2.0 Ma low and mid-latitude regions of earth were warmer than modern state, hinting the plausibility of stronger variation in monsoon dynamics in the warmer environments.

To further solidify the analysis, in Fig. 3.13 I have studied the windowed cross-correlation ρ between σ_S and the other three parameters viz., e^{400} (400 kyr cycle of e), e (the full e i.e., 100 kyr cycle) and δ^{18}O. In Fig. 3.13 (A-B) we can see high positive correlations between σ_S and e^{400}, and also between σ_S and e. Suggesting that the dynamical complexity of the monsoon responds in an almost linear way to the changes of the solar insolation. The relationship be-

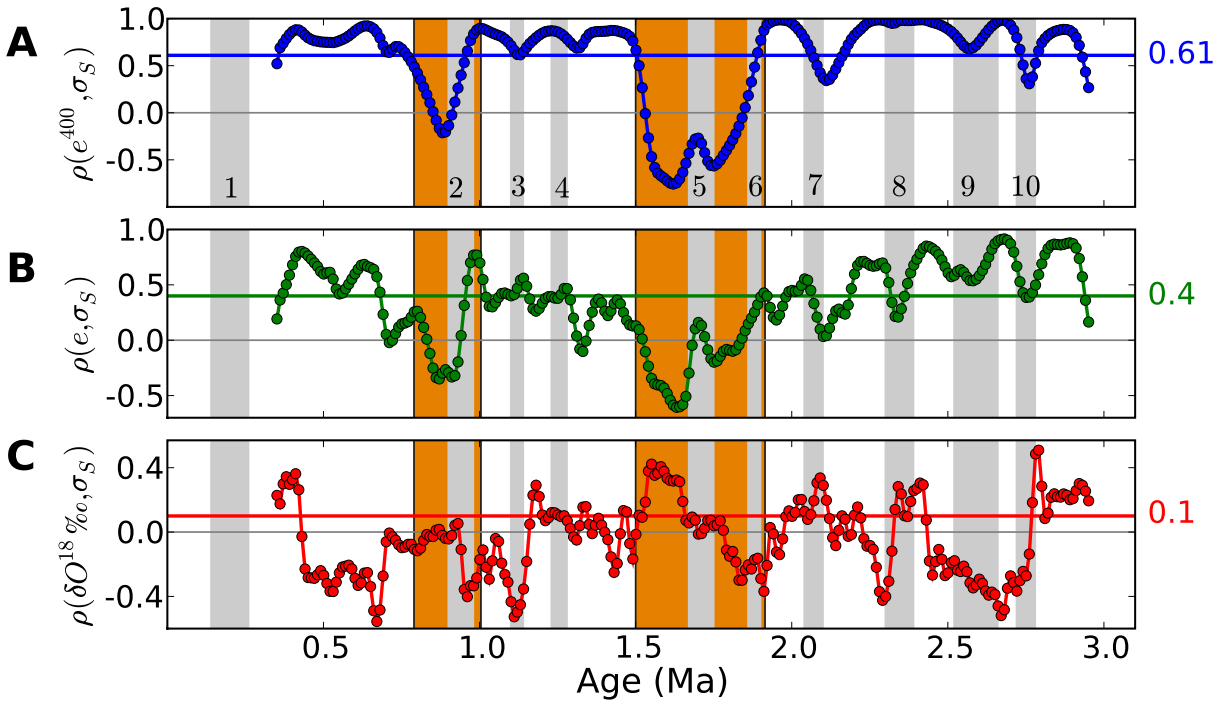

Figure 3.13 Windowed correlation: the cross-correlation between σ_S and e^{400} (A), e (B) and δ^{18}O record (‰) (C). The correlation coefficient ρ is calculated over 25 points using 99% overlap. The horizontal coloured lines correspond to cross-correlation for the each indicated parameter over the whole series. The values are quoted in the same colour on right hand side. The two vertical orange bands, indicate sharp drop in $\rho(e^{400},\sigma_S)$ and $\rho(e,\sigma_S)$. Grey bands are same as in Fig.3.11. Transitions 3,7,8,10 also show sharp changes in $\rho(e,\delta^{18}\text{O})$

tween σ_S and e suggests that the solar insolation acts as an external varying forcing parameter for the south Asian summer monsoon system, where slight variations induce large changes in the dynamical complexity of the monsoon. Also, we observe two very sharp fluctuations in $\rho(e^{400},\sigma_S)$ and $\rho(e,\sigma_S)$ (orange vertical bands in Fig.3.13), where these ρ simultaneously change sign. One such transition occurred around $0.8-1.0$ Ma, where $\rho(e^{400},\sigma_S)$ and $\rho(e,\sigma_S)$ have values very close to zero. This means that the dynamical responses of monsoon became independent of solar forcing during this period. A plausible cause for the emergence of such a feature could be the known fact that this was the period of transition in the cyclicality in NH ice sheets [*Raymo (1994)*]. The variation in NH ice sheet could modulate monsoon via the changes in patterns of planetary albedo (see Fig. 2.2). The other transition happened around $1.5-1.9$Ma (see orange band in Fig. 3.13). Here the values of $\rho(e^{400},\sigma_S)$ and $\rho(e,\sigma_S)$ become large negative, implying that monsoon dynamics responded in characteristically opposite way to its usual response to solar forcing and its variations. A possible reason could be the development of strong Walker circulation during this period [*Ravelo (2006)*]. The influence of the intensification of Walker circulation could come via the changing sea surface

temperature (see Fig. 2.2). All these exceptions suggest that major changes in internal forcing of the climate system like glaciation cycles and oceanic circulations could significantly disturb the basic response mechanism of the monsoon to the solar insolation. We observe in Fig.3.13 C that correlation between the NH ice cover (δ^{18}O record) and monsoon dynamics (σ_S) is rather weak and with some sharp rises and drops to higher positive and negative correlations corresponding to identified transition bands 3,7,8 and 10. This indicates that the influence of the NH ice cover on the monsoon dynamics is more episodic, i.e., there are a few events when we observe changes in NH ice cover influencing monsoon dynamics in some way.

3.5.1. A note about treatment of missing values, embedding parameters and window size

Apart from the shortness of the data lengths, the other central problems which surround palaeoclimate data analysis are irregular sampling and missing values. The LGS data set I have used in the Sec. 3.5, to study monsoonal evolution is regularly sampled but has a large amount of missing values. To test the performance of the method in the case of missing values I carry out the following procedure to create a toy model.

First I replace the points in the LGS data set by the values of the x variable of the same time length from the Rössler system (as described in Sec. 3.3.2 and also see Eq. (3.14). In turn, I introduced missing values exactly at the same time positions as in the LGS data set. Next, I introduce dynamical transitions to the Rössler system by changing one of its parameters abruptly as done in the for Fig. 3.8 (B). Then I try to identify these transitions using the presented method. All the parameters in the method like embedding dimension, delay, window sizes, significance level remain the same as I have used in the observational data. I treat the missing values as described in the Sec. 3.5 i.e., I replace the missing values with flags (e.g., large number 9999).

In Fig. 3.14 we can clearly distinguish four different dynamical regimes. I have defined a transition as the point where σ_S crosses from significantly lower values to higher values and vice-versa.

I use a rather high embedding dimension, which is due to the fact that the systems I consider have one of their parameters varying with time (such as Rössler system with abrupt and nonlinear transitions and Bakers' map with drift). This converts the systems into non-autonomous systems and Taken's theorem is not valid for such systems. Hence, I cannot take the embedding dimension of $2m+1$ as prescribed by the Taken's theorem. Though there is no specific embedding theorem for such systems, heuristic arguments in *Hegger et al. (2000)* state that a proper choice for the embedding dimension should be larger than $2(m+P)$ where P is the number of time varying parameters of the system. It has been suggested that this technique of "overembedding" a time series helps in overcoming both nonstationarity and noise effects. I continue using high embedding dimension for the LGS data set too, as it is a proxy of a system which not only has high dimensions but also a large parameter space. So, $2(m+P)$ must be a large number. In the LGS data set apart from visible non-stationarity (a

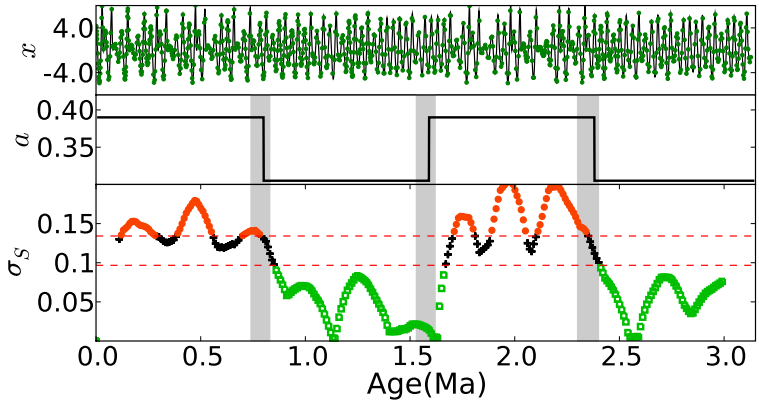

Figure 3.14 Test the effect of missing values on the method: Green points represent the time points sampled from the dynamics (black line) of Rössler system. The time position of sampling comes from the LGS data set. The artificial abrupt transition are introduced by changing the parameter a in Eq. 3.14 from value of 0.32 to 0.39. A fourth order Runge-Kutta integrator with time step of 0.001 was used. The sampling time between two consecutive points is 0.2.

strong trend towards increasing monsoon), the time series also has a high amount of noise which is also visible by eye and via its power spectrum. By taking high embedding I treat the gaps as merely a form of temporal noise. It helps in such a case to have high embedding dimension as I have showed with the example above.

In Fig. 3.5 I show a quick convergence of $\sigma_{\Delta_j'}$ on taking large enough window sizes, which in turn gives the measure dependence on the structure of the attractor through the effective dimension D_p. By taking overlapping windows, I avoid reducing the amount of data appreciably. The presented method differs in one basic aspect from other methods based on recurrence properties. In most of them one first takes a window over the data (or embedded vectors) and then calculate some measure based on the recurrence property. This brings the relationship between windowing, dimension and delay. In my method I follow a different approach : first of all the recurrence distances of a point over the whole time series are calculated and, then, by comparing each consecutive time point I calculate a measure $S_{j|j+1}$ for each point. Till this step I have no windowing. In the next step I calculate the fluctuations in this measure by taking windows. The way I have defined the significance test the window size now helps in resolving time scales on which we wish to see the transitions. The only task of windowing is to give control over resolving time scales for transitions.

3.6 Summary

In summary, I propose a new approach, viz. the fluctuation of similarity, for identifying transitions of regimes of distinct dynamical complexity in a short time series. I show that standard deviation of these fluctuations has dependence on the complexity of the dynamics. I have also

derived the relationship between the Lyapunov spectrum and the similarity measure. Hence relating the newly introduced measure with the underlying dynamical invariants of the system. Additionally I provide an application of the method to a palaeo-pleistocene record of the south Asian monsoonal system. The above results suggest that complexity in the dynamics of south Asian summer monsoon has linear dependence on the variations in solar insolation due to changes in orbital parameters. The existence 400 kyr cycle is very apparent in the LGS data set analysed. Episodic influences of changes in the NH ice sheets and oceanic circulations have also been observed. I stress that the presented method has broad applications for a general purpose as I have demonstrated by both several prototypical model systems and experimental time series.

Chapter **4**

Spatial correlations, structures and directionalities in monsoonal rainfall extremes over south Asia

The invalid assumption that correlation implies cause is probably among the two or three most serious and common errors of human reasoning.

– Stephen Jay Gould*

4.1 Introduction

As we have discussed in the Sec. 2.3, monsoonal rainfall manifests itself over the south Asian region with immense spatial variability and complexity. The Indian monsoonal system covers some of the wettest to the driest regions on the Earth. The north east and western coast along the Western Ghats are the wettest regions, whereas north west is the driest. We observe that rainfall regions on an annual scale ranges from below 150 mm per season to above 2100 mm per season during summer monsoon (JJAS months) (see Fig. 2.3). Several factors derive this spatial variability and complexity ranging from large scale monsoonal circulation to synoptic and meso-scale weather systems and their interaction with the complex topography of the region. In the next two chapters my aim will be to extract the finer details of spatial correlations in rainfall events that might emerge due to the interaction among different atmospheric processes on distinct time scales. Hence, not only providing new physical insights into these interactions but also presenting some new methodological developments in the analysis of large spatio-temporal data sets.

In this chapter I will introduce event synchronization, a measure developed for estimating nonlinear correlations between the rain events occurring at two distinct geographical locations. Analysis of these correlations can provide vital information about the atmospheric processes that are responsible for the generation of rainfall during the active phase of ISM. Here,

* He was an American palaeontologist, evolutionary biologist, and historian of science.

I will attempt to analyse them using the methodology of hierarchal clustering. The clusters thus found will represent the spatial regions where the specific type of rain events occur either synchronously or in a lag synchronised way. In this way I have attempted to regionalise the precipitation field, but not based on similarity in monthly variability as has been previously endeavoured with several different methods [*Iyengar and Basak (1994)*, *Gadgil et al. (1993)*, *Gutiérrez et al. (2006)*]. Rather based on the similarity of dynamics of daily rainfall, which is determined by estimating the strength of synchronous rainfall activity over large parts of the land. Further using the delay behaviour of rainfall events, we estimate the directionalities related to the progress of such type of rainfall events. A comparison of the delay direction with winds during ISM is also presented. One of the important climatological aims here will be to understand the spatial manifestation of the inter-seasonal oscillations (ISO) i.e., the "Active Phase" and "Break Phase". During an active period heavy rainfall appears over most parts of India; in a break period the phase reverses and rainfall is recorded only along the Himalayan foothills, south-east peninsular India [*Singh et al. (1992)*, *Krishnmurthy and Shukla (2000)*]. I will attempt to identify the regions where the active phase of monsoon brings rainfall in form of higher order of synchronization.

In the next section I will introduce the data sets used in the analysis Sec. 4.2. Discussion on the climatology of the region has already been provided in Chapter 2. In the Sec. 4.3 I will present the concept of event synchronization, the algorithm used for clustering and the method used to derive directionalities from the data . Then I will proceed with results, discussion and summary in Sec. 4.4.

4.2 Data

The main data set I have used is a daily precipitation described in the Appendix A Sec. A.1.2. This particular data set will be referred to as APHRO–V0902 in the following text. For the purpose of comparing the results I have also used another high resolution precipitation data set described in the Appendix A Sec. A.1.3. It will be referred as IMD-D. For further comparison with wind directions during certain type of rainfall events, I have used the data set described in Appendix A Sec. A.2.

Next I will introduce the special methods used in the data analysis for identifying those regions of rainfall which show a higher level of coherence in terms of rainfall activity.

4.3 Methods

4.3.1. Event synchronization (ES)

In this section I will describe the concept of ES [*Quiroga et al. (2002)*]. Let R_i^j be the net amount of rainfall received at grid point j on the i^{th} day during the JJAS months of any year in the data. To classify events into a "heavy rainfall event" or a rainfall event which is a result of

increased monsoonal activity, the threshold T_j on a rain event at the grid point j is calculated by taking the α_j percentile of rain events occurring at grid j of all wet days ($R_i^j > 0$). Next, I determine the time indices of events $R_i^j \geq T_j$ in the data. Let such an event occur at time t_l^j at grid point j and t_m^k at grid point k, l and m are the time indices assigned to the events $l = 1,2.....,s_j, m = 1,2,....s_k$. where, s_j and s_k are total number of such events occurring at grid point j and k, respectively and within a minimal separation time $\pm\theta_{lm}^{jk}$ for each pair of events (l,m), where it is defined as

$$\theta_{lm}^{jk} = \min\{t_{l+1}^j - t_l^j, t_l^j - t_{l-1}^j, t_{m+1}^k - t_m^k, t_m^k - t_{m-1}^k\}/2 \tag{4.1}$$

θ_{lm}^{jk} is the minimum time between two succeeding rainfall events. We need to count the number of times an event occurs at j after it appears at k and vice versa. This is achieved by defining the quantities $c(j|k)$ and $c(k|j)$:

$$c(j|k) = \sum_{l=1}^{s_j}\sum_{m=1}^{s_k} J_{lm}^{jk} \tag{4.2}$$

with

$$J_{lm}^{jk} = \begin{cases} 1 & if \quad \tau'^{jk} < t_l^j - t_m^k < \tau^{jk} \\ 1/2 & if \quad t_l^j - t_m^k = \tau'^{jk} \\ 0 & else, \end{cases} \tag{4.3}$$

Where τ^{jk} is the upper limit of the range of time delays and τ'^{jk} is the lower limit. These time delay limits will be used to explicitly take into account the different time scales on which monsoonal precipitation can evolve(as in Sec. 5.3). If not explicitly mentioned then we have

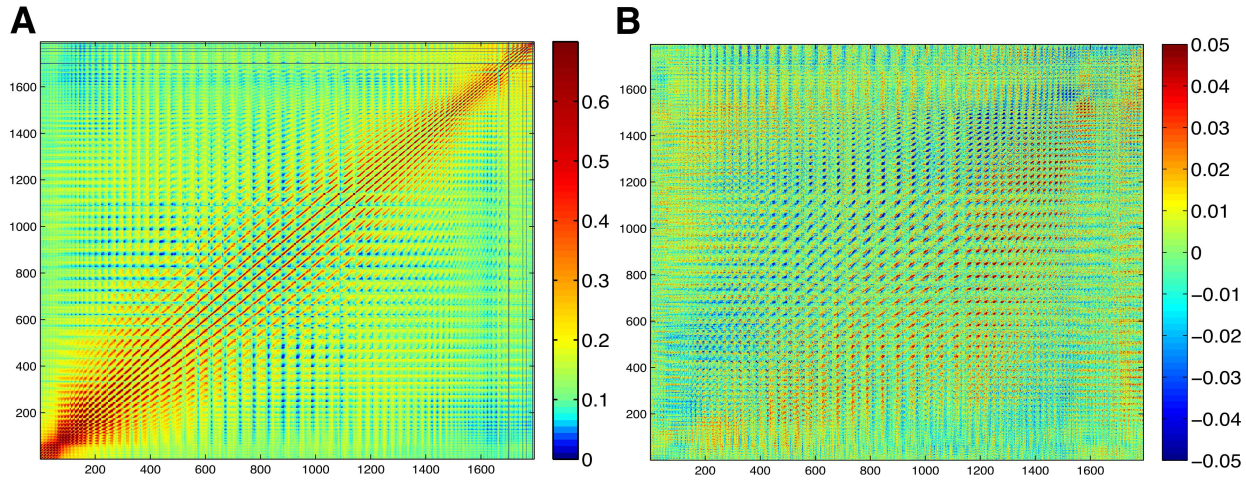

Figure 4.1 (A) It is the matrix Q, for $\alpha = 94\%$. Note the higher values along the diagonal, as the grid points along the diagonal are closer in space. (B) matrix q, for $\alpha = 94\%$. q is a anti symmetric matrix.

estimated the delays from minimum separation time, Eq. (4.1), i.e. $\tau^{jk} = \theta^{jk}_{lm}$ and $\tau'^{jk} = 0$ (Sec.4.4.1, 4.4.2 and Chapter. 5).

Similarly we can also define $c(k|j)$ and from these quantities we obtain

$$Q_{jk} = \frac{c(j|k) + c(k|j)}{\sqrt{(s_j s_k)}} \tag{4.4}$$

$$q_{jk} = \frac{c(j|k) - c(k|j)}{\sqrt{(s_j s_k)}} \tag{4.5}$$

Q_{jk} is the measure of event synchronization between grid points j and k. The q_{jk} measures the delay behaviour. These measures are normalised to $0 \leq Q_{jk} \leq 1$ and $-1 \leq q_{jk} \leq 1$. $Q_{jk} = 1$ implies complete synchronization and $q_{jk} = 1$ means that events at k precedes events in j. Along the diagonals of Q_{jk} we find higher values, as the elements along the diagonals are closer in space (Fig 4.1).

4.3.2. Clustering

To achieve the task of regionalisation, I have used the hierarchical clustering algorithm for forming clusters out of the matrix Q_{jk} [*Jain and Dubes (1988)*]. Hierarchical clustering can be considered as building a hierarchical tree (or dendrogram) where at each step I merge the leaf nodes sequentially based on a linkage criterion. Each of these leaf nodes is a cluster. All the clusters are fused together at the root of the hierarchical tree. In the sense of graph

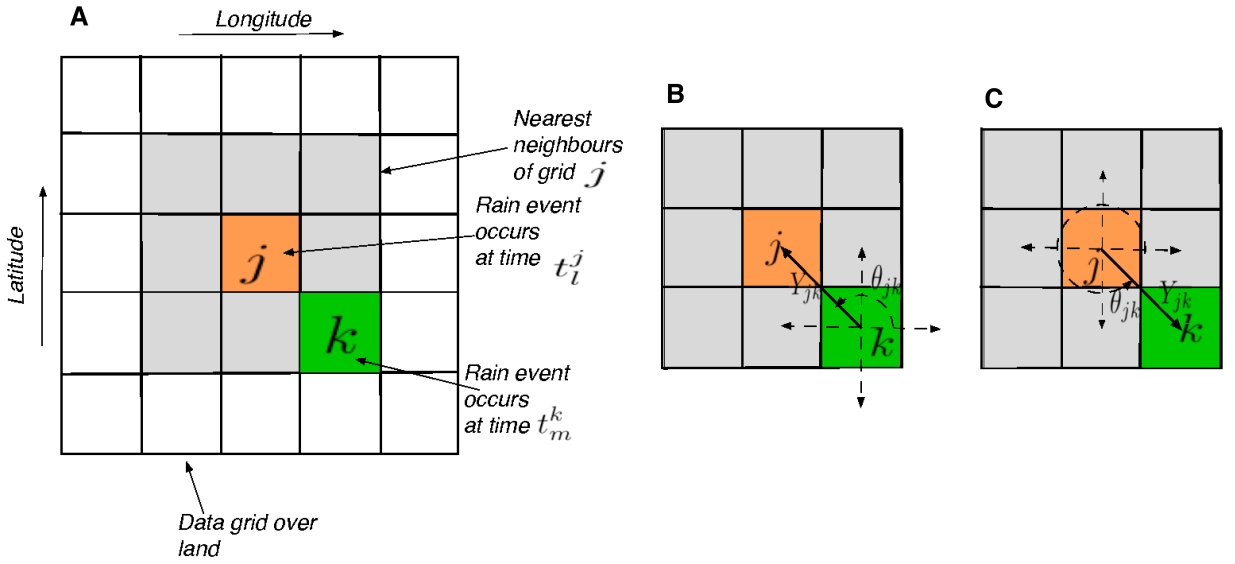

Figure 4.2 (A) Show the arrangement of data grid over the land, used for the estimation of Q_{jk} and q_{jk} , (B)-(C) Diagrammatic representation for calculating time delay patterns at a grid j.

theory a hierarchical tree is an acyclic connected graph where each node has zero or more children nodes and at most one parent node. As I have large data to cluster (around 1,800 grid points in APHRO–V0902 and 350 in IMD-D) I have used the "average linkage" as the linkage criterion for merging the nodes of the tree. In the average linkage method the distance L between two clusters, say U and V, is defined as

$$L = \frac{1}{|U| \cdot |V|} \sum_{j \in U} \sum_{k \in V} d(j,k) \tag{4.6}$$

Where j and k are the indices for the grid points, $|U|$ and $|V|$ stands for number of elements in set U and V respectively, and $d(j,k) = 1 - Q_{jk}$ is the dissimilarity matrix . The clusters which have the smallest value of L are merged together first. The hierarchical tree so formed is cut at the second highest value of the inconsistency coefficient. At the highest value all the leaf nodes of the hierarchal tree are merged together. The inconsistency coefficient is given by the relation

$$I_l = \frac{z_l - \mu_l}{\sigma_l} \tag{4.7}$$

where μ_l is the mean of the heights (height of a node is the length of the longest downward path to a leaf from that node) of all the links included in the calculation and σ_l is the standard deviation. z_l is the distance between the links joined at the level l (level of a node is the length of the path to its root). After obtaining the tree, I keep only those clusters for further analysis, which contain more than or equal to 3 grid points. The smaller clusters are discarded.

4.3.3. Time-delay patterns of events

Eq. 4.5 is used to extract the time delay patterns of events. Remember that $q_{jk} > 0$ means that the events in k precede events in j and for $q_{jk} < 0$ vice-versa. So using the sign q_{jk}, it is possible to determine the delay direction of events occurring at location j and k. The time delay patterns are determined as follows. In accordance with the convention that the negative q_{jk} means time delay is from j to k and vice-versa for positive q_{jk}, I try to calculate the average direction to the nearest neighbours. Say a vector $\mathbf{Y_{jk}} = (x_{jk}, y_{jk})$ gives the direction between two nearest neighbour grids (see Fig. 4.2) and $x_{jk} = \cos\theta_{jk}$ and $y_{jk} = \sin\theta_{jk}$, where θ_{jk} could have discrete values $\{0, \pi/4, \pi/2, 3\pi/4, 5\pi/4, 3\pi/2, 7\pi/4, 2\pi\}$. Then the average direction will be

$$\mathbf{Y'_{jk}} = \frac{\sum_{\substack{nearest\\ neighbours\\ of\, j}} q_{jk} \mathbf{Y_{jk}}}{\sum_{\substack{nearest\\ neighbours\\ of\, j}} q_{jk}} \tag{4.8}$$

In Sec. 4.4.2 I will further discuss the directionality involved in the moisture transport during the active phase of ISM at the each grid point, and compare it with wind data from NCEP/NCAR reanalysis.

4.4 Results and Discussion

In the following sections I would discuss the results of the above procedures on the two different data sets used for this study.

4.4.1. Spatially coherent zones of heavy rain events

For calculating the ES I have used two different thresholds of $\alpha = 94\%$ percentile and $\alpha = 90\%$ percentile. These thresholds could be interpreted as very heavy to heavy rainfall events. My motivation for choosing these thresholds is based on the fact that such rainfall events at a given location can only occur when monsoon is in its active phase over a large spatial region containing this particular location. Such events can be scattered over many locations within a large spatial region occurring either at the same time or with some delay. To get these regions where monsoonal activity is more coherent compared to any other region, I applied a clustering analysis to the ES measure of this coherent activity. The results of clustering are shown in Fig. 4.3 where clusters containing at least 3 or more grid points are shown in colour and the remaining grid points are marked in black. I have marked some of the larger clusters by numbers for easy reference. A summary of the analysis is given in Tab. 5.1. I do not expect to generate exactly equal structures for both data sets used in the analysis, as spatial resolution, spatial extent and time period are different (see Sec. 4.2]. Nevertheless, I find some striking similarities in the structures in all the (Fig. 4.3(A)-(D)). In all cases studied in clustering analysis I see a cluster forming along the strong orographic barrier of Western Ghats on the west coast of India (cluster number 4 in Fig. 4.3(A)-(D)) pointing the important role played by orographic barriers in generation of rainfall. Interestingly, the whole of north east (cluster number 10 in Fig 4.3 (A)-(D)) emerge as a single cluster even though the region comprises of a very variable topography, including many high mountain ranges. I also

Data Used	Threshold α percentile	I_l	Fig no.	Total Clusters	Clusters containing greater than 3 grid points
APHRO–V0902	94%	1.547	4.3(A)	67	38
APHRO–V0902	90%	1.547	4.3(B)	45	27
IMD–D	94%	1.538	4.3(C)	14	10
IMD–D	90%	1.538	4.3(D)	19	14

Table 4.1 Summary of results of Cluster analysis

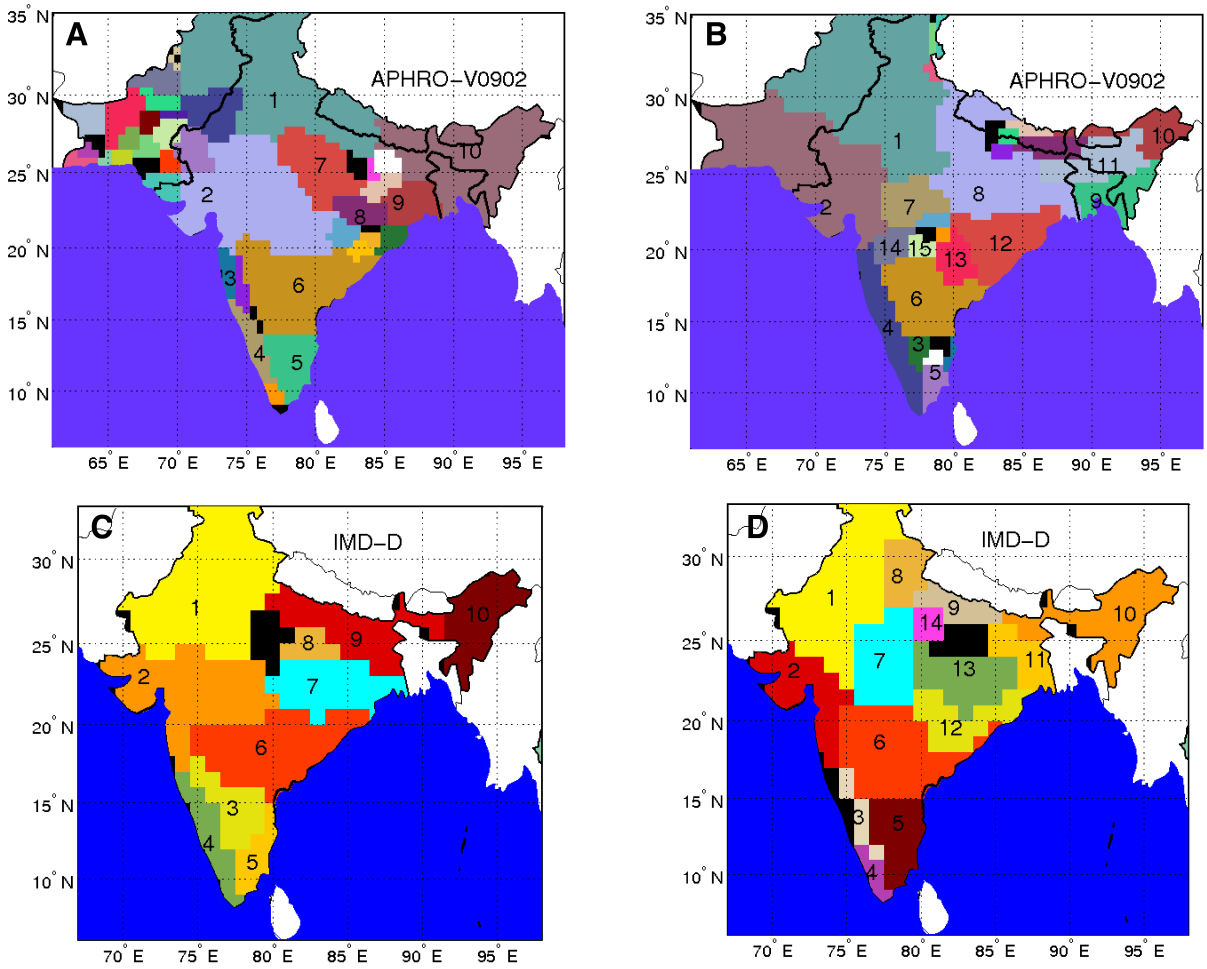

Figure 4.3 Different colours here gives the different clusters obtained using Hierarchal clustering algorithm using matrix for Q as dissimilarity matrix. (A) Data used is APHRO–V0902 , threshold $\alpha = 94\%$, (B) Data used is APHRO–V0902 , threshold $\alpha = 90\%$,(C) Data used is IMD-D, threshold $\alpha = 94\%$,(D) Data used is IMD-D , threshold $\alpha = 90\%$

observe many other similarities in the results like formation of separate clusters in south east peninsular India (cluster number 5 in Fig. 4.3 (A)-(D)).

The basic motivation for the above analysis was to identify regions where monsoonal rainfall dynamics is more coherent or connected than others. So, I have been able to identify the following major dynamical regions of monsoonal activity over the Indian subcontinent, these are: i) along the west coast (cluster number 4), ii) north east of India and parts of Bangladesh (cluster number 10), iii) north west of the subcontinent (cluster number 1), iv) western and adjoining parts of central India (cluster number 2), v) south east peninsular India (cluster number 5). Some of the clusters are large in size covering thousands of kilometres. A possible mechanism for the existence of such long range spatial correlations could be due to clustering of synoptic activity during the active phase of monsoon [*Goswami et al. (2003)*]. This type of monsoonal rainfall activity seems to be a result of large spatial scale atmospheric

activity, like the formation and clustering of large low pressure systems. In one monsoon season there could be $3-4$ active or break periods and slowly evolving between each other [*Lawrence and Webster (2002)*]. Very heavy rain events are associated with the active phase of monsoon. From the presented method I could obtain the regions where monsoon gets active simultaneously or within some delay. During the active phase of monsoon, large parts of the country experience heavy rainfall (above normal rainfall). But some parts like the south east peninsular India are out of phase with it [*Krishnmurthy and Shukla (2000)*, *Singh et al. (1992)*], and receive rain during the break phase. I also see in the above clusters that south west peninsular India is always a small separate cluster. Different rain producing monsoonal systems can evolve at many different time scales, in Sec. 4.4.3 I make an explicit distinction between different modes of rainfall activity present in monsoon and estimate coherent zones based on time scales involved in monsoonal precipitation.

Coherent rainfall zones over India have been previously recognised by *Gadgil et al. (1993)*. The data used in their work was monthly rainfall from 459 stations spread over India for the period 1907-1981. The delineation of the rainfall zones was based on the variation of monthly rainfall. They have identified 31 zones (limited to the political boundary of India), those are smaller in size than the ones identified in presented work. In present work I had maximum of 14 to 19 clusters when I considered data only for India (see Fig 4.3 (C) and (D) and Tab. 5.1). Most of the basic structures I have found and mention above were also present in *Gadgil et al. (1993)* but few of them defragmented into smaller zones. This difference seems to emerge due to two main reasons – (1) Dissimilarities in the spatial and temporal resolution and spatial extent of the data analysed in the present study. I used spatially interpolated daily rainfall data compared to the monthly rainfall station data in *Gadgil et al. (1993)*. (2) In the presented study I have aimed at understanding the dynamical coherence in the daily rainfall activity during monsoon season. For measuring the dynamical coherence I have used a nonlinear measure of correlations rather than coherence in variation of monthly rainfall as was done in *Gadgil et al. (1993)*.

4.4.2. Time delay directions

The method of ES as defined before, allows for obtaining time delay patterns of events. The time delay patterns could be further used in studying the paths of rain events and their moisture sources or paths of synoptic scale atmospheric disturbances and their evolution. This is significant in the light that Indian summer monsoon has two major branches or sources of moisture. The first one is the Bay of Bengal in the east and the second one is Arabian sea in the west. In Fig. 4.4 I show the time delay patterns obtained employing the approach discussed in Sec.4.3.3. I have observed similar patterns for both data sets. I have also compared these results with wind directions as calculated from the reanalysis data at 850hPa height for the same threshold Fig. 4.4 (C). To obtain wind directions I have taken the mean of the wind directions over the days when there were events above an α percentile in the precipitation data set of NCEP/NCAR reanalysis data sets, i.e. the time indices are also derived from the

same NCEP/NCAR reanalysis data set. The directions estimated by the proposed method are almost identical to the wind directions (Fig. 4.4) derived from reanalysis data. I have obtained similar results for the other thresholds too. The similarity between time delay patterns and wind directions indicate that the above method allows to estimate the paths of moisture movement over the land mass or the directions involved in atmospheric circulations during moisture transport in the monsoon season. It is interesting to note that the moisture source of heavy rainfall events over central Indian region is the Bay of Bengal. It's known that rainfall over the central part of India is received by north ward movement of low pressure areas from the Bay of Bengal [*Lal et al. (1995)*]. Therefore I find a strong directionality of arrows from the Bay of Bengal to north over central India (Fig.4.4).

4.4.3. Separating time scales

The Monsoon's inter seasonal oscillations (ISO) are another intriguing feature of monsoon dynamics. These oscillations are part of the internal dynamics of monsoon and also govern its inter seasonal variability (ISV). The ISO is composed of a hierarchy of quasi-periods like 3 to 7 days, 10 to 20 days and 30 to 60 days [*Waliser (2006)*, *Ding and Sikka (2006)*]. The 3 to 7 days mode is caused by the oscillation of monsoon trough over the Indo-Gangetic plains. Both 10-20 days [*Krishnamurti and Ardunay (1980)*] (quasi-biweekly oscillation (QBM)) and 30-60 days [*Sikka and Gadgil (1980)*] (Madden-Julian oscillations (MJO)) are associated with planetary scale waves.

As a first exercise I try to identify the important time delays in the data at a particular threshold on the rain events. For this purpose I use the IMD-D data set and calculate the frequency distribution of the delays $p(n_i{}^d) = p(|t_l^j - t_m^k|)$ for a given distance d between the grid points j and k on the lattice where $n = |l - m|$, $n = 1,2...122$ (days) and i is an index

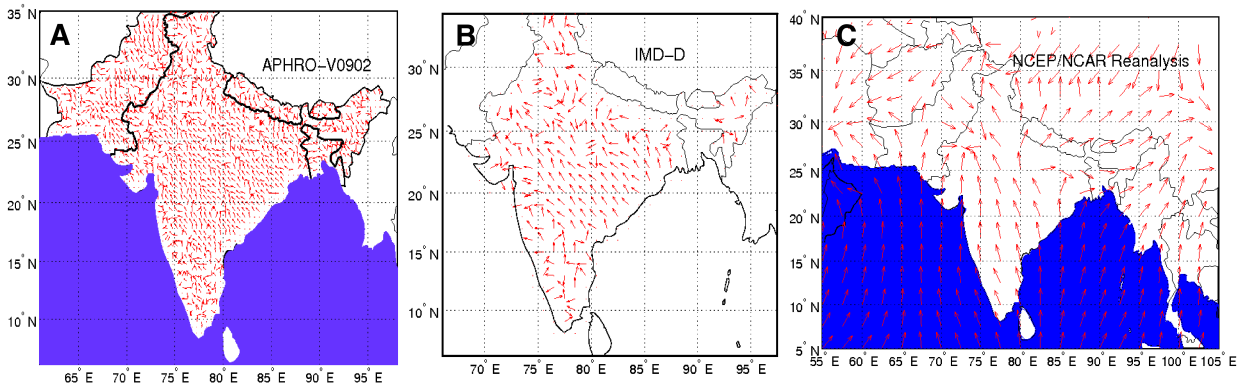

Figure 4.4 (A) Time delay patterns as obtained form APHRO–V0902 for $\alpha = 94\%$ (B) Time delay patterns as obtained form IMD-D for $\alpha = 94\%$ (C) Mean wind direction at the height of 850 hPa during rain events above the threshold of $\alpha = 94\%$, derived from NCEP/NCAR data set with time indices for the events calculated from NCEP/NCAR precipitation data

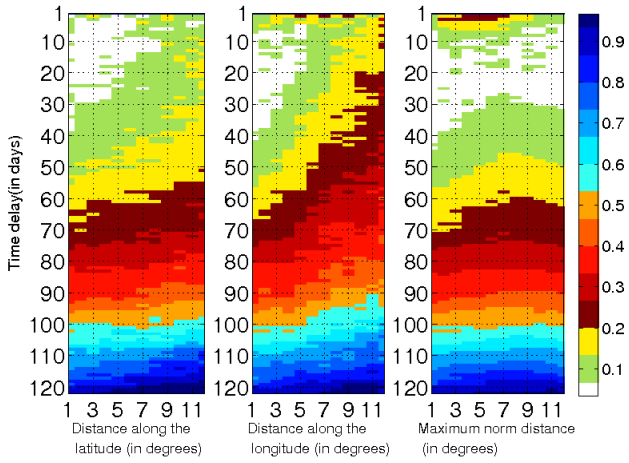

Figure 4.5 Shannon entropy for the distribution of each delay for all comparisons between grid with a given distance. The data used to generate above image is IMD-D.

for a specific comparison between the grid points j and k satisfying the distance criterion $(maxnorm(j,k) = d)$. To calculate the relative significance (not in a strict statistical sense) of delays, I calculate the Shannon's entropy for each delay at $S_n^d = \sum_{i=1}^{r} p(n_i^d) \ln p(n_i^d)$. Clearly S_n^d is a measure of how uniform the distribution is for a particular delay over all r comparisons. The more uniform the particular delay the more it has relative importance in the system. Further we normalise S_n^d by, $S = \frac{|S_n^d - S_n^{d\,max}|}{S_n^{d\,max}}$ where $S_n^{d\,max}$ is the maximum of S_n^d over all the delays. In Fig 4.5 the lowest values (white colour) of S gives the most significant delays. I observe that the most important delay lies between 10 to 40 days with a maximum spatial extent for 10 to 20 days. From this I conclude that the 10 to 20 days mode is the most influential one and has largest spatial scales.

For the purpose of study on the spatial organisation of the rainfall field and the formation of coherent zones, I use four different delay ranges: 3 to 7 days, 10 to 20 days (QBM), 30 to 60 days (MJO), and 0 to 7 days to account for the fluctuations in synoptic systems such as lows and depressions . Applying these time slices in Eq. 4.3 I calculate the ES matrix Q and perform a clustering analysis on it. This allows me to determine the spatial scales under explicit consideration of the aforementioned time scales.

In Fig. 4.6 (A and B) I consider $\tau'^{jk} = 10$ and $\tau^{jk} = 20$, corresponding to the QBM. In this case I get geographically discontinuous clusters. I find a strong coherence between the west coast of peninsular India, parts of central India and the Northeast of India (see green colour along the west coast and north east of India in Fig. 4.6(A) and yellow colour in Fig. 4.6(B) marked as 1a, 1b and 1c). The black colour means the formation of very small clusters or defragmentation. Another important aspect to note is the formation of a huge cluster formed over parts of central India, north India and Pakistan Fig. 4.6(A). These findings suggest that two zones of rainfall activity exist on these times scales, one along the west coast, parts of central India and north east and another one consists of the Indo-Gangetic plains. The

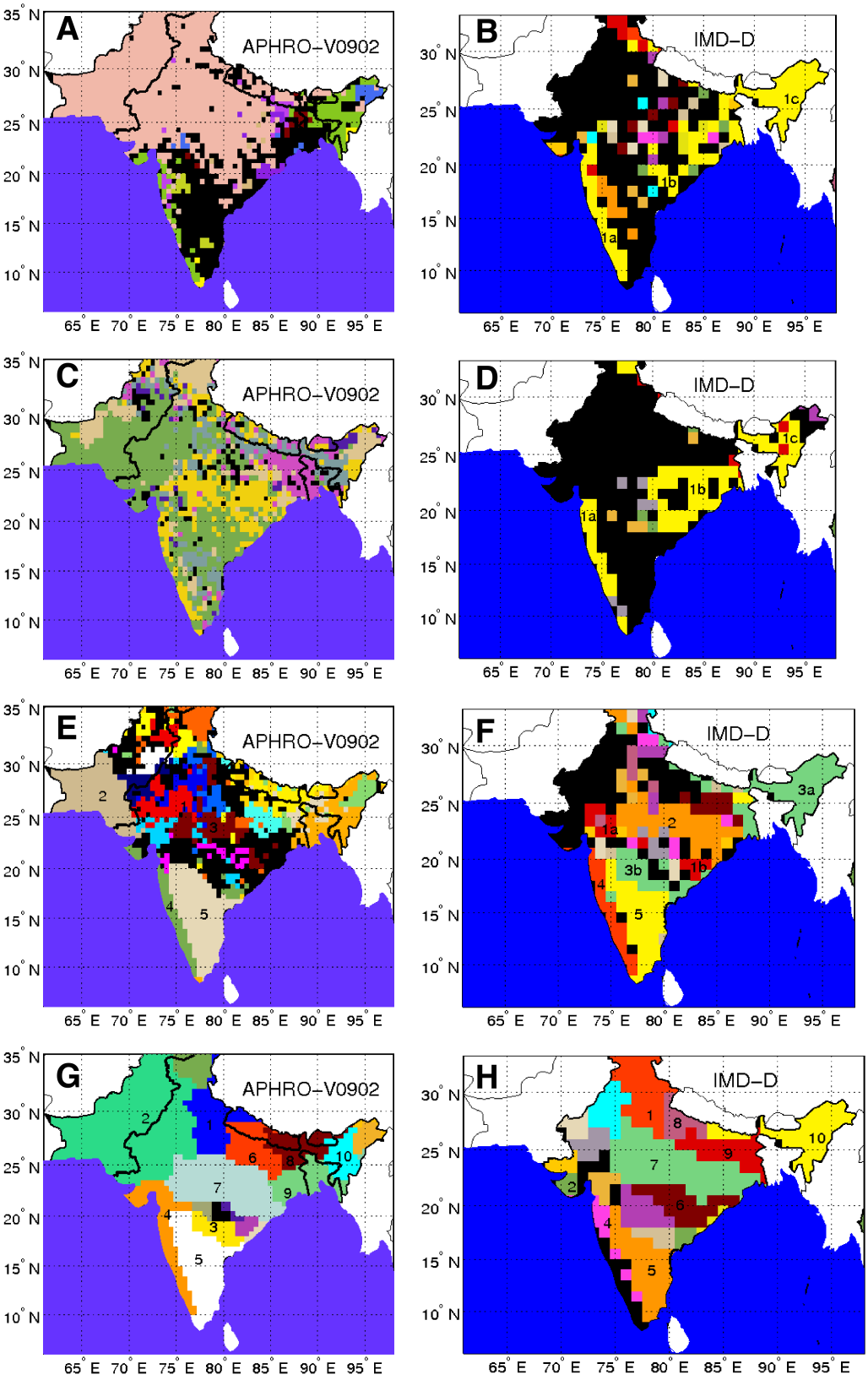

Figure 4.6 Results of clustering at different delays. (A-B) Time delay 10–20 days, (C-D) Time delay 30–60 days, (E-F) Time delay 3–7 days, (G-H) Time delay 0–7 days. Data sets used are indicated in the figure.

detected spatial scales are now rather large, as also expected because this particular mode is associated with the planetary scale phenomena of westward moving waves originating in the Pacific.

For the case of the MJO scale I take $\tau'^{jk} = 30$ and $\tau^{jk} = 60$. Again I find a formation of geographically discontinuous clusters, with largest coherence emerging between the west coast of peninsular India, parts of central India, the Northeast of India and the foothills of the Himalayas (see yellow colour along the west coast, along Himalayas, central India and parts of north east of India in Fig. 4.6(C) and yellow colour in Fig. 4.6(D) marked as 1a,1b and 1c). The high coherence in rain intensities over the west coast and central India on MJO time scales has also been reported by *Singh et al. (1992)*. However the coherence observed between the Himalayan foothills and the north east has not yet been known. One striking dissimilarity between the results for the two data set is the formation of a single cluster covering the east peninsular India and Indo-Gangetic plains (green colour in Fig 4.6(C)) in APHRO-V090 as compared to IMD-D Fig. 4.6(D) where the region is defragmented into very small clusters. The clustering algorithm is not able to fully resolve the structures in this particular case of the IMD-D data set.

In the next case I consider the time scale $\tau'^{jk} = 3$ and $\tau^{jk} = 7$. Now I do not find too much geographically discontinuity in clusters (Fig. 4.6 E - F). Unlike larger time scales, there does not exist very long range correlations in this case. I observe clear and strong clusters along both east and west peninsular India, the former being the larger one (Fig. 4.6 E - F) (clusters 4 and 5)). I find a strong separate clustering over central India and north east too. Finally, I take a temporal thresholds of $\tau'^{jk} = 0$ and $\tau^{jk} = 7$, putting together all the synoptic scale activities and fluctuations into one single time scale. Obviously, very long range correlations and geographically discontinuous spatial correlations as observed above do not exist, I also have similar observation for these time scales in the Fig 4.5. The last two time scales are indicative for spatially local variations in monsoonal rainfall due to synoptic scale activities like monsoonal troughs, depressions and lows. Even with all the dissimilarities of IMD-D and APPHRO-V090 I see striking similarities in the clusters in Fig. 4.6(G -H). The clustering

Data Used	Threshold α percentile	time delay range	I_l	Fig no.	Total Clusters	Clusters plotted
IMD–D	94%	$10-20$ days	1.1496	4.6(B)	119	14
APPHRO-V0902	94%	$10-20$ days	1.1546	4.6(A)	239	10
IMD–D	94%	$30-60$ days	1.1467	4.6(D)	159	6
APPHRO-V0902	94%	$30-60$ days	1.1546	4.6(C)	28	6
IMD–D	94%	$3-7$ days	1.1545	4.6(F)	69	15
APPHRO-V0902	94%	$3-7$ days	1.1546	4.6(E)	213	19
IMD–D	94%	$0-7$ days	1.148	4.6(H)	29	18
APPHRO-V0902	94%	$0-7$ days	1.15469	4.6(G)	32	19

Table 4.2 Summary of results of the Cluster analysis

analysis is summarised in Tab. 4.2.

From the above results I can conclude that during monsoon the rainfall field exhibits an obvious spatial organisation. This study provides a qualitative view of this spatial organisation. It has a strong dependence on the time scales of the different rain producing processes and systems. At QBM and MJO time scales, I have long range spatial correlations transacting even couples of 1000 km, where at shorter time scales of 3 to 7 days and 0 to 7 days the spatial scales are smaller. The interaction of topography and monsoonal dynamics is also apparent in the shape of the clusters.

4.4.4. Remarks

By using event synchronization I have estimated different dynamical zones of monsoon and how they reorganise themselves on different time scales. Above results strongly point towards the existence of spatial organisation in the rainfall field due to the interaction of physico-geographical factors with the dynamics of the monsoon. The spatial correlations studied above could only emerge if there is a recurring tendency of heavy rain events to occur over the same zone in lag synchronised form. This in turn could lead to recurring floods over the same zone. An example for this case is the north east (cluster number 10 in Fig. 4.3) of the subcontinent and parts of Bangladesh, which suffers from regular floods during monsoons. In accordance with the observation these regions do emerge as separate clusters in presented study and even at different time scales of monsoonal evolution. Hence, such spatial organisation also opens the possibility for predictions of a probable spatial extent of monsoonal rainfall activity within a certain time delay, if some sporadic heavy rain events are reported in a part of these coherent rainfall zones. Another important application could also be in palaeoclimate studies where data are usually collected from few sparse locations. Using above classification of regions one could relate a location with larger dynamical region of past monsoonal rainfall. Furthermore, one could spatially extrapolate such data to the dynamical region it belongs to. In the light of the above facts, a further study of the spatial organisation of monsoonal rainfall is necessary to both understand climatological causes and its application in prediction of extreme monsoonal events like floods and droughts. I will attempt such a study in the next chapter. organisation

4.5 Summary

Here I have presented a method based on event synchronization of rain events to identify regions where the dynamics of monsoonal rainfall is more coherent or homogenous. I have applied this method on two separate data sets of different spatial resolution and spatial extent and also different time periods. The regions obtained in both data sets are strikingly similar emphasising the underlying structure in the physical processes responsible for the generation of rainfall activity, whether it is the atmospheric circulation itself or the topography of the region. Furthermore, I have developed a method to construct patterns of time delay of rainfall

events. I showed that these time delay patterns seem to follow the prevalent wind directions during the period. Hence, we can establish the path ways of moisture movement over land during monsoon. At distinct time scales, I observe the emergence of spatial correlations on very different spatial scales. At the time scales of QBM and MJO, which are related to the planetary scale waves, I have seen very long range spatial correlations crossing over even the geographically discontinuous regions. Above results also seem to be consistent with the observation of recurring extreme events (floods) over these regions. I suggest that the possibility of employing such approach in the prediction of extreme events at different time scales should be explored.

Chapter **5**

Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks

A system is a network of interdependent components that work together to try to accomplish the aim of the system.

– William Edwards Deming*

5.1 Introduction

Indian Summer monsoon (ISM) rainfall over South Asia is the result of the interaction of several complex atmospheric processes evolving at many different spatial and temporal scales (e.g., *Webster (1987)*). Apart from the influences of the interplay of synoptic scale weather phenomena, the ISM rainfall patterns are also modulated by the steep topography of the Himalayas (e.g., *Bookhagen and Burbank (2010)*). Hence monsoonal rainfall has highly intricate spatiotemporal patterns. Here, I will analyse these spatiotemporal ISM rainfall patterns over South Asia employing nonlinear correlation measures called event synchronization and complex networks.

The methodology of complex networks has emerged as an important mathematical tool in the analysis of complex systems in the last decade and has been applied to a wide variety of disciplines within the natural and social sciences (e.g., *Watts and Strogatz (1998)*, *Newman (2003)* and *Albert and Barabási (2002)*). The spatiotemporal structure of complex networks, dynamics on them, and phenomena such as network synchronization have been of large interest to the nonlinear-dynamics but also to the climate communities (e.g., *Boccaletti et al. (2006)*, *Arenas et al. (2008)*). In recent years, the tools of complex network theory has also found an application in the data driven analysis of the global climate system (e.g., *Donges et al. (2009)*, *Tsonis et al. (2006)*,*Tsonis and Swanson (2008)*, *Yamasaki et al. (2008)* and *Donges et al. (2011)*).

* He was an American statistician and author.

The correlation structure of climate variables and their teleconnections have been studied employing this mathematical tool. Other applications of this approach in climate data analysis include the identification of community structures in the climate system [*Tsonis et al. (2010)*] and the linkages of different regional climate phenomena [*Steinhaeuser et al. (2010)*], leading to the discovery of a new dynamical mechanism for major climate shifts [*Swanson and Tsonis (2009)*]. A similar approach based on a shared-nearest neighbour method has been used to discover climate indices from sea surface temperature data [*Steinbach et al. (2003)*]. Here, I apply a complex networks approach to a specific regional climatic phenomenon of the ISM and study the spatiotemporal pattern of rainfall.

Although rainfall has a significant impact on society, agriculture and fresh-water generation in this region [*Bookhagen and Burbank (2010)*], it is not easy to decipher its dynamics due to its spatiotemporal complexity and involved small-scale processes. Furthermore, rainfall is a point process with large spatial and temporal discontinuities ranging from very weak to strong events within small temporal and spatial scales [*Wulf et al. (2010)*]. Nonlinear correlations called event synchronization can overcome these difficulties (e.g., *Quiroga et al. (2002)*). The methodologies of complex networks and event synchronization take into account the nonlinearities existing in the correlation structure of the rainfall field. For the analysis presented here, I have considered the rainfall events at the 90th and 94th percentiles and thus focused on the extreme rainfall events and the associated atmospheric processes. By further developing the methodology and building upon previous investigations [*Malik et al. (2010)*], I also attempt to study the evolution of monsoonal rainfall pattern over the last decades. The key advantage of applying complex network theory is that it does not require details of several of climatic variables and indices one may have to analyse to study spatiotemporal rainfall patterns.

This study provides several critical insights into the underlying atmospheric processes responsible for the evolution of the ISM and related extreme rainfall events. A description of climatic setting of Indian Summer Monsoon (ISM) is already provided in detail in Chapter 2 of the this thesis. Therefore, I will next introduce the data, methods and then results.

5.2 Data

In this study, I have used daily gridded rainfall data from 1951 to 2007. A detailed description of this data set is given in Appendix A Sec. A.1.1. I have extracted the data for the South Asian region (Fig. 5.1) with a horizontal resolution of 0.5 degree ($\sim$ 55km). I will refer to this data set as APHRO–V01003R1. I have also employed the zonal (u) and meridional (v) wind components and surface rainfall data from the NCEP/NCAR reanalysis data set from 1951 to 2004 with 2.5 degree resolution (See Appendix A Sec. A.2 for detailed description).

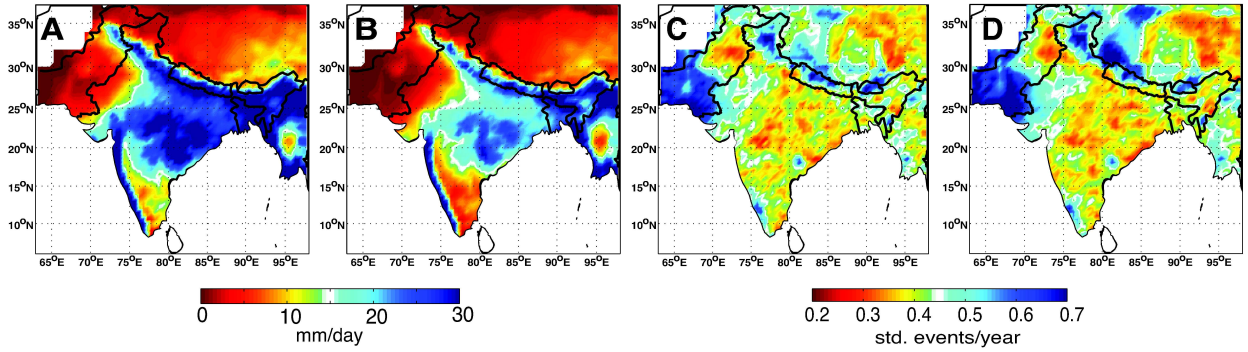

Figure 5.1 Rainfall thresholds and their annual variability calculated from 57 years of data (1951-2007). I show the annual daily rainfall amount for thresholds of $\alpha = 94\%$ (A) and $\alpha = 90\%$ (B). Note the generally high daily rainfall amounts in the Ganges plain and at the orographic barriers of the eastern Swats and the southern Himalayas. I show the standard deviation from the mean number of rainfall events per year for $\alpha = 94\%$ (C) and $\alpha = 90\%$ (D). Areas in blue indicate high inter-annual variability and are generally spatially disconnected from the mean annual rainfall pattern.

5.3 Methodology

In this section I will first introduce the nonlinear correlation measure of event synchronization; second I will provide a description of network construction methodology and third, I list the details of network measures and terminology used in this work.

5.3.1. Event synchronization (ES)

I have introduced event synchronization (ES) in Sec. 4.3.1, as a nonlinear correlation to measure the strength of synchronization of rain events between two different grid points and their delay behaviour. ES has been previously introduced in *Quiroga et al. (2002)* and modified by *Malik et al. (2010)*. Here too, ES serves as the basis for constructing the complex networks. Only those rain events and their time indices are considered, which are above a certain α percentile of all the wet days during the 4 summer monsoon months of JJAS (June, July, August and September). An α percentile threshold for each grid point gives a unique net amount of rainfall per day as the threshold for each grid point. For $\alpha = 94\%$ and $\alpha = 90\%$ the thresholds on net rainfall amount per day is shown in Fig. 5.1 (A) and (B) respectively. Also, the annual standard deviation of the number of events per year is plotted in the Fig. 5.1 (C) and (D) for both thresholds. Thresholds used in this study are considered to be extreme rainfall events [*Groisman et al. (1999)*, *Kripalani and Kulkarni (1999)*, *Goswami et al. (2006)*]. I have chosen these thresholds, because rainfall at or above them only occurs during the active phase of the ISM. Hence these thresholds are extremely useful in studying the spatial structures of underlying atmospheric processes which are responsible for the active phase of monsoon. Events with the above mentioned thresholds are less effected by the sampling uncertainty, as compared to very high thresholds and hence provide more robustness to calculation of ES. This is one of the approaches for characterising extreme

rainfall events, an alternative approach is to describe extreme daily rainfall events by suitable statistical models of extreme value theory [*Coles (2001)*]. A detailed analysis between two approaches in context of the ISM can be found in for e.g., *May (2004a)* and *W.May (2004)*. The mathematical description for the ES is provided in Sec. 4.3.1. All throughout this chapter I have used delays defined using minimum separation time i.e., Eq. 4.1.

So, finally I get two matrices: Q_{ij} which is the measure of the *strength of event synchronization* between grid points i and j. It is normalised to $0 \leq Q \leq 1$ implying $Q = 1$ for complete synchronization. The second matrix q_{ij} measures the *delay behaviour* and $-1 \leq q \leq 1$. And $q_{ij} = 1$ implies that an event at i always precedes an event at j. The matrix q_{ij} open up several unique possibilities to analyse delay directions. Q_{ij} is a square symmetric matrix and q_{ij} is square anti-symmetric matrix. ES has been very specifically designed to calculate nonlinear correlations among time series with events. And better suited than other correlation measure such as cross-correlation (linear) and mutual information (nonlinear) for measuring correlations among bivariate time series with events defined on them. There exist some other measures for estimating dependencies among events too. A distinctively different but similarly analytical rigourous method to measure dependencies between extremes based on extreme value theory is provided in *Coles et al. (1999)*. In this approach, it is not required to pre define events like in ES but it has no option of analysing delays and their directions.

5.3.2. Constructing adjacency matrices

To construct the network out of the ES matrices described above, I treat a grid point over land as the vertex of the network and any edge between them will be referred to as a link. In other words, vertex and grid point have the same meaning in the following discussion. The links between different grid points exist if the strength of the ES is above a certain predefined threshold. Directionality to the links is introduced using the delay direction from the sign of q. Let us say a grid point is connected to k other different grid points. Then I call k the *degree* of the grid point where k is an integer between 0 and $N-1$, where N is the total number of grid points. For the purpose of constructing the network I have used a fixed global link density K. In the case of undirected networks it is related to the probability $P(k)$ – the number of grid points having k connections as $\frac{1}{N-1}\sum_{k=1}^{k=k_{max}} P(k)k = K$, where k_{max} is the maximum of degree. I will only analyse the minimalist correlation structure of the ISM rainfall. In order to obtain it, I have taken the value of $K = 0.05$, i.e. I assume that only 5% of the total grid points are connected. These links represent the 5% strongest correlations. The underlying assumption is that the extracted minimalist correlation structure contains the statistically most significant and essential features of correlations in the ISM rainfall field. Thereby, I analyse only the statistically most significant correlations and, in turn, I will be able to remove much of the redundant information. From a meteorological perspective, I thus analyse the most persistent atmospheric features during an ISM season responsible for generation of extreme rainfall events. I calculate a threshold θ_{ij}^{Q} on Q_{ij} by setting $K = 0.05$. Next we, then convert Q_{ij} into

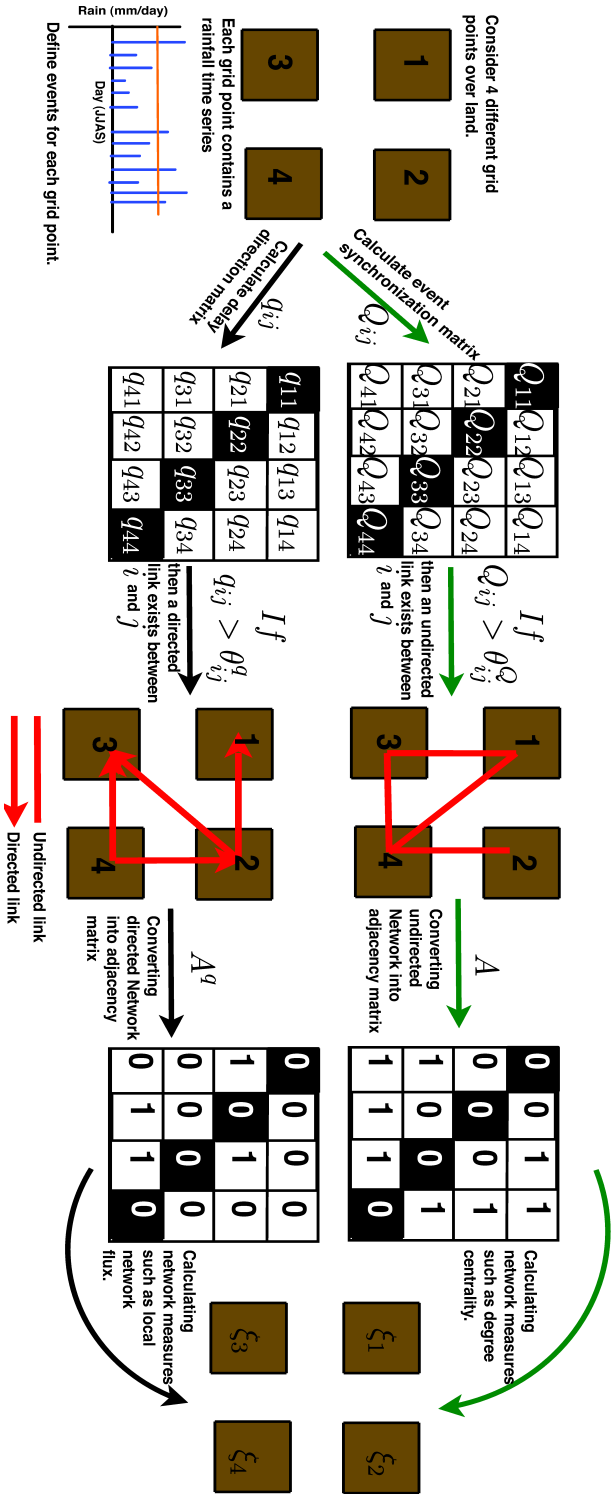

Figure 5.2 Schematic flow diagram showing the steps involved in the construction of the network (read from top to bottom) and in calculating some network measure ξ. First, I define events for a rainfall time series at the each grid point and then calculate event synchronization matrix Q_{ij} and delay direction matrix q_{ij}. Depending on matrix characteristics, I establish undirected or directed links and convert the network to an adjacency matrix. And from the adjacency matrix I estimate the network measure ξ.

a binary matrix called the *adjacency matrix A* where,

$$A_{ij} = \begin{cases} 1 & if \quad Q_{ij} > \theta_{ij}^{Q} \\ 0 & else, \end{cases} \tag{5.1}$$

and A is a symmetric matrix such as Q. Similarly, q_{ij} can also be constructed into an adjacency matrix A^q with a difference that now

$$A_{ij}^{q} = \begin{cases} 1 & if \quad q_{ij} > |\theta_{ij}^{q}| \\ 0 & else, \end{cases} \tag{5.2}$$

We now have sense of direction in A^q, i.e. $A_{ij}^{q} = 1$ means that the link is from i to j and not vice-versa. Therefore A^q is not a symmetric matrix. A further adjacency matrix A^{Qq} can be constructed as,

$$A_{ij}^{Qq} = A_{ij}^{q} A_{ij} \tag{5.3}$$

and it has the characteristic of having both, the information of direction and strength of the link. Again A^{Qq} will not be a symmetric matrix. In Fig. 5.2 I have provided a schematic representation of the steps involved in the construction of a network, starting from rainfall time series at each grid point.

5.3.3. Complex network measures

I use several basic measures from complex network theory to characterise the rainfall network constructed from the ES matrices [*Boccaletti et al. (2006)*] (c.f Fig. 5.2] The simplest measure is the *degree centrality* of grid point j, C_{Dj}, which is given as

$$C_{Dj} = \frac{\sum_{i=1}^{N} A_{ij}}{N-1} \tag{5.4}$$

where N as above is the total number of grid points. C_{Dj} measures the number of grid points linked to a particular grid point j. A grid point having higher degree centrality is expected to have higher influence on the functioning of the network. In the presented study grid points where critical atmospheric processes responsible for the development of ISM rainfall take place should show higher degree centrality. The *local clustering coefficient* $\mathscr{C}_j$ for the grid point j tells us the probability if two different connected grid points that are also connected to the same third grid point. In the language of graph theory this relates to how close the neighbours of a vertex are to a complete graph (also called clique). Let us say that grid point j has k_j links, i.e, its is connected it with k_j other nodes. If each of these grid points are also connected to each other then I need $\frac{k_j(k_j-1)}{2}$ links. If the actual number of links existing is $\mathscr{E}_j$

then

$$\mathscr{C}_j = \frac{2\mathscr{E}_j}{k_j(k_j-1)} \tag{5.5}$$

In the presented approach, I will use $\mathscr{C}_j$ to estimate the spatial continuity of rainfall fields. An additional sophisticated centrality measure is the *closeness centrality* C_{Cj}. The mathematical definition used here measures the network vulnerability. Grid points with a high value C_{Cj} are very critical for the functioning of the network [*Dangalchev (2006)*]. Let us say that the mean *geodesic distance*, i.e. the shortest path between a grid point j to all the other grid points i connected to it is $d(j,i)$ then closeness centrality is

$$C_{Cj} = \sum_{i \in \{V \setminus j\}} 2^{-d(j,i)} \tag{5.6}$$

where $\{V \setminus j\}$ is the set of all the vertices or grid points excluding j. The physical interpretation of this measure in the present context is that it gives the speed of information propagation. For example, any perturbation in the system travels fastest to the vertices with highest vales of C_{Cj}. I have normalized C_{Cj} between 0 and 1 by dividing it with the maximum of C_{Cj}.

Betweenness centrality C_{Bj} is the sum of the ratio of the number of shortest paths between two vertices passing through a particular grid point to the total number of shortest paths between those two vertices. Mathematically it is given by

$$C_{Bj} = \sum_{j \neq s \neq t \in V} \frac{\sigma_j(l,m)}{\sigma(l,m)} \tag{5.7}$$

where $\sigma_j(l,m)$ is the number of shortest paths between l and m passing through j. Physically C_{Bj} indicate the information pathways if we assume that the information travels using the shortest path. In the presented case, I hypothesise that rain events (or in the wider sense water vapour) are the quantity travelling through the network. This hypothesis is subject to further research in order to provide robust and solid physical interpretations.

I can also measure the length of the links in physical units by defining the *geographical distance*. If two grid points i and j are connected then the length of this link L_{ij} can be calculated using the formula for spherical earth projected on to plane, i.e.

$$L_{ij} = R\sqrt{(\delta\phi_{ij})^2 + (\cos(\phi_m)\delta\lambda_{ij})^2} \tag{5.8}$$

where $\delta\phi_{ij}$ and $\delta\lambda_{ij}$ are differences in latitude and longitude in radians between grid point i and j, ϕ_m is the mean of the latitudes of i and j and R is the radius of the earth.

5.3.4. Directed networks : local network flux

Directed networks are the networks where every link has a sense of direction, i.e., either it is outgoing or incoming. The indegree k^{in} is the number of links incident on a grid point; -

the outdegree k^{out} is the number of links leaving a grid point. I define the local network flux as the difference of the two, i.e. $\Delta k = k^{in} - k^{out}$. A strong positive value of Δk will indicate accumulation of moisture at the grid point, i.e. it will be a moisture sink. To calculate the local network flux, I make use of the matrix A^q evaluated as defined in Sec. 5.3.2.

5.3.5. Identifying anomalous monsoon years

The above described measures can help us in studying the spatial structures of rainfall fields and their properties but not its temporal evolution. In the following section, I develop and present a new scheme, which uncovers some details of the temporal evolution of the ISM rainfall patterns. I apply this to the same data set over last 6 decades and provide new insights into the spatiotemporal complexity of monsoonal rainfall. The underlying assumption is that the above constructed network has the minimum essential correlation structure of the rainfall field. I thus suggest that extreme events occur within this structure and will deviate from this synoptic structure only if the ISM is abnormal. I can use this observation to find anomalous behaviours of the ISM and to identify regions where the ISM rainfall has the most intricate spatial structure. Usually, the method employed for discovering anomalous monsoon behaviour is based on a standard deviation of a rainfall index, which may be biased by inhomogeneous spatial rainfall distributions. The method described in the following paragraphs is not impacted by the large spatiotemporal discrepancies in rainfall data.

Let us assume that a rain event occurs at some grid point i. This will make any other grid point j vulnerable to such an event too, if there exists a link from i to j. This type of vulnerability must be inversely depended on the distance from i to j. Apart from the link strength its directionality is also important, as only those grid points will be vulnerable which

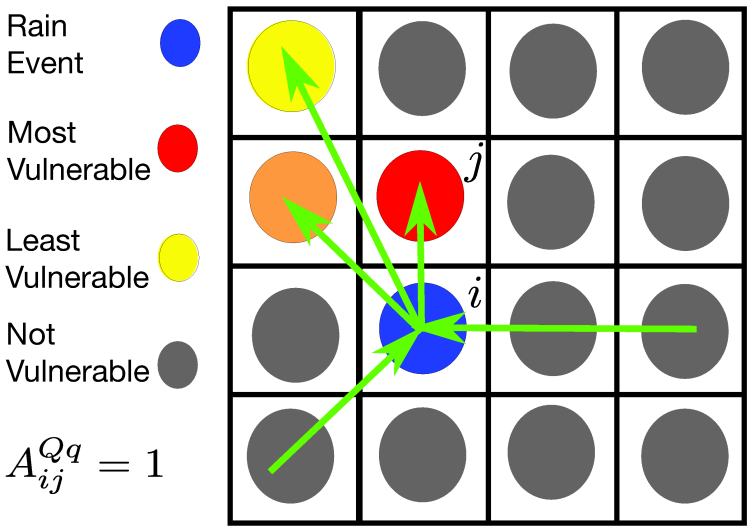

Figure 5.3 A simple schematic representation of the calculation of ρ_j (Eq. 5.9). The matrix depicts the square grid points over land. j is most vulnerable as it is geographically closest to i. Only incident links contribute to the vulnerability of a grid point and may have an extreme rain event.

have incoming links from i. All this information can be obtained from the adjacency matrix A^{Qq}. A schematic explanation of the above is presented in the diagram in Fig. 5.3. I can now write the *vulnerability* of grid point j receiving rainfall to be

$$\rho_j = \frac{\sum_{i \in V(t)} \frac{1}{L_{ij}} A_{ij}^{Qq}}{\sum_{i=1}^{N} \frac{1}{L_{ij}} A_{ij}^{Qq}} \tag{5.9}$$

where $V(t)$ the set of grid points where rain event of the type α happened at time t and N is the total number of grid points. L_{ij} is the geographical distance between grid points i and j. ρ_j is calculated from the first half of the data set and predicted for the second half. I set a pre-assigned value for ρ_j to find the prediction accuracy and assess the evolution of monsoonal rainfall patterns and also their spatial complexity.

5.4 Results and Discussion

First, I discuss the degree centrality and its distribution obtained from the adjacency matrix A. Then I present the analysis of spatial scales and clustering coefficients. Further, I introduce the results on centrality measures, followed by an attempt to visualise the links within the network. At the end of the discussion, I analyse the spatiotemporal evolution of ISM rainfall patterns using the new scheme described in Sec. 5.3.5.

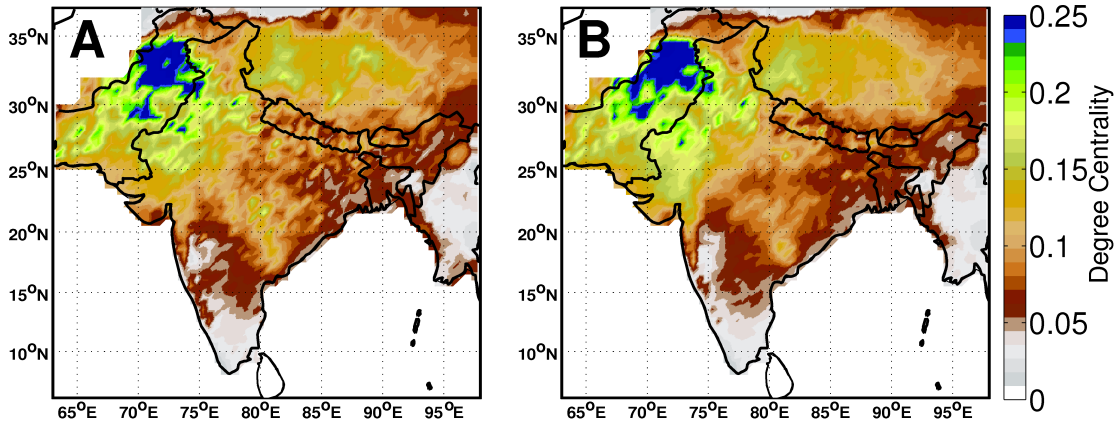

Figure 5.4 Degree centrality C_{Dj} (A) $\alpha = 94\%$ (B) $\alpha = 90\%$. Obtained from the matrix A it gives the number of links to a grid point and it is normalized between 0 and 1 by dividing it by $N-1$ i.e. the total possible number of links to grid point. Note higher C_{Dj} in Northwest Pakistan and lower values on the southern Indian peninsular.

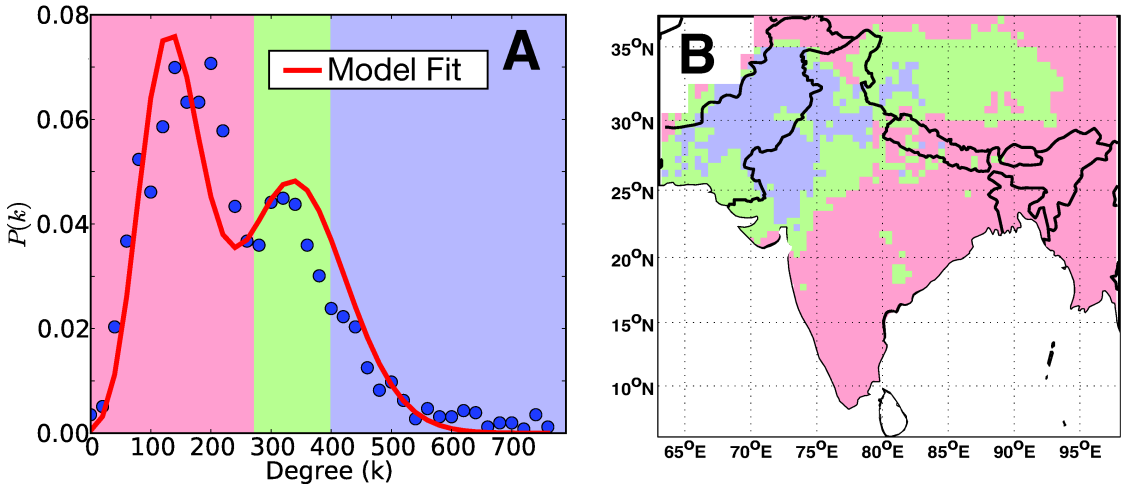

Figure 5.5 Spatial degree distribution. (A) Measured degree distributions are shown in blue dots and red line indicates the model fit. (B) depicts the spatial distribution of coloured regions shown in (A). The region in blue and green receive rainfall only during the large-scale spatial activity of monsoon, i.e. during its active phases, whereas the regions in pink also receive rainfall during break phases.

5.4.1. Degree centrality and Degree distribution

The spatial patterns of degree centrality C_{D_j} are very similar for both thresholds of $\alpha = 94\%$ and $\alpha = 90\%$. Higher degrees are observed in northwest Pakistan and lowest values occur in southeast India (Fig. 5.4). This suggests that higher degree emerges mainly due to longer spatial connections in these regions. This potentially can be related to the large spatial scale of ISM rainfall over these regions. To distinguish between regions where monsoonal rainfall is also due to localised or synoptic (large scale) activity of ISM, I show the distribution of the degree in Fig. 5.5 (A) for the case of $\alpha = 90\%$. I obtain a bimodal distribution and a model fit of the type

$$P(k) = \frac{{n_r}^{n_k} \exp(-n_r)}{n_k!} + \frac{{n_d}^{n_k} \exp(-n_d)}{n_k!} \tag{5.10}$$

where I found $n_r = 17.4$, $n_d = 8.0$ and $n_k = k/40$ (red curve in Fig. 5.5 (A)). Hence, $P(k)$ is sum of two poissonian distributions with different means. This implies that there must exist two different kinds of regions with their own characteristic number of links or distinct spatial scales of rainfall. Previous work has documented that the ISM has two modes, the active phase and the break phase. Therefore, it is possible that certain regions continue to receive rainfall during the break phases but certain other regions receive rainfall only during the active phases of ISM especially areas bounded by the monsoonal trough. To construct such a division of regions, I divide grid points into different zones (Fig. 5.5 (B)). This figure depicts the spatial regions associated with the degree distribution. Rainfall in central Pakistan northwestern India, and partly in the western Tibetan Plateau (Zada Basin) is expected to be the result of large spatial monsoonal activity. Heavy rainfall processes in the northwestern Indian subcontinent are associated with atmospheric interactions between western disturbances

and the Indian monsoon system. Where the trough existing in mid latitude westerlies penetrates southward and interacts with monsoonal trough over the Indo-Pakistan region causing recurvature of depressions and lows in $75°E - 78°E$ belt [*Ding and Sikka (2006)*]. This complex interaction is fundamentally composed of cold westerly winds interacting with warm, moisture-laden monsoonal winds over a very dry and hot region during the peak of summer with low pressure fields. Hence such interaction should lead to volatile convective instabilities in the atmosphere and which should not only produce extreme convective rainfall but also have far reaching effects on the internal dynamics of ISM. It is understood that this interaction first enhances monsoonal rainfall over the northwest and may ultimately lead to withdrawal of monsoonal trough to foothills of Himalayas i.e a break phase of ISM (*Ding and Sikka (2006)* and references therein). A recent study detailing the causes of floods in North West Pakistan during late July- August 2010 also hints that a similar mechanism was responsible for these floods [*Hong et al. (2011)*]. I will go into further details of this interaction and its possible influences in the next sections.

5.4.2. Median length of links

I provide further insight into the spatial scales involved in these regions by analysing the geographical length of these links. I will use the formula introduced in Eq. 5.8. This distance metric gives us the advantage of expressing spatial scales in length units. The median of L_{ij} for each grid point is shown in Fig. 5.6 (A). Clearly the characteristic scale of monsoonal rainfall for the 90th percentile seems to be below 250 km for most of the region. Also, some larger

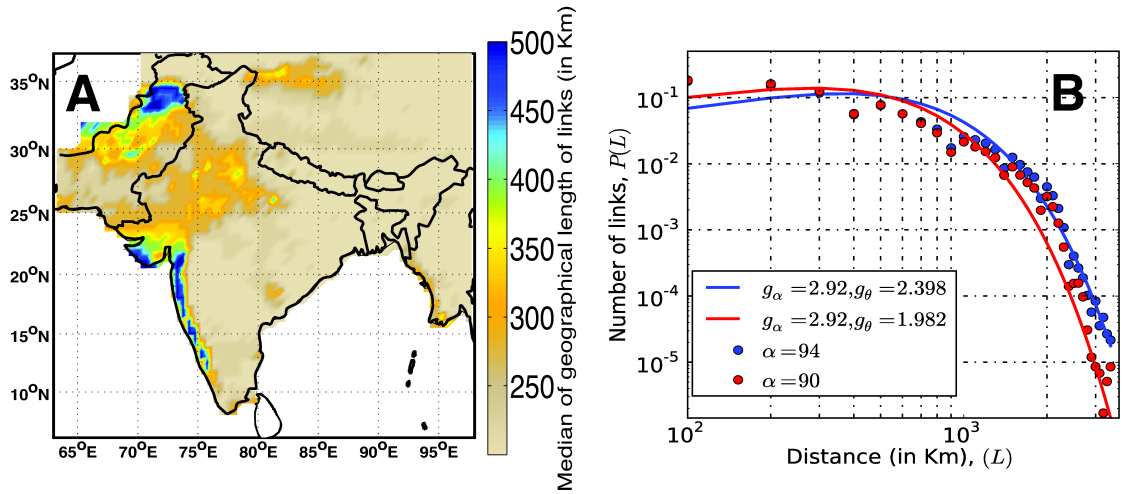

Figure 5.6 (A) Median of geographical length (km) of links calculated using formula for spherical Earth projected onto a plane. (B) The distribution of length scales for $\alpha = 94\%$ (blue) and $\alpha = 90\%$ (red) and their corresponding fits. Note that more extreme rainfall events ($\alpha = 94\%$) have longer characteristic spatial scales as the blue curve results in higher numbers of links at longer distances. However, the longest recorded spatial scales are associated with $\alpha = 90\%$, because the red curve has a longer tail.

spatial scales above 500 km exist, such as in northwest Pakistan and the southwestern coast of peninsular India. I observe that most of the larger spatial scales exist in the region with medium to high degree distributions (green and blue colors in Fig. 5.5 (B)). This supports the statement that these regions receive rainfall from large spatial monsoon activity stretching over distances of 500 km, a characteristic for the active monsoon phase. Fig. 5.6 (A) also provides some supplementary information to Fig. 5.4 and Fig. 5.5, about the south west coast of peninsular India along. Due to the significant topographic barrier of the western Ghats (cf. Fig. 2.4) this region has a generally smaller degree centrality (see Fig. 5.4). However, I observe high spatial scales for this region (Fig . 5.6 (A)) and associated these with rainfall coeval with rainfall south of the Himalayas during the active monsoon phases.

Next, I analyse if the spatial scales follow an analytical form. For this purpose, I plot the distribution of distance versus the number of links (Fig . 5.6). The fitted bold lines are gamma distribution of the form

$$P(L) = {n_L}^{(g_\alpha - 1)} \frac{\exp(-n_L/\theta)}{\Gamma(g_\alpha)\theta^{g_\alpha}} \tag{5.11}$$

where $n_L = L/100.0$ and the values of other parameter are given in the legend of the Fig. 5.6 (B) for the fitted curves. Fig. 5.6 (B) indicates that I observe longer spatial scales in the smaller events. Although characteristic scales must be larger for stronger rainfall events at $\alpha = 94\%$, because these values are higher for most part of the distribution (blue points in Fig. 5.6 (B)). The analytical form implies that more extreme rain events are more spatially localised.

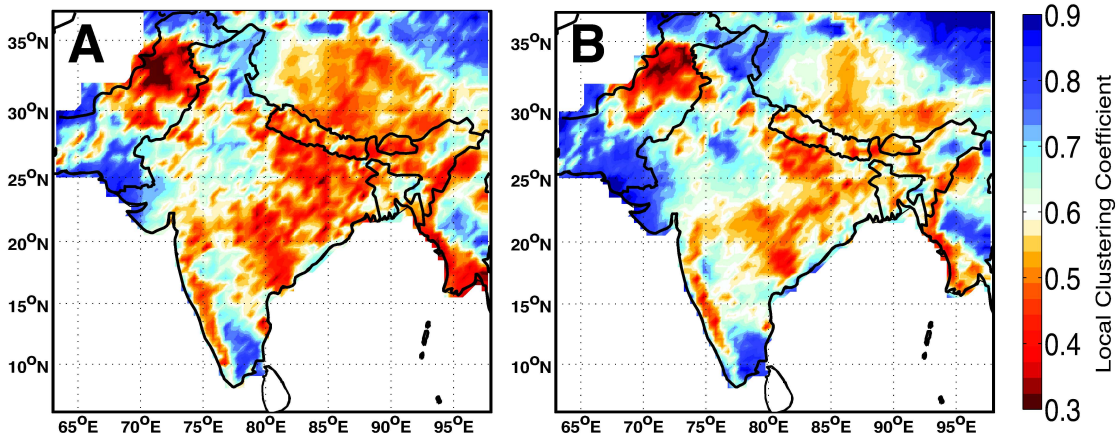

Figure 5.7 Local clustering coefficient $\mathscr{C}_j$ (A) $\alpha = 94\%$ (B) $\alpha = 90\%$. Red colours indicate that the rainfall field is less spatially continuous, i.e. it is fragmented. In contrast, blue colours outline more spatially continuous rainfall fields. I observe that $\mathscr{C}_j$ is independent of the spatial scales involved in the rainfall (compare with Fig . 5.6 (A)).

5.4.3. Clustering coefficient

The local clustering coefficient $\mathscr{C}_j$ shows the spatial organisation of rainfall with respect to a reference grid point. The field may have a large spatial extent but whether it is highly fragmented or spatially continuous cannot be inferred from the above measures and analysis. However, I can derive this additional information from the values of $\mathscr{C}_j$. I associate lower values of the clustering coefficient with more fragmented or spatially discontinuous rainfall fields, whereas larger values represent clustered activity. Importantly, this measure is independent of the involved spatial scales. The spatial pattern of $\mathscr{C}_j$ for the two rainfall threshold I am using is very similar to each other (Fig. 5.7 (A) and (B)). Large areas of low clustering values $\mathscr{C}_j$ related to fragmented rainfall are located in northwest Pakistan and in central and eastern India. I observe higher clustering coefficients in south Pakistan, parts of the Tibetan plateau, and in the northwestern and southeastern parts of India (cf. Fig.2.4 and 5.7). This suggests that stronger rainfall events are more spatially clustered in these regions. Northwest Pakistan has lower values of clustering coefficients, which indicates that this region receives rainfall due to the large spatial activity of the ISM, but with a spatially fragmented rainfall field. Also, comparing the two panels in Fig. 5.7 (A) and (B) indicates that stronger rainfall events are more spatially fragmented than smaller ones.

5.4.4. Centrality measures

Closeness centrality C_{Cj} has been introduced in Sec. 5.3.3 and it can be employed in the task of identifying the grid points which perform a critical role in the functioning of this network structure. C_{Cj} is plotted in Fig. 5.8 (A) and shows that regions of highest closeness centrality lies in the northwestern subcontinent with a focus on northwest Pakistan. As described in Sec. 5.3.3 this indicates that the information travels fastest to and from these points. Any pertur-

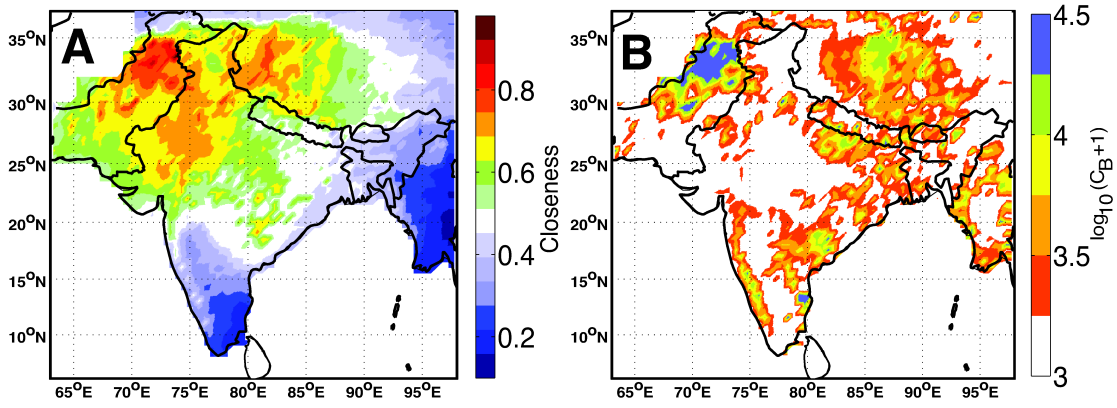

Figure 5.8 (A) Closeness centrality C_{Cj}. I observe high C_{Cj} in the northwestern parts of the subcontinent suggesting the importance of atmospheric processes in modulating the ISM activity. (B) Betweenness centrality C_{Bj}. Higher values of C_{Bj} represent the moisture transport pathways over the land during the active phase of the ISM. For both (A) and (B) I chose $\alpha = 90\%$.

bation occurring in this region will effect the monsoonal rainfall patterns at rapid temporal scales. The existence of any atmospheric instability over this region will have an immediate and widespread effect over ISM rainfall. It has also been observed that this region is near the boundary between two synoptic systems: the westerly and the monsoonal trough. Spatial fluctuations in the westerly trough and its southward penetration could lead to an interaction resulting in a complex modulation of the ISM activity over rest of the land mass [*Ding and Sikka (2006)*]. To provide further evidence that this may be an important mechanism during the most active phase of the ISM, I make the following calculation. First, I identify a set of 50 grid points with 50 top most values of C_{Cj}. I found that all the grid points in this set have $C_{Cj} > 0.7585$, (see Fig. 5.8(A)). Next, I calculate the linear cross correlation r between the total number of events above the threshold $\alpha = 90\%$ that occurred within a set of 50 grid points and the remainder of the land mass. The value r was found to be 0.535 (refer Table. 5.1) and $C_{Cj} > 0.7585$. Most of these 50 grid points are located in northwest Pakistan (see Fig. 5.8 (A)). This indicates that interaction of western disturbances with the ISM plays an important role in generation of the extreme rainfall events over large parts of the Indian subcontinent. The above results indicate that an influence of such interactions must be far more spatially extensive then limited to the northwestern parts of the subcontinent. I will further analyse the underlying atmospheric mechanism in the next section.

I have plotted the betweenness centrality C_{Bj} to analyse and visualise its spatial structure (Fig. 5.8 (B)). Higher values of C_{Bj} are observed over large parts of Tibet, the east coast of peninsular India, parts of central India, the central Gangetic plains, northwest Pakistan and along the western Ghats. From a mathematical point of view, the higher values of C_{Bj} highlight the main pathways of information travel in a network. In the above case, moisture is the quantity assumed to be traveling through the network and therefore the analysis highlights the main pathways of moisture transport during the ISM. Pathways of moisture transport are modulated and facilitated by the existence of deep convection and the underlying topography [*Roe (2005)*, *Webster et al. (1998)*, *Bookhagen and Burbank (2006)*, *Bookhagen and Burbank (2010)*, *Bookhagen (2010)*]. Hence, higher values of C_{Bj} in Fig. 5.8 (B) represent the regions where deep convection ceases to exist during the active phase of monsoon.

5.4.5. Visualising links of complex networks for the study region

The links that are obtained from the matrix A, and are non directional. In Fig. 5.9 I have plotted the number of links in relation to 50 grid points selected for a particular region (highlighted by a gridded, red colour matrix in Fig. 5.9). The number of links to other regions are shown by the colour scale ranging from 0 to 50. In Fig. 5.9 (A) I observe that links over northwest Pakistan extend deep into central India and also connect to parts of Tibet and to almost the entire western and central Himalayan region. As stated previously, this indicates that stronger rainfall events over northwest Pakistan are likely the result of large spatial scale monsoonal activity. In Fig. 5.9 (A) I am able to visualise this spatial extent and spatial structure of monsoonal activity. Clearly the spatial structure in Fig. 5.9 (A) re-emphasises the

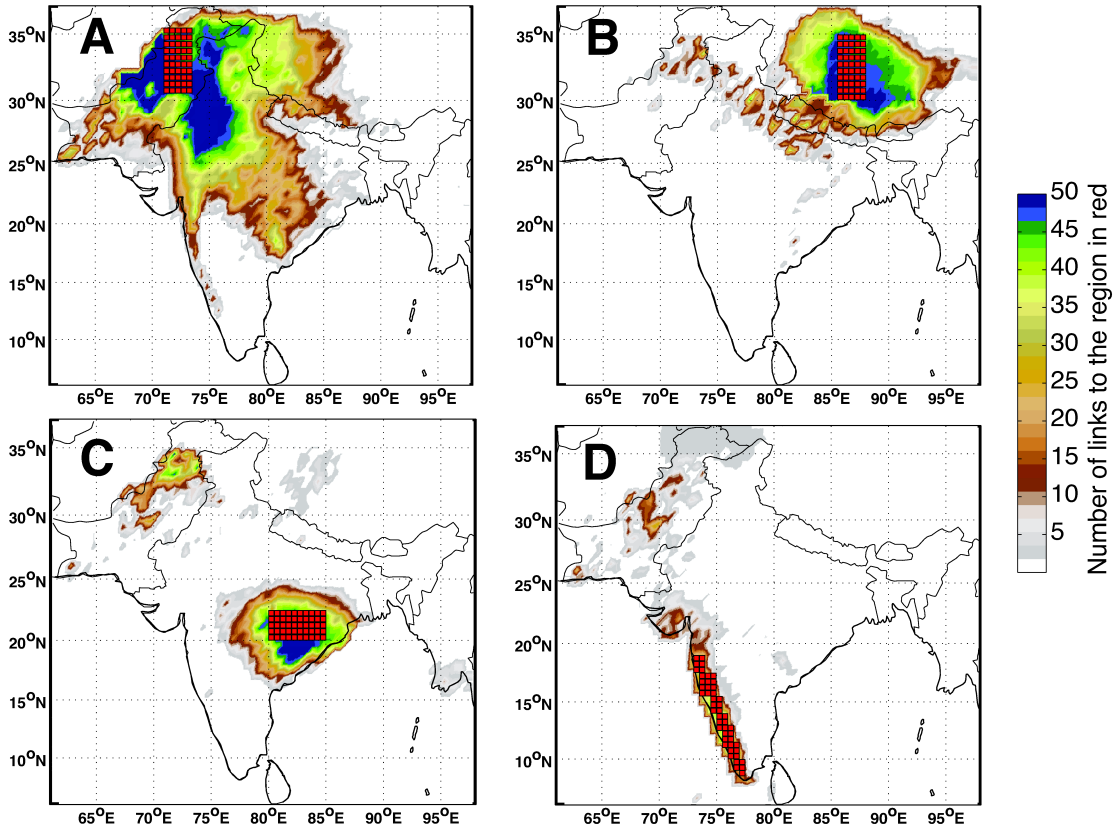

Figure 5.9 Links between a set of 50 reference grid points (gridded red matrix) to other grid points (colourbar) at $\alpha = 90\%$. Note the spatially extensive links for a reference area in northwestern Pakistan (A). In contrast, extreme rainfall linkages in the western Ghats (D) have a limited spatial extent.

significance of the mechanism mentioned in Sec. 5.4.3. It indicates the extensive influence on the extreme rain events that occur in other parts of the subcontinent. The number of links on the Tibetan plateau are high and they appear to be much more localised (see Fig. 5.9 (B)), as high topographic barriers exist to the west and south. In Fig. 5.9 (C) I observe an interesting feature for the central Indian region: there are two major geographically disconnected regions with high number of links. This may happen due to formation of deep convective cells over the region of NW Pakistan and adjoining regions and is dominating during the active phase of the ISM. We also observe some localisation in the south west India Fig. 5.9 (D). This localisation is caused by the topographic barriers formed by the western Ghats. We also observe a few long-range connections from this region to almost the entire west coast of India and western parts of the subcontinent. These links may emerge due to the existence of an offshore trough along the west coast of India and embedded mesoscale vortices during active phase of monsoon [*Ding and Sikka (2006)*]. This synoptic configuration facilitates convection along the west coast and the orographic forcing of western Ghats acts as boundary to this convection.

Region	No. of grid points	Ref. Fig.	r
Tibet	50	5.9 (B)	0.4495
NW Pakistan	50	5.9 (A)	0.4675
Central India	50	5.9 (D)	0.3182
West coast of India	50	5.9 (C)	0.2992
Closeness > 0.7585	50	5.8 (A)	0.5351
log(Betweenness+1) > 4.358	50	5.8 (B)	0.5325
Degree centrality > 0.26	50	5.4 (B)	0.4893

Table 5.1 Linear correlation r between extreme rain events in a region and to the remainder of the subcontinent for $\alpha = 90\%$

In Table 5.1 I provide the linear cross-correlation r between rain events at $\alpha = 90\%$ occurring in the marked regions during the ISM season to similar events that occur in the rest of the region (Fig. 5.9). We observe that the highest linear cross-correlation are the ones in northwest Pakistan. This strengthens the argument that extreme rainfall events in this region are due to the large spatial activity of the ISM. The interaction of western disturbances and the ISM could be an intensifying monsoonal activity over the Indian subcontinent [*Dimri (2004)*]. However, no detailed study, at least to my knowledge, exists about this interaction and the penetration depth across the Himalayas and Tibet.

An immediate practical application of my methodology can be derived from Fig. 5.9 and Table 5.1: These regions outline areas that are best suited for palaeoclimatic proxies that reconstruct the Holocene and historical monsoonal activity. Above results suggest that the best region lies in the northwestern Indian subcontinent centred in northwest Pakistan and within the northwestern Himalayas because only very strong and significant active phases of the ISM transport rainfall to this region.

5.4.6. Directed networks

To obtain the sinks of moisture over land, I plot the local network flux Δk as defined in Sec. 5.3.4 in Fig. 5.10. Higher positive values exhibit moisture sinks over land and are highlighted in red colours in Fig. 5.10. One of the pronounced moisture sinks spreads from the central to lower Gangetic plains along the foothills of the central Himalayas. This region is known for the formation of a high number of low-pressure systems (monsoonal lows and depressions) [*Mooley and Shukla (1989)*]. The large spread of sinks is related to the movement of low-pressure systems over land, as they are not constant spatial feature [*Sikka (1977)*]. The extent of the red areas in the map indicates the region in which most of these systems form. A second major moisture sink shown in Fig. 5.10 is located in Pakistan near the western boundary of the average monsoonal-trough location [*Ding and Sikka (2006)*]. Some other

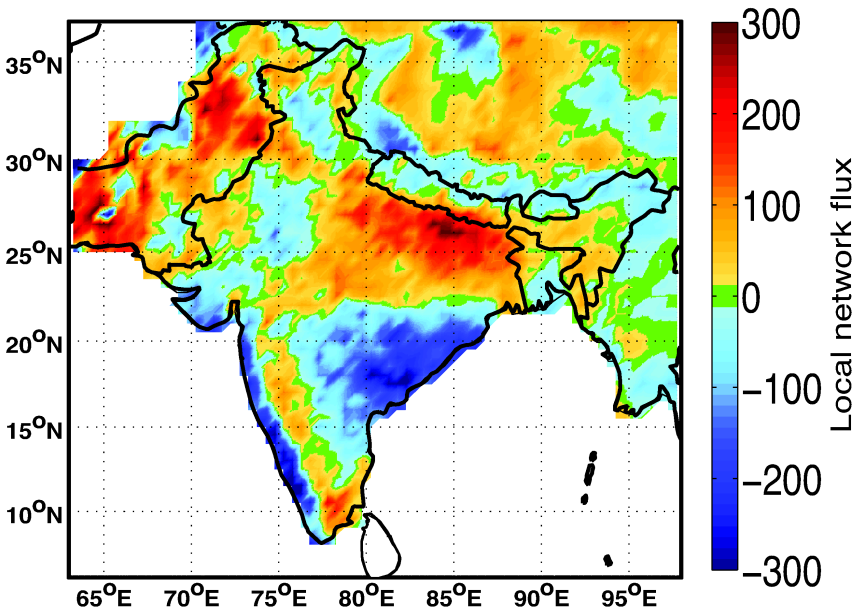

Figure 5.10 Local network flux obtained from directed networks i.e. matrix A^q, $\alpha = 90\%$.

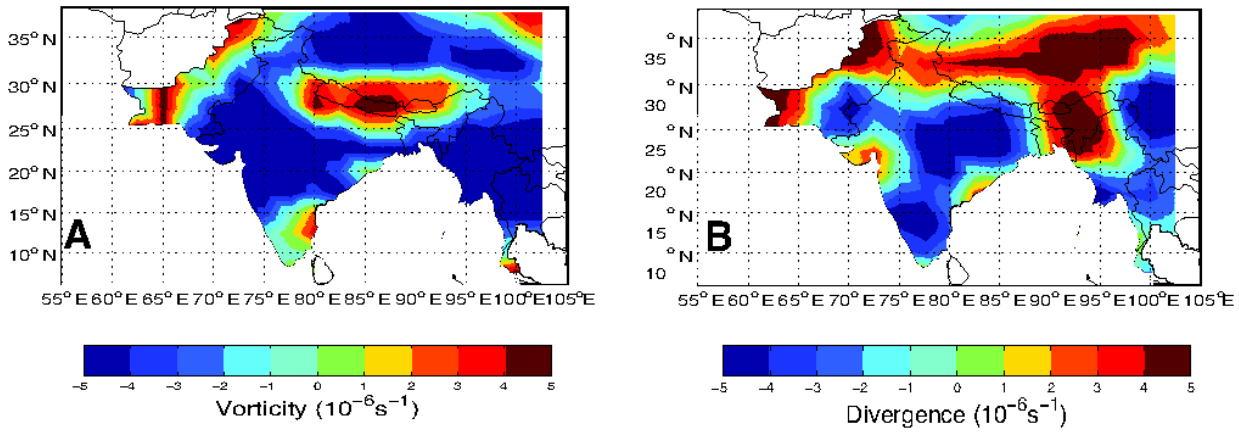

Figure 5.11 Relative vorticity (A) and Divergence (B) obtained from the NCEP/NCAR reanalysis data set for wind at the height of 850 hPa. We observe high vorticity over the central Gangetic plains and adjacent regions in Nepal, Tibet, and parts of Pakistan. When comparing (A) with Fig. 5.10, most high vorticity regions correspond to moisture sinks. The divergence zone over northern Pakistan shows the western boundary of the monsoonal trough during extreme events considered in the study.

minor sinks are observed on the Tibetan plateau and along the western Ghats. These sinks are rainfall accumulation from different directions, i.e. moisture convergence zones and do not necessarily indicate higher rainfall amounts. This type of moisture convergence is likely generated by the underlying heat balance and orographic effects over these regions [*Bhide et al. (1997)*]. The regions with high negative values of local network flux are the regions, which are closer to the moisture sources.

In this study I have considered extreme rain events (90% and 94% percentile), which are usually a result of convective rainfall processes [*Schumacher and Houze (2003)*]. The mesoscale convective systems (MCS) play an important role for generating these rainfall amounts[*Johnson (2006)*, *Houze et al. (2007)*]. Therefore, the identified sinks and sources can be associated to spatial patterns of MCS within the ISM region. Apart from the underlying heat balance over the land, the vorticity and divergence of wind vectors can also provide information about the spatial structures of the MCS and corresponding locations of sources and sinks of moisture (moisture convergence zones). To better understand the origin of moisture sinks over land, I present a comparison of Fig. 5.10 with wind vorticity and divergence (NCEP / NCAR data) (Fig. 5.10). For obtaining the time indices of the wind vector, I have used the same thresholding procedure as used for APHRO–V1003R1 and thus I maintain a statistical consistency. The relative vorticity (ζ) and divergence (δ) were computed using the formulas

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial x} \quad , \quad \delta = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \tag{5.12}$$

where u and v are zonal and meridional wind components at the height of 850 hPa. I observe positive vorticity over the central Gangetic plains extending over the central Himalayas and Tibet (Fig. 5.11 A). Also, positive vorticity is observed over Pakistan and southeast India. These are also the two regions where I observe higher local network flux. Positive vorticity is associated with cyclonic rotation and can often be related to low pressure areas, which drive the moisture inward into the land. Divergence shows convergent winds over the central Gangetic plains and in central India (Fig. 5.11 B). The shape of the divergence zone in Pakistan appears to show the boundary of monsoonal trough during these events. The similarity in spatiotemporal extent of divergence and vorticity during the extreme rainfall events suggests that these structures are required for causing extreme rainfall events.

5.4.7. Identifying anomalous monsoon years

In this section, I will present an application and description of the mathematical scheme developed in Sec. 5.3.5 to (i) identify regions in which ISM rainfall has the most intricate and temporally unstable structure, to (ii) identify anomalous monsoon years and their deviation from the normal spatial structure of the ISM rainfall field, and to (iii) decipher the temporal evolution of spatial structure of the ISM rainfall.

The first requirement for using the scheme described in Sec.5.3.5 is to determine a thresh-

old on ρ_j (c.f. Eq. 5.9). Then I will use this threshold to ascertain the accuracy of my prediction. Prediction based on ρ_j will be in binary space, i.e. if ρ_j is greater than a predetermined threshold, an event (1) occurs at grid point j, else no event (0) occurs at j on this day. We thus need two different kinds of ratios to examine the quality of the prediction. First, I have $\frac{\varepsilon_p}{E_p}$, which is the ratio of the correctly predicted events (ε_p) to the total number of events predicted (E_p). Second, I will use a ratio $\frac{E_p}{E_T}$, i.e. total number of predicted events to the total number of events that occurred (E_T). I will use a quantile value of ρ_j, i.e. $Q[\rho]$ as a threshold on ρ_j.

The value of $Q[\rho]$ is chosen such that $\frac{\varepsilon_p}{E_p} \sim \frac{E_p}{E_T}$ has the best accuracy in the first half of the data. The curves between the above two ratios are shown in Fig. 5.12 (A) and (B) without delay and with one-day delay, respectively. The red lines indicate the points where $\frac{\varepsilon_p}{E_p} \sim \frac{E_p}{E_T}$ is satisfied. The delay of one day was introduced by changing the set $V(t)$ to $V(t-1)$ in Eq. 5.9. The value of $Q[\rho]$ has been estimated from the first half of the data set and the second half is predicted. I will use the symbol ε for the accuracy of prediction and it is given by $\varepsilon = \frac{\varepsilon_p}{E_p}$ where $\frac{\varepsilon_p}{E_p} \sim \frac{E_p}{E_T}$. In Fig.5.12 (A) we observe $\varepsilon = \frac{\varepsilon_p}{E_p} \sim \frac{E_p}{E_T} \sim 0.7$, i.e. 70% of the events were correctly predicted when I did not use a time delay. The curve in Fig. 5.12 (B) suggests that the scheme described above does not have practical usefulness for prediction of extreme rainfall events. Introducing a delay of 1 day, we observe that the accuracy drops to 30%. I will only be using the ε (also referred as accuracy in this text) without delay as a measure of spatiotemporal intricacy of the ISM rainfall field. The accuracy (ε) will be higher

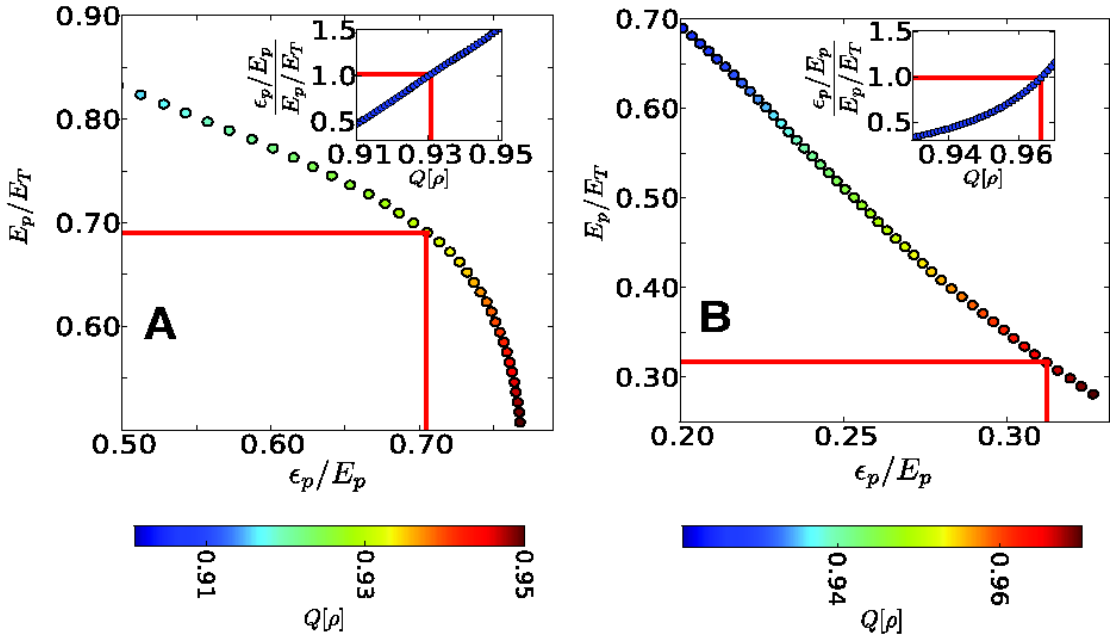

Figure 5.12 Estimation of ρ_j from the first half of the data set. The value of the ρ_j used is one where the condition $\frac{\varepsilon_p}{E_p} \sim \frac{E_p}{E_T}$ is satisfied. (A) no delay, (B) one-day delay.

if the correlation structure of extreme rainfall events obtained using the above methodology is least intricate and more temporally stable. The accuracy metric ε is known as *probability of detection* (POD). Using the same threshold on $Q[\rho]$ I compare the quality of prediction with other skill scores in Fig. 5.14. I have used the *threat score* (TH) which is defined as $TH = \frac{\varepsilon_p}{\varepsilon_p+\varepsilon_f+\varepsilon_m}$. Where ε_f the number of false alarms, i.e. predicted events that did not occur and ε_m is the number of misses, i.e. no predicted event but an actual event occurred. Also, $E_P = \varepsilon_p + \varepsilon_f$ and $E_T = \varepsilon_p + \varepsilon_m$. TH is more balanced score and it ranges between 0 and 1 (Fig. 5.14 (A)). An additional similar skill score is the *equitable threat score* (ETH), which is defined as $ETH = \frac{\varepsilon_p-\varepsilon_r}{\varepsilon_p+\varepsilon_f+\varepsilon_m-\varepsilon_r}$. Where $\varepsilon_r = \frac{E_P E_T}{N_d}$, N_d is the number of predicted days. ε_r gives the number of prediction correct by chance (Fig. 5.14 (B)). We observe that the basic pattern of prediction quality remains almost the same for all measures. However, values of prediction are slightly lower for most regions based on TH and ETH as compared to POD.

Following the procedure described above, I calculated ε for each grid point. I have plotted ε in map view and we observe that for most of the subcontinent the accuracy (ε) was above 70% and even reaches up to 100% in places (Fig. 5.13). We also clearly show that regions with high and complex topography such as the Himalayas and the western Ghats are characterised by the lowest ε values. From this spatial pattern I infer that these regions are characterised by the most intricate rainfall patterns and high temporal fluctuations.

Next, I employ the above scheme to analyse the anomalous behaviour of the ISM. My basic assumption for this task is that the complex-network construct inherits the most essential

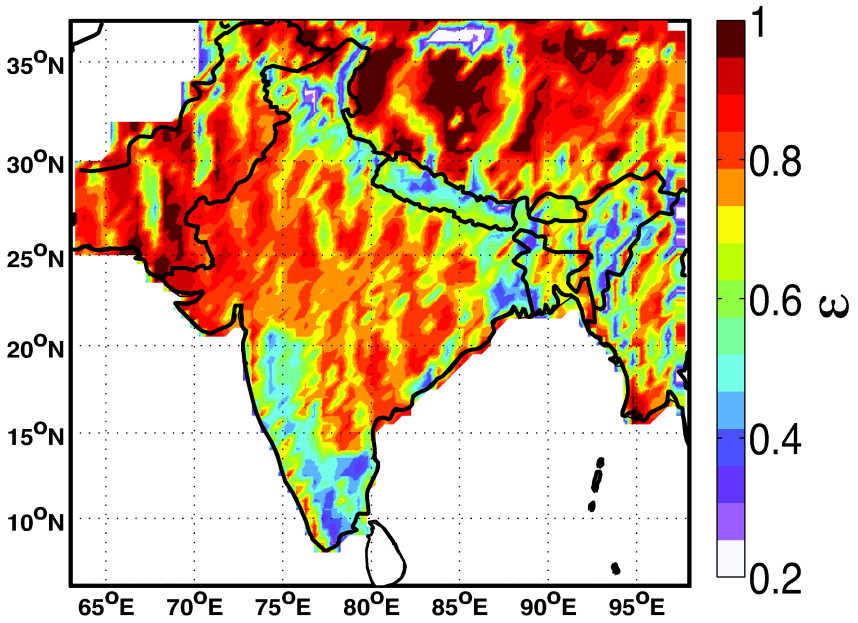

Figure 5.13 The accuracy of prediction ε on a scale from 0 to 1 for the last half of the data set over the map, calculated without delay in the prediction scheme. The lower values of ε indicate higher intricacies in rainfall patterns. Observe lower values of ε along the Himalayas and the western Ghats in southern peninsular India.

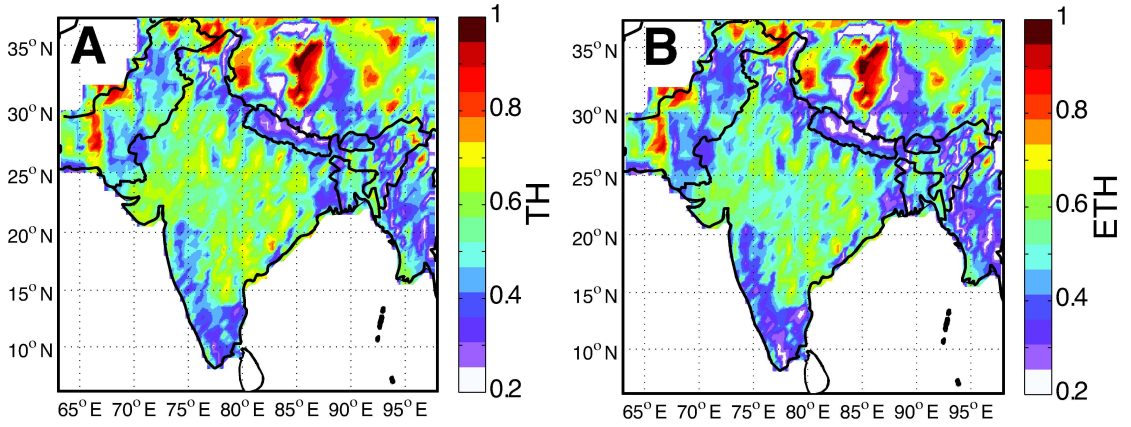

Figure 5.14 Comparison of the prediction quality using several skill scores (see text for explanation). (A) The threat score (TH). (B) The equitable threat score (ETH). The basic spatial prediction pattern remains the same regardless of skill score (cf. Fig. 5.12). However, we note the values derived form TH and ETH are slightly lower than ε for almost all the regions.

structures and patterns of extreme rainfall events. Hence, any considerable deviation from this essential structure should suggest anomalous behaviour of the ISM – an exceptionally weak or strong, or an abnormal monsoon. I will consider prediction accuracy ε per year for the entire region to be the measure of this deviation. In Fig. 5.15 (A) I plot a scatter plot of prediction accuracy ε per year for the entire region versus the z-score of AIMRI (All-India Monsoon Rainfall Index) [*Parthasarathy et al. (1995)*]. The significance band of ε was obtained by bootstrapping the spatial sum of ε for each day [*Davison and Hinkley (2006)*]. That is, I randomly draw 122 days with replacement from the spatial sum over ε for all the grid points. We observe that the prediction accuracy ε is lower for weak monsoon years and is high for strong monsoon years (Fig. 5.15 (A)). All the weak monsoon years are below the upper limit of the significance band and similarly all the strong monsoon years are above the upper limit of the significance band. This indicates that the prediction accuracy ε can distinguish between normal and abnormal monsoon. I emphasise that ε tends to distinguish a monsoon year in terms of the spatial organisation and structure of the extreme rain events (90th percentile) in a particular year. Years lying within the significance band are characterised by no significant change in spatial organisation and structure of the extreme rain events.

E^{α} is the total number of events above the threshold $\alpha = 90\%$ in a year over the whole subcontinent. The cross correlation between E^{α} and AIMRI was found to be 0.83. I use E^{α} instead of AIMRI for further analysis as AIMRI is limited to the year 2000. In Fig. 5.15 (B) I plot a scatter plot of E^{α} and ε. It is thought that there exists a dynamical coupling between El Niño Southern Oscillation (ENSO) and ISM. This coupling has been of wide interest in the scientific literature (e.g., *Kumar et al. (2006), Kumar et al. (1999), Mokhov et al. (2011), Maraun and Kurths (2005)*). Thus, I attempt to identify possible influences of ENSO on the spatial structure of the ISM rainfall field. I have indicated the El Niño and La Niña years in Fig. 5.15 (B) and we observe many events where the ISM has been weak during El Niño years (left to

the vertical red lines in Fig . 5.15 (B)) and the accuracy has been low. When a strong monsoon has followed La Niña (right to the vertical blue line in Fig . 5.15 (B)) there is a generally higher accuracy. We also observe several exceptions to this rule: some (El Niño years) are not only within the significance band but also above the significance band. In summary, above results suggest that El Niño and La Niña are not resulting in a complete breakdown of the spatial structure of monsoonal rainfall or in its complete reorganisation. ENSO's influence on ISM rainfall pattern appears to be highly complex. These exceptions could be caused by the temporal evolution of the suggested weakening of the coupling between ENSO and ISM [*Kumar et al. (1999)*]. The dynamical nature of this coupling could also be another reason, as it has been hypothesised that ENSO-ISM relationship is a bi-directional phenomenon rather than ENSO directly influencing ISM only[*Mokhov et al. (2011)*].

I can also take this analysis a step further to understand how the ISM has evolved during the time span of the data set. This will give additional information about the changes in the organisation and structure of the ISM extreme rainfall field during times of global warming [*Goswami et al. (2006)*, *Ramanathan et al. (2005)*, *Levermann et al. (2009)*, *Zickfeld et al. (2005)*]. For this purpose I plot the temporal evolution of ε in Fig. 5.15 (C). Between $1951 - 2000$, we do not observe any drastic trend in the values of ε. It is merely fluctuating between the higher values for strong monsoon and the opposite for weak monsoon years. However, a small change in the mean level of fluctuation is observed around the late 1960's. Interestingly, we observe a characteristically distinct evolution of ε since the year 2000 with ε not following the previously determined rule with high values for strong monsoon and the opposite for weak monsoon. It is continually and strongly increasing, for example the years 2005, 2006, and 2007 show the highest values of ε. This indicates that some basic structural change may have occurred in the rainfall patterns over the Indian subcontinent during this period. Increase in ε can only be due to considerable increase in spatial correlations during this period and it is known that increasing spatial correlations are an early warning of approaching "tipping points" or abrupt dynamical transitions in dynamical systems [*Lenton (2011),Lenton et al. (2008)*]. I speculate that we are reading a tipping point in this system, which has some particular precursor in the dynamics [*Lenton et al. (2008)*, *Zickfeld et al. (2005)*, *Levermann et al. (2009)*]. A plausible reason for this is a shift in the distribution of magnitude and frequency of extreme rainfall events, caused either by rapid changes in surface heat fluxes due to the increase in aerosol content [*Ramanathan et al. (2005)*] or that some inherent monsoonal dynamics has changed during this time [*Levermann et al. (2009)*]. A few recent studies suggested a significant increasing trend in the frequency and magnitude of extreme rain events and a significant decreasing trend in the frequency of moderate events over central India during the monsoon seasons from 1951 to 2000 [*Goswami et al. (2006)*] and 1901-2004 [*Rajeevan et al. (2008)*]. My above arguments are only speculative and subject to further research.

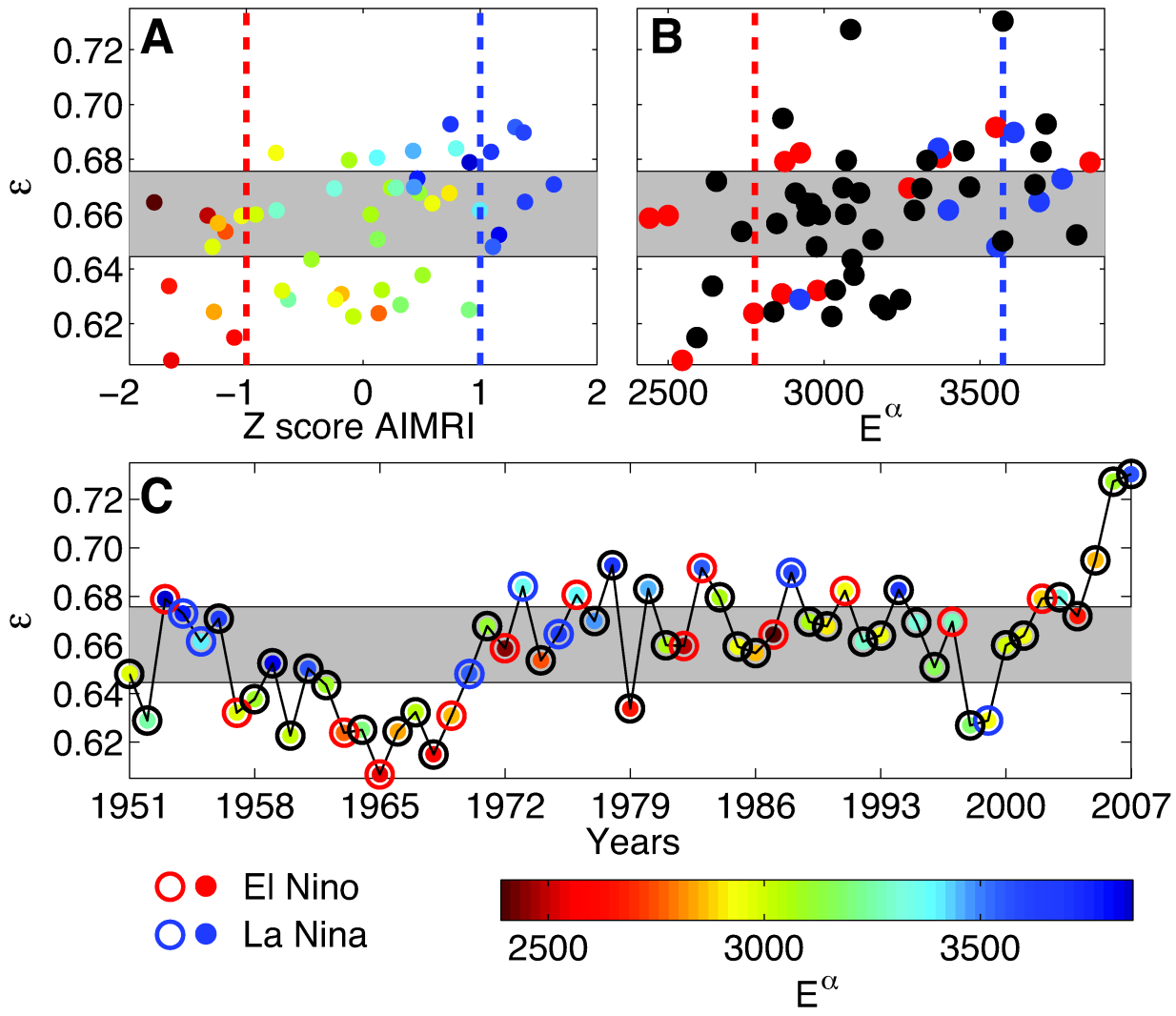

Figure 5.15 (A) Scatter plot between AIMRI (All-India Monsoon Rainfall Index) and ε (accuracy of prediction). The red dashed vertical line is the -1 sigma standard deviation of AIMRI, i.e. to the left of it are the weakest monsoons years and the blue dashed vertical line is the $+1$ standard deviation of AIMRI, i.e. to the right of it are the strongest monsoon years. The colour code gives E^{α}, which is the total number of events above the used threshold α in a year over the entire region. We observe that years with large number of events accumulate at the top of the significance band indicating that ε is higher for stronger monsoon years and lower for weak monsoon years, i.e. years with less events accumulating at the bottom of the significance band. (B) Scatter plot between E^{α} and ε and different colours representing the El Niño years (red), La Niña years (blue) and Non El Niño/ La Niña years (black). Left to the red vertical line are the weakest 15% monsoon years in terms of E^{α} and right to the blue line are top 15% of the monsoon years in terms of E^{α}. (C) The temporal evolution of ε. El Niño years (red circles), La Niña years (blue circles) and Non El Niño / La Niña years (black circles). Colours within the circles are E^{α} values. We note a strong increase in ε since the year 2000 with highest values of ε occurring for the last three years.

5.5 Summary of key findings

I have analysed the spatial structure and organisation of the monsoonal rainfall field. I have obtained the correlation structure of extreme rainfall events ($>$ 90th and $>$ 94th percentiles) employing nonlinear correlation of event synchronization. Furthermore, I have carried out a comprehensive spatiotemporal analysis using a complex network approach. Some of the findings provide new insights into the interaction of atmospheric processes responsible for the generation of extreme rainfall events during the ISM. The presented methodological approach is strengthened by reiterating previous findings and observation of the climatic features of the ISM. Because of the somewhat extensive and new methodology, I synthesise and list my important findings:

1. It has been previously shown that there exist two distinctive phases of activity within during the ISM [*Webster et al. (1998)*]: (1) the active phase, during which large areas of the Indian subcontinent receive extensive rainfall and (2) the break phase during which most regions receive no to very little rainfall. With the presented approach, I have been able to document the spatial manifestation of these phases and are thus able to support and validate presented approach. In addition, presented approach provides new insights and I was able to identify regions, which receive rainfall only during the most active phase of the ISM. In a second step, I was able to provide a quantitative measure of the median length scale involved in rainfall during the active phase of monsoon. This analysis of geographical-length scales shows that spatial scales above the 90th and 94th percentile rain events follow a gamma distribution. I determined that median length scale in these events are up to 250 km for most of the region.

2. I was able to identify the structure and organisation of the rain field in terms of its spatial discontinuity. Using clustering coefficients, I have determined that in northwest and southeast India, south Pakistan and in parts of the Tibetan plateau rainfall activity of the ISM occurs in more defragmented forms as compared to other parts of the subcontinent. This may be related to the fact that extreme rainfall events (at the 90% and 94% percentiles) in these regions are often associated with localised convective rainfall cells. .

3. The presented approach using centrality measures (degree, closeness, betweenness) suggests that atmospheric processed in northwest Pakistan play crucial role in the generation of large rainfall events over other parts of the Indian subcontinent during the ISM. In northwest Pakistan, mid-latitude westerlies interact with the monsoonal trough in a very dry and hot region enabling formation of convection instabilities. This particular interaction is also said to be responsible for generation of the ISM break phase. Hence I was able to establish with above methodology the importance of this particular mechanism on the internal dynamics of monsoon.

4. I have identified that the central Gangetic plains and parts of Pakistan are the major moisture sinks characterised by high amounts of moisture accumulation during the ISM. The location of these moisture sinks was also found to be consistent with vorticity and divergence of wind vector during the ISM. The central Gangetic plains are known for the formation high

number of monsoonal depressions during the ISM season.

5. I have developed a methodology based on causalities of rain events and the complex network approach to identify anomalous monsoon years. I find that regions with high topography and relief have the most intricate and unstable ISM rainfall patterns. Using a similar approach, I studied the temporal evolution of the ISM and its linkages to El Nino Southern Oscillation (ENSO). This analysis reveals that ENSO is not always resulting in a complete breakdown of the spatial ISM rainfall structure nor in its complete reorganisation. This findings supports the previously established understanding that the coupling between ENSO and ISM is of complex dynamical nature.

6. My analysis reveals that since the year 2000 the ISM exhibits characteristically different rainfall patterns as compared to the time period from 1951 to 2000. This is a new insight into the evolving complexity of monsoonal precipitation in a warming environment. As it has been stated in previous studies, there exists an increase in magnitude and frequency in extreme rainfall events during the ISM, while moderate rainfall events are decreasing. This characteristic feature in ISM rainfall distribution could be responsible for the observed change in rainfall patterns since 2000.

5.6 Conclusion

In this study, I have presented an analysis of the spatiotemporal Indian Summer Monsoon (ISM) rainfall distribution using nonlinear methods and complex networks. Above study improves previous analysis because it specifically takes into account the temporal disparities and spatial complexity of ISM rainfall. The analysis provides new insights into the interaction of different atmospheric processes responsible for the generation of extreme rainfall events (at the 90% and 94% percentile) during the ISM. In summary, this study not only opens up new opportunities for meteorologist to look at regional climate using tools from complex networks, but also provides new and valuable insights into the phenomena of ISM rainfall.

Appendix A

Data Sets Used

" Data! Data! Data! " he cried impatiently. "I can't make bricks without clay." †

A.1 Rainfall

A.1.1. APHRO-V01003R1

This data set is gridded daily rainfall data from 1951 to 2007, developed as part of the project – Asian Rainfall Highly Resolved Observational Data Integration Towards the Evaluation of Water Resources(APHRODITE) [*Yatagai et al. (2009)*]. It is freely available from the website — http://www. chikyu.ac.jp/precip/. I have extracted the data for the South Asian region (see Fig. 2.3) with a horizontal resolution of 0.5 degree ($\sim$ 55km) (APHRO-V1003R1). I have referred to this data set as APHRO-V01003R1 in the thesis.

A.1.2. APHRO-V0902

This data set is the older version of the data set listed above in A.1.1. In time it extended only from 1961-2004 and was also developed as part of the project – Asian Precipitation Highly Resolved Observational Data Integration Towards the Evaluation of Water Resources(APHRODITE) [*Yatagai et al. (2009)*]. It is freely downloadable from the website – http://www.chikyu.ac.jp/precip/. I have extracted the data for the South Asian region from the 0.5 degree resolution data set (APHRO-V0902) for monsoon Asia. I have referred this data set as APHRO-V0902 in the thesis. See Fig.4.3 (A) for example.

† Stated by the character *Sherlock Holmes* in the detective story *The Adventure of the Copper Beeches* written by the scottish author Sir Arthur Conan Doyle (1859 - 1930)

A.1.3. IMD-D

I have also used high resolution precipitation data set collected and developed by the Indian Metrological Department (IMD,Pune) from 1951-2004 [*Rajeevan et al. (2006)*]. It has a daily, one degree latitude, and one degree longitude resolution. But its spatial coverage includes only the political boundary of India. In the thesis I have referred this data set as IMD-D. In all the above data sets the density of rain gauge stations varies over the whole subcontinent. There are certain regions which donot seem to be well represented, for example one of such regions is the state of Jammu and Kashmir in IMD-D. Where the number of stations are 2 to 4 but the area interpolated is 23 grid points which is roughly equivalent to 2500 km^2. Hence I have tried to exclude this region in our analysis in IMD-D. For the description of the interpolation method, the exact locations of rain gauge stations, and the quality of raw data used in IMD-D see the references *Rajeevan et al. (2006)*, and *Rajeevan et al. (2005)*.

A.2 Wind

I have employed the zonal (u) and meridional (v) wind components from the NCEP/NCAR reanalysis data set from 1951 to 2004 with 2.5 degree resolution, provided by the NOAA and available at http://www.esrl.noaa.gov/psd/. To get the correct time indices of the extreme events, I have also used the daily precipitation data from the same source.

A.3 Topography

For the topography I have used ETOPO1- the 1 arc minute gridded global relief model of Earth's surface that integrates land topography and ocean bathymetry. It is provided by NOAA. For further description see http://www.ngdc.noaa.gov/mgg/global/global.html

A.4 Palaeoclimate records

A.4.1. Orbital time scales

A.4.2. Lithogenic grain size (LGS)

This data comes from the Ocean drilling programme site 722 in the Arabian Sea. It is a proxy of wind intensity (carrying capacity) over geological time scales. The data is publicly available and originally published in *Clemens et al. (1996)*. Suggested data citation from the ftp server: Clemens, S.C., et al. 2005. Arabian Sea Summer Monsoon Proxy Data for the Past 3.5 Myr. IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series # 2005-082. NOAA/NCDC Paleoclimatology Program, Boulder CO, USA.

A.4.3. Global ice volume: δ^{18}O

The data comes from the Ocean drilling programme site 659 from North Atlantic (18 deg 05′N, 21°02′W, 3070 m water depth). The data is originally published in *Tiedemann et al. (1994)*. And available at PANGAEA at the address : http://doi.pangaea.de/10.1594/PANGAEA.696121 Also see the reference on the ftp of server NOAA/NGDC Paleoclimatology Program: Raymo, 1997, Major Climate Terminations Data, IGBP PAGES/World Data Center-A for Paleoclimatology, Data Contribution Series # 97-024,NOAA/NGDC Paleoclimatology Program, Boulder CO, USA.

Appendix B

Trends in ISM rainfall

Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.

– H.G.Wells

B.1 Trends in IMD-D data set

In this appendix I will present some of the supplementary results dealing with the issue of changes that are emerging in the spatio-temporal behaviour of monsoonal precipitation over India. This type of study has high significance in the light of increasing trend in global surface temperature and changing characteristics of precipitation at the global scale due to atmosphere warming [*Allan and Soden (2008)*, *Trenberth et al. (2003)*]. The data set used for this analysis is IMD-D from 1951-2006. Please refer to the Appendix A and Sec. A.1.3 for further description.

B.2 Trends in mean annual rainfall and Variance

First, we will take a look at the trends that exist in the mean annual rainfall and its variance during the ISM. The *Mann-Kendall Test* is used for identifying trends and their statistical significance [*Kendall (1955),Mann (1945)*]. The description of *Mann-Kendall Test* is as follows :- Let say R_{ik} is the total precipitation received at grid point k in the year i and i = 1,2...,n . Considering a pair R_{ik} and R_{jk} we first inspect to find out whether $R_{ik} > R_{jk}$ or $R_{ik} < R_{jk}$. Representing number of years for which $R_{ik} > R_{jk}$ by a_k and number of years for $R_{ik} < R_{jk}$ by b_k we then define

$$S_k = a_k - b_k \tag{B.1}$$

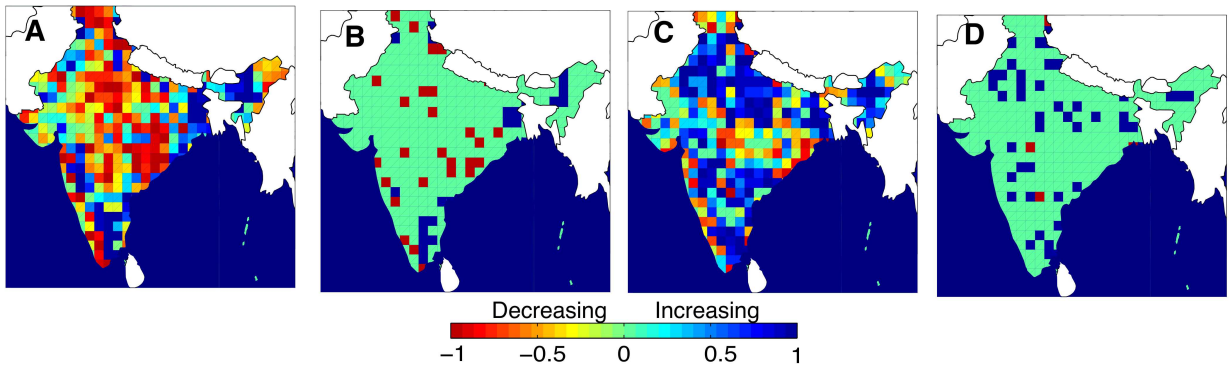

Figure B.1 (A) Linear trends in the mean annual rainfall from 1951-2006 , the colour bar represents the probability of accepting the null hypothesis. The negative values are decreasing trend and positive values are increasing. (B) I have shown only those grid points which satisfy probability criterion of the significance test , red points are decreasing trends and blue points are increasing trends. (C) Trend in the variance of annual rainfall from 1951-2006. Colour bar is the same as for (A). (D) I have shown those grid points which satisfy probability criterion of the significance test, red points are decreasing trends and blue points are increasing trends for the annual variance.

and then quantity Z_k follows a normal distribution, where it is defined as,

$$Z_k = \begin{cases} (S_k - 1)/\sigma_{s_k} & if \quad S_k > 0 \\ 0 & if \quad S_k = 0 \\ (S_k + 1)/\sigma_{s_k} & if \quad S_k < 0 \end{cases} \tag{B.2}$$

$$\sigma_{s_k} = \sqrt{\frac{n(n-1)(2n+5)}{18}}$$

The significance of trends is tested by comparing Z_k with standard normal variate z at the

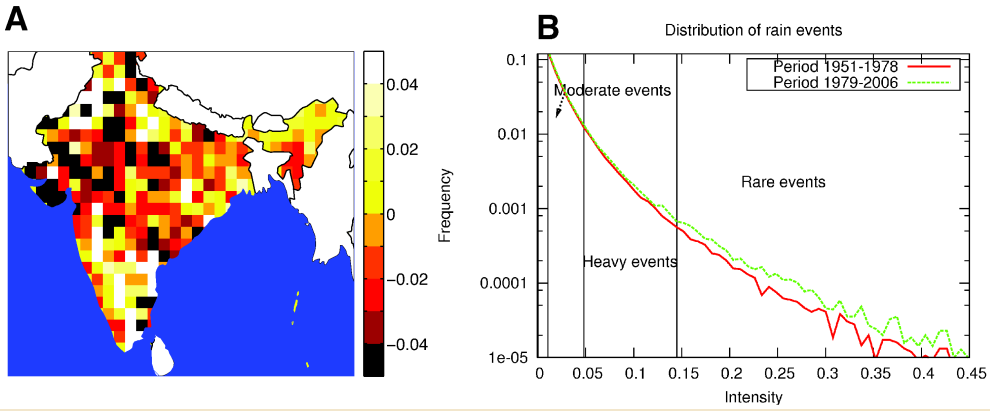

Figure B.2 (A) Relative entropy change over the two halves of the data. Observe the changes in parts of central India and north west India. (B) Distribution of daily rainfall intensity for the whole land mass for the two halves of the data set. Note the clustering of rare events (extreme events) in the second half of the data set.

desired significance level. Therefore the null hypothesis H_0, that there is a trend, is accepted if $\alpha_k = 2P(Z_k \geq |z|)$ is less than level of desired significance. The direction of trend is given by the sign of Z_k.

In the Fig B.1 (A) we plot probability $p_k = 1 - \alpha_k$ for $Z_k > 0$ and $p_k = \alpha_k - 1$ for $Z_k < 0$. Here the desired significance level is $\alpha = 0.05$ which means the we accept the H_0 if $p_k \geq 0.95$ for increasing trend or $p_k \leq -0.95$ for decreasing trend. One of the feature to note in Fig B.1 (A) is that majority of the grid points show a decreasing trend and few of them show statistical significant trend as well (See Fig B.1 (B)). To find out the trends in variance we calculate variance over window size of 5 years with three 3 years overlap. We use the same Mann-Kendall test as described above. The result is presented in the Fig B.1 (C-D). Here we observe that majority of grid points have a trend of increasing variance, many among them are significant (Fig B.1 (D)). Indicating that annual rainfall is becoming more variable. As we observe only small number of grid points showing a significance trend in the mean rainfall, therefore it seems that monsoonal rainfall is stable, at least at the annual scale, for the time period considered in the above data set.

B.3 The changing characteristics of rainfall events

Next, I will attempt to extract any feature that may have appeared over the last decades in the rainfall events, especially in regard to the extreme events, as they have enormous societal influences.

B.3.1. Entropy change of the rain event distribution

In this section I present a calculation of relative entropy change over the two halves of the data. For this purpose we take precipitation annual anomaly values -

$$z_{ij} = \frac{P_{ij} - \mu_j}{\sigma_{P_{ij}}} \tag{B.3}$$

where $\mu_j = \frac{1}{N}\sum_{i=1}^{N} P_{ij}$, $\sigma_{P_{ij}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_{ij} - \mu)^2}$ and $N = 122$ days corresponding to four monsoon months. P_{ij} is the net rainfall on i^{th}th day and j^{th} location. Now the Shannon entropy [*Press et al. (1992)*]for the distribution p_i is defined as

$$H(z) = \sum_{i=1}^{n} p_i \ln p_i. \tag{B.4}$$

The edges of the bins for the distribution are for the calculation of entropy are calculated using the whole 56 year data. Let the entropy for first half is H_1 and second half is H_2.We plot the quantity $\Delta H = \frac{H_2 - H_1}{H_1}$ i.e., relative change in entropy, below in Fig B.2 (A). We observe that large parts in central India and north west India show a decrease in entropy: which could be

possible in only one scenario that there is some kind of preference for certain rainfall events to occur in the second half of the data. We study this question in further detail below.

B.3.2. Rainfall events

A very important aspect to study are the characteristics of individual rainfall events, meaning their intensity and frequency. In a recent study with the same data set [*Goswami et al. (2006)*), it has been shown that extreme events are on increase over central India. However, a pre-defined threshold on net rainfall events was used in this study. I present similar findings, but it seems that this phenomena has far more complex spatial features. I try to use two different approaches to understand and visualise the above mentioned phenomena. Both the approaches are entirely different from *Goswami et al. (2006)*. The first one is based on choosing threshold on intensity of rainfall. Intensity of a rainfall event is relative to the place, for example a heavy rainfall event will be different for a place receiving more than 2000 mm of rainfall to the place which receives 100 mm of rainfall during a monsoon season. Keeping this in mind I define the intensity I_{ij} of a rainfall event at a particular location (grid point) j and i^{th} day as follows,

$$I_{ij} = \frac{R_{ij}}{M_j} \tag{B.5}$$

where R_{ij} is rainfall on i^{th} day at j^{th} grid point and M_j is the mean summer monsoon (JJAS) rainfall at j^{th} grid point. On the basis of frequency distribution of I_{ij} I define the following three types of rainfall events i.e., the top 0.5% as *rare rainfall events (or extreme rainfall events)*, the next 5% as *heavy rainfall events* and the next 30% as *moderate rainfall events*. In the Fig. B.2 (B), I have shown the result for the whole land mass by mixing data from all the grid points. In the Fig. B.2 (B), we could clearly observe the clustering at the end of the distribution, indicating a increase in heavy and extreme rainfall events.

A second approach could be to use quantile of net rainfall as thresholds. Then use trend analysis to evaluate the statistical significance of the change. For example an extreme or a rare rainfall event could be considered to be the 99% quantile of all the rainfall events in the 56 year at a particular grid point. Then I count number of such events happening over a period of years, I have taken this period equal to 4 years. The number of events happening in this period say is equal to N_{ij}^k, where $k = 1,2,..,14$. I use Mann-Kendall test for trend detection and significance on N_{ij}^k. The result from such approach is present in Fig B.3. We also repeat this exercise for heavy events 95% quantile and also for more moderate rain events at 50% quantile and 65% quantile. We see in Fig B.3 that there is considerable decrease in smaller rainfall events and it is also statistically significant. It is important to note that these events are major source of precipitation during the monsoonal season. We also find that there is increase in heavy and extreme events but there are very small regions where it is significant. Thus, I propose that the changes in monsoonal precipitation is not homogenous over the whole of India. This may be due to finer scale processes effecting the monsoonal rainfall, such as in

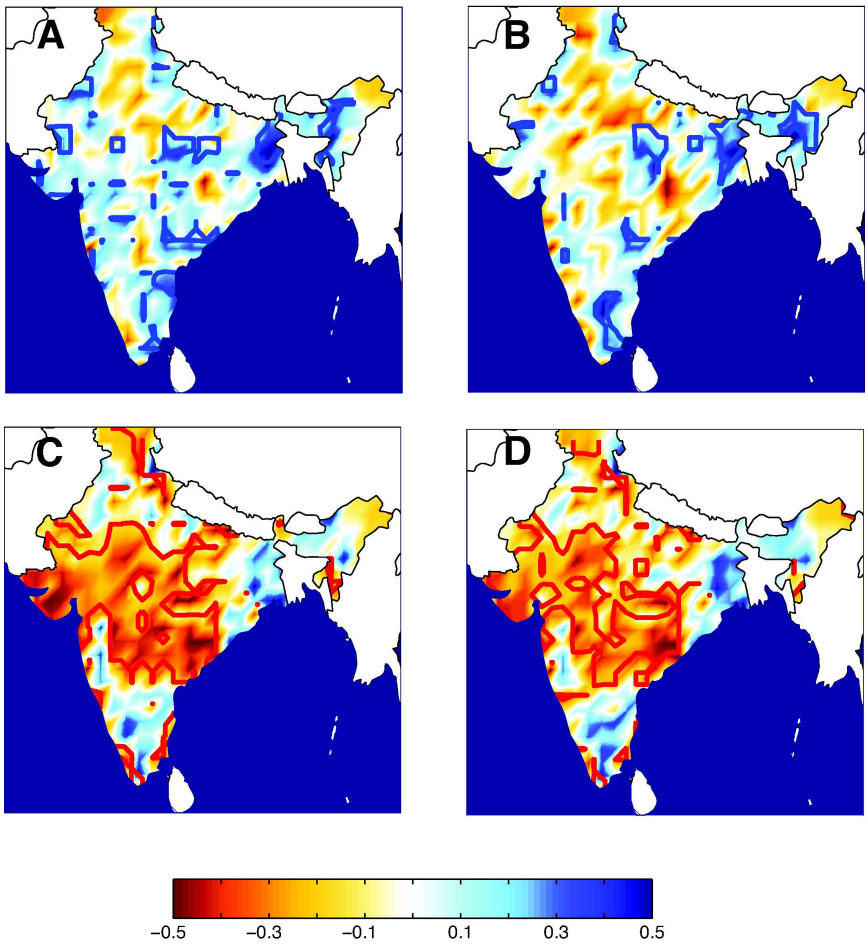

Figure B.3 The colour gives the direction of trend ($\tau = \frac{2S_k}{n(n-1)}$ in the Mann-Kendall test), the curved lines gives the significant regions. The colour of the curved line corresponds to the colour bar. (A) Extreme rainfall events 99% quantile, (B) Heavy Rainfall Events, 95% quantile, (C) Moderate Rainfall Events,50% quantile (D) Moderate Rainfall Events,65% quantile

homogenous spread of aerosols.

B.3.3. Droughts

The above results compel us to look into another extreme weather condition viz, droughts. As monsoon is the main source of precipitation, availability of water in India is directly dependent on monsoonal precipitation. Thus, it is also of high interest to see whether the frequency of droughts over the period concerned in this data set has changed. For this purpose I define a condition at a particular grid point as drought, if the total JJAS rainfall is below 25% the long term mean. Now again I divide the data into two halves . Let say ${Y_d}^1$ is the number of years below the threshold in the first half and ${Y_d}^2$ in the second half. Then change in mean recurrent time of droughts over these two halves will be,

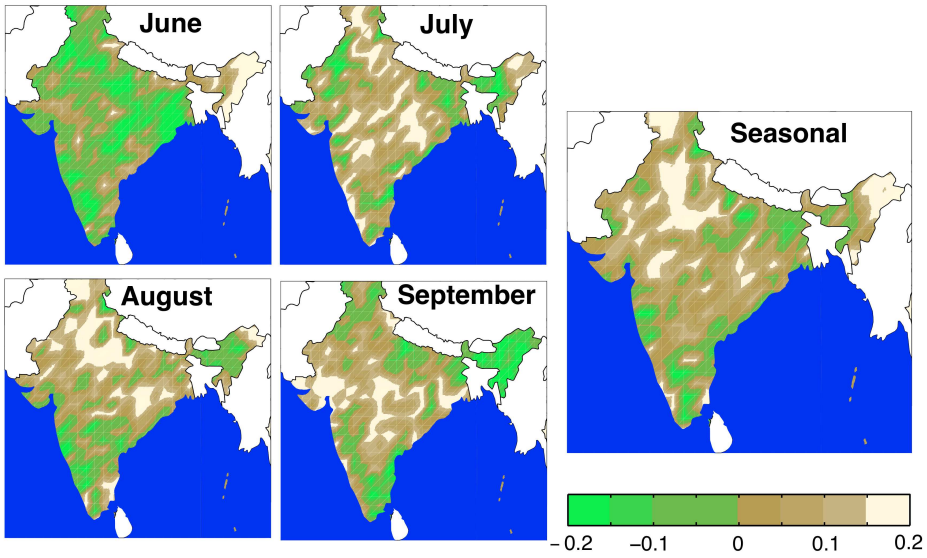

Figure B.4 The recurrence of droughts: the colour here represents the change in mean recurrent time and the colourbar is the same for all the figures. The month is indicated in the figure itself. Positive values indicate higher recurrences of droughts and negative values the reverse case.

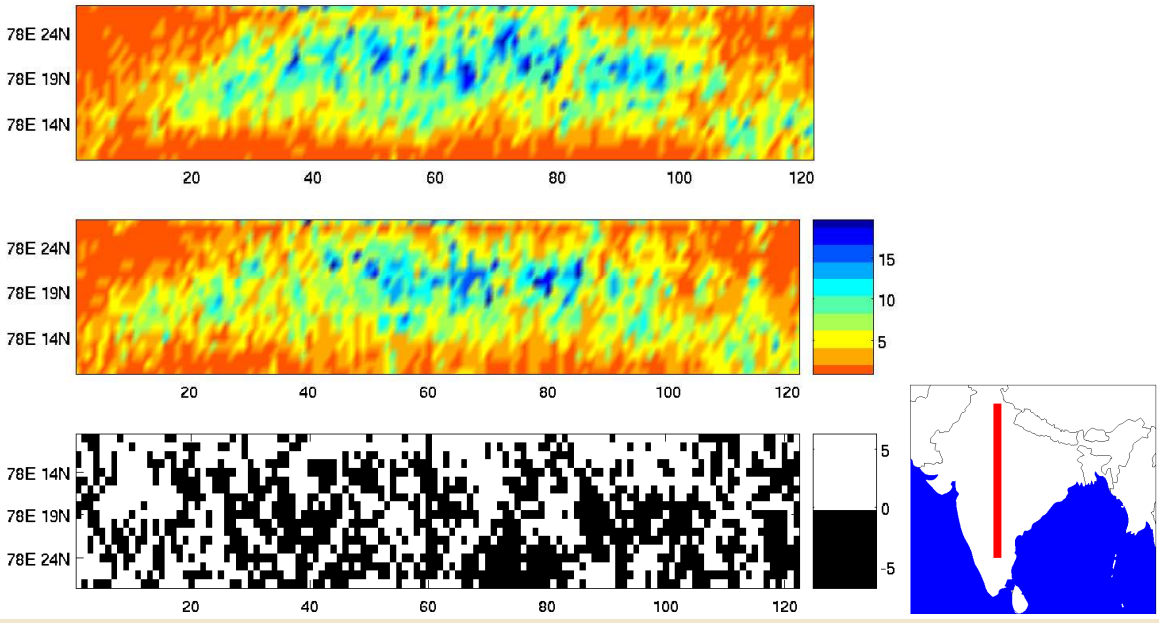

Figure B.5 The pattern of daily mean rainfall along the latitudinal strip shown in the map. The unit of colour bar is mm/day. The top curve is the first half, the middle the second half and bottom the difference.

$$\Delta Y = \frac{1}{Y}(Y_d^2 - Y_d^1) \tag{B.6}$$

where Y is total number of years in each half, so here $Y = 28$ and from Eq. B.6 we see that $\Delta Y \in [-1,1]$. Thus, a negative value of ΔY will imply that there is decrease in the incidences of drought whereas positive values will indicate that there is an increase in the incidences of drought. In the Fig. B.4 we observer that large parts India show a increase in recurrence of droughts . I repeated this exercise for monthly precipitation as well. The result for that too are presented in Fig. B.4. An interesting result is obtained for July and September, the months of monsoon progress and retreat. Large parts of west India, central India and to some extent the south received far lesser rainfall during these months. It seems that both the progress and the retreat of monsoon are taking place in an abnormal way. To look into this question I try another exercise . I take grid points in a strip extending from $14^{\circ}N$ to $31^{\circ}N$ at $78^{\circ}E$ (see small map in Fig B.5). Calculating daily mean over 28 years for the JJAS i.e., $X_{ij}^1 = \sum_{l=1}^{28} \frac{R_{ij}^l}{28}$ and $X_{ij}^2 = \sum_{l=29}^{56} \frac{R_{ij}^l}{28}$, where R_{ij}^l is rainfall at at a particular location (grid point) j^{th} and i^{th} day in l^{th} year. X_{ij}^1 is the daily mean of the first half and X_{ij}^2 for second half (top curve in Fig B.5 is the first half and middle curve is the second half). Then I plot this mean in colour on plot with days in the X-axis and location of the grid point on the y-axis. Visual inspection of these curves doesn't show any significant evolution in the patterns. But if we see the difference between the two curves obtained by subtracting first half from the second, $\Delta X = X_{ij}^2 - X_{ij}^1$. We observe many black point indicating ($X_{ij}^1 > X_{ij}^2$) indicating decrease in rainfall. This analysis reflects that total annual rainfall may be statistically stable but occurrence of below normal monsoon years seems to have increased over this period.

B.4 Summary

In summation, in this appendix I have presented an analysis of high resolution rainfall gridded data set from 1951 to 2006 for India that discerns the emerging characteristics in the spatio-temporal behaviour of monsoonal precipitation in the warming atmosphere. We observe that changes have extremely complex spatial features and it seems that monsoon is showing inhomogenous changes in space. A preference for certain type of rain events is also observed which probably is manifesting in more pronounced drought conditions. All this possibly means that small scale processes are influencing monsoon more rather than some large scale drastic change like breakdown or weakening of monsoonal circulation.

Bibliography

R. Agnihotri, K. Dutta, R. Bhushan, and B. L. K. Somayajulu. Evidence for solar forcing on the indian monsoon during the last millennium. *Earth and Planetary Science Letters*, 198: 521–527, 2002.

R. Albert and A. L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74: 47, 2002.

S. Albeverio, V. Jentsch, and H. Kantz, editors. *Extreme Events in Nature and Society*. Springer, 2006.

Richard P. Allan and Brian J. Soden. Atmospheric warming and the amplification of precipitation extremes. *Science*, 321:1481–1484, 2008.

M. A. Altabet, M. J. Higginson, and D.W Murray. The effect of millennial-scale changes in arabian sea denitrification on atmo- spheric co2. *Nature*, 415:159–162, 2002.

A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. S. Zhou. Synchronization in complex networks. *Phys. Rep.*, 469(3):93–153, 2008.

J. Arnhold, P. Grassberger, K. Lehnertz, and C.E. Elger. A robust method for decting interdependences: application to intracranially recorded eeg. *Physica D*, 134:419–430, 1999.

K. Ashok, Z. Guan, and T. Yamagata. Impact of the indian ocean dipole on the relationship between the indian monsoon rainfall and enso. *Geophys. Res. Lett.*, 113:4499–4502, 2001.

J. Beer, W. Mende, and R. Stellmacher. The role of the sun in climate forcing. *Quat. Sci. Rev.*, 19:403–415, 2000.

A. Berger and M.F. Loutre. Insolation values for the climate of the last 10 million years. *Quat. Sci. Rev.*, 10(4):297–317, 1991.

W. H. Berger and U. von Rad. Decadal to millennial cyclicity in varves and turbidites from the arabian sea: hypothesis of tidal origin. *Global and Planetary Change*, 34:313–325, 2002.

U. V. Bhide, R. Muzamdar, S. P. Ghanekar, D. K. Paul, T. C. Chen, and G. V. Rao. A diagnostic study on heat sources and moisture sinks in the monsoon trough area during active break phase of the indian monsoon of 1979. *Tellus, Ser. A - Dyn. Meteorol. Oceanol.*, 49:455–473, 1997.

S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. Complex networks: Structure and dynamics. *Phys. Rep.*, 424(4-5):175–308, 2006.

B. Bookhagen. Appearance of extreme monsoonal rainfall events and their impact on erosion in the himalaya. *Geomatics, Natural Hazards and Risk*, 1:37–50, 2010.

B. Bookhagen and D. W. Burbank. Topography,relief, and trmm-derived rainfall variation along the himalaya. *Geophys. Res. Lett.*, 33:L08405, 2006.

B. Bookhagen and D. W. Burbank. Toward a complete himalayan hydrological budget: Spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. *J. Geophys. Res.*, 115:F03019, 2010.

D. L. Cadet and B. C. Diehl. Interannual variability of surface field over the indian ocean during recent decades. *Mon. Weather Rev.*, 112:1921, 1984.

M. A Cane and P. Molnar. Closing of the indonesian seaway as a precursor to east african aridification around 304 million years ago. *Nature*, 411:157–162, 2001.

M.A. Cane and A.C. Clement. A role for the tropical pacific coupled ocean-atmosphere system on milankovich and millenial timescales. part ii: Global impacts. In P. U. Clark, R. S. Webb, and L. D. Keigwin, editors, *Mechanisms of Global Climate Change at Millennial Time Scales*, pages 373–383. American Geophysical Union, Washington DC, 1999.

B. K. Chakrabarti and M. Acharyya. *Rev. Mod. Phys.*, 71:847, 1999.

C. D. Charles, D. E. Hunter, and R. G. Fairbanks. Interaction between the enso and the asian monsoon in a coral record of tropical climate. *Science*, 277:925–928, 1997.

P. Chatterjee and B. N. Goswami. Structure, genesis and scale selection of the tropical quasi-biweekly mode. *Quat. J. Roy. Meteorol. Soc.*, 130:1171–1194, 2004.

C. O. Clark, J. E. Cole, and P. J. Webster. Indian ocean sst and indian summer rainfall: Predictive relationship and their decadal variability. *J. Climate*, 13:2503–2519, 2000.

S. Clemens. Extending the historical record by proxy. In B. Wang, editor, *The Asian Monsoon*. Springer, 2006.

Steven Clemens, Warren Prell, David Murray, Graham Shimmield, and Graham Weedon. Forcing mechanisms of the indian ocean monsoon. *Nature*, 353:1991, 1991.

Steven C. Clemens, David W. Murray, and Warren L. Prell. Nonstationary phase of the plio-pleistocene asian monsoon. *Science*, 274:943–947, 1996.

R. Cohen and S. Havlin, editors. *Complex Networks: Structure, Robustness and Function*. Cambridge, 2010.

S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, 2001.

S. Coles, J. Heffernan, and J. Tawn. Dependence measures for extreme value analyses. *Extremes*, 2:4:339–365, 1999.

A. Dai and T. M. L Wigley. Global patterns of enso induced precipitation. *Geophys. Res. Lett.*, 27:1283–1286, 2000.

C. Dangalchev. Residual closeness in networks. *Phisica A*, 365(2):556–564, 2006.

W. Dansgaard, S. J Johnsen, H. B. Clausen, D. Dahl-Jensen, N. Gundestrup, C. U Hammer, C. S. Hvidberg, J. P. Steffensen, A. E. Sveinbjörndottir, J. Jouzel, and G. Bond. Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature*, 364:218–220, 1993.

A C Davison and D Hinkley. *Bootstrap Methods and their Application*. Cambridge, 8th edition, 2006.

Peter B. deMenocal. Plio-pleistocene african climate. *Science*, 270:53–58, 1995.

Peter B. deMenocal. African climate change and faunal evolution during the pliocene-pleistocene. *Earth Plan. Sci. Lett.*, 220:3–24, 2004.

A. P. Dimri. Impact of horizontal model resolution and orography on the simulation of a western disturbance and its associated precipitation. *Meteorological Applications*, 11:115–127, 2004.

Y. Ding and D. R. Sikka. Synoptic systems and weather. In B. Wang, editor, *The Asian Monsoon*. Springer, 2006.

J. F. Donges, Y. Zou, N. Marwan, and J. Kurths. The backbone of the climate network. *Europhys. Lett.*, 87(4):48007, 2009.

J. F. Donges, H. C. H. Schultz, N. Marwan, Y. Zou, and J. Kurths. Investigating the topology of interacting networks - theory and application to coupled climate subnetworks. *Eur. Phys. J. B*, 2011. doi: 10.1140/epjb/e2011-10899-1.

S. S Dugam, S. B. Kakade, and R. K. Verma. Interannual and long-term variability in the north atlantic oscillation and indian summer monsoon rainfall. *Theor. Appl. Climatol.*, 58:21–29, 1997.

J. D. Farmer, E. Ott, and J. A. Yorke. The dimension of chaotic attractors. *Physica D*, 7(1-3): 153–180, 1983.

J. Fasullo and P. J. Webster. Hydrological signatures relating the asian summer monsoon and enso. *J. Climate*, 15:3085–3095, 2002.

J. Fasullo and P. J. Webster. A hydrological definition of indian monsoon onset and withdrawal. *J. Climate*, 16:3200–3211, 2003.

S. Gadgil. The indian monsoon and its variability. *Annu. Rev. Earth and Planet. Sci.*, 31: 429–67, 2003.

S. Gadgil and K. R. Kumar. The asian monsoon: agriculture and economy. In B. Wang, editor, *The Asian Monsoon*. Springer, 2006.

S. Gadgil, Yadumani, and N.V Joshi. Coherent rainfall zones of the indian region. *International Journal of Climatology*, 13:547–566, 1993.

A. Ganopolski and S. Rahmstorf. Abrupt glacial climate changes due to stochastic resonance. *Phys. Rev. Lett.*, 88(3):038501–1–038501–4, 2001.

J. B. Gao. Recurrence time statistics for chaotic systems and their applications. *Phys. Rev. Lett.*, 83(16):3178–3181, Oct 1999. doi: 10.1103/PhysRevLett.83.3178.

B. N. Goswami. Interannual variations of indian summer monsoon in a gcm: External conditions versus internal feedbacks. *J. Climate*, 11:501–522, 1998.

B. N. Goswami, R. S. Ajayamohan, Prince K. Xavier, and D. Sengupta. Clustering of synoptic activity by indian summer monsoon interseasonal oscillations. *Geophys. Res. Lett.*, 30:1431, 2003.

B. N. Goswami, V. Venugopal, D. Sengupta, M. S. Madhusoodanan, and Prince K. Xavier. Increasing trend of extreme rain events over india in a warming environment. *Science*, 314: 1442 – 1445, 2006.

P. Grassberger, R. Badii, and A. Politi. Scaling laws for invariant measures on hyperbolic and nonhyperbolic attractors. *Jour. of Stats. Phys.*, 51(1/2):135, 1988.

P. Y. Groisman, T. R. Karl, D. R. Easterling, R. W. Knight, P. F. Jamason, K. J. Hennessy, R. Suppiah, C. M. Page, J. Wibig, and K. Fortuniak et al. Changes in the probability of heavy precipitation: Important indicators of climatic change. *Climate Change*, 42(1):243–283, 1999.

A. K. Gupta, D. M. Anderson, and J. T. Overpeck. Abrupt changes in the asian southwest monsoon during the holocene and their links to the north atlantic ocean. *Nature*, 421: 354–357, 2003.

J.M. Gutiérrez, A.Galván, A.S Cofin no, and C.Primo. Chaos game characterization of temporal precipitation variability : Application to regionalization. *Fractals*, 14:87–99, 2006.

S. Hastenrath. *Climate Dynamics of the Tropics: An Updated Edition of Climate and Circulation of the Tropics*. Kluwer Academic, Dordrect, The Netherlands, 1987.

Rainer Hegger, Holger Kantz, Lorenzo Matassini, and Thomas Schreiber. Coping with nonstationarity by overembedding. *Phys. Rev. Lett.*, 84(18):4092–4095, 2000.

C. C. Hong, H. H. Hsu, N. H. Lin, and H. Chi. Roles of european blocking and tropical ? extratropical interaction in the 2010 pakistan flooding. *Geophys. Res. Lett.*, 38:L13806, 2011.

R. A. Houze, D. C. Wilton, and B. F. Smull. Monsoon convection in the himalayan region as seen by the trmm precipitation radar. *Q. J. R. Meteorol. Soc.*, 133:1389 –1411, 2007.

R.N. Iyengar and P. Basak. Regionalization of indian monsoon rainfall and long term variability signals. *International Journal of Climatology*, 14:1095–1114, 1994.

A Jain and R Dubes. *Algorithms for Clustering Data*. NJ: Prentice-Hall, 1988.

R. H. Johnson. Mesoscale processes. In B. Wang, editor, *The Asian Monsoon*. Springer, 2006.

H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, second edition, 2004.

M.G Kendall. *Rank Correlation Methods*. Griffin, 1955.

R. N. Keshavamurthy. Power-spectra of large scale disturbances of the indian south-west monsoon. *Indian J. Meteorol. Geophys.*, 24:117–124, 1973.

R. H. Kripalani and A. Kulkarni. Rainfall variability over south-east asia - connections with indian monsoon and enso extremes: New perspectives. *Int. J. Climatology*, 17(11):1155–1168, 1999.

V. Krishnamurthy and B. N. Goswami. Indian monsoon-enso relationship on interdecadal timescale. *J. Climate*, 13:579–595, 2000.

T. N. Krishnamurti and P. Ardunay. The 10 to 20 day westward propagating mode and "breaks in the monsoons". *Tellus*, 32:15–26, 1980.

V. Krishnmurthy and J. Shukla. Interseasonal and inter annual variability of rainfall over india. *J. Clim*, 13:4366–4377, 2000.

D. Kroon and G. Ganssen. Northern indian ocean upwelling cells and stable isotope composition of living planktonic foraminifera. *Deep-Sea Research I*, 36:1219–1236, 1989.

K. K. Kumar, B. Rajagopalan, and M. Cane. On the weakening relationship between the indian monsoon and enso. *Science*, 284(5423):2156–2159, 1999.

K. K. Kumar, B. Rajagopalan, M. Hoerling, G. Bates, and M. Cane. Unraveling the mystery of indian monsoon failure during el niño. *Science*, 314(5796):115–119, 2006.

J. E. Kutzbach, W. L. Prell, and W. F. Ruddiman. Sensitivity of eurasian climate to surface uplift of the tibetan plateau. *J. of Geol.*, 101:177–190, 1993.

Ying-Cheng Lai. Transition from strange nonchaotic to strange chaotic attractors. *Phys. Rev. E*, 53(1):57–65, 1996.

M. Lal, L. Bengtsson, U. Cubash, M. Esch, and U. Schlese. Synoptic scale disturbances of the indian summer monsoon as simulated in a high resolution climate model. *Clim. Res.*, 5: 243–258, 1995.

J. C. Larrasoaña, A. P. Roberts, E. J. Rohling, M.Winklhofer, and R. Wehausen. Three million years of monsoon variability over the northern sahara. *Clim. Dyn.*, 21:689–698, 2003.

J. Laskar. The limits of earth orbital calculations for geological time-scale use. *Philos. T. R. Soc. Lon. A*, 357:1735–1759, 1999.

J. Laskar, F. Jouzel, and F. Boudin. Orbital, precessional, and insolation quantities for the earth from 20 myr to +10 myr. *Astron. Astrophys*, 270:522–533, 1993.

J. Laskar, M. Gastineau, F. Joutel, P. Robutel, B. Levrard, and A. Correria. A long term numerical solution for the insolation quantities of earth. *Astron. and Astrophys.*, 428:261–285, 2004.

D. M. Lawrence and P. J. Webster. The boreal summer interseasonal oscillation: Relation between northward and eastward movement of convection. *J. Atm. Sci.*, 59:1593–1606, 2002.

K. Lehnertz and C. E. Elger. *Phys. Rev. Lett.*, 80:5019, 1998.

T. M. Lenton. Early warning of climate tipping points. *Nature Clim. Change*, 1:201–209, 2011.

T. M. Lenton, H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber. Tipping elements in the earth's climate system. *PNAS*, 105(6):1786 –1793, 2008.

A. Levermann, J. Schewe, V. Petoukhov, and H Held. Basic mechanism for abrupt monsoon transitions. *PNAS*, 106(49):20572–20577, 2009.

W. T. Liu, A. Zhang, and J. K. B. Bishop. Evaporation and solar irradiance as regulators of sea surface temperature in annual and interannual changes. *J. Geo. Res.*, 99(C6):12623–12637, 1994.

R. A Madden and P. R. Julian. Detection of a 40-50 day oscillation in the zonal wind in the tropical pacific. *J. Atmos. Sci.*, 28:702–708, 1971.

N. Malik, N. Marwan, and J. Kurths. Spatial structures and directionalities in monsoonal precipitation over south asia. *Nonlin. Proc. Geophy.*, 17:371–381, 2010.

N. Malik, B. Bookhagen, N. Marwan, and J. Kurths. Analysis of spatial and temporal extreme monsoonal rainfall over south asia using complex networks. *Clim. Dyn.*, 2011a. doi: DOI: 10.1007/s00382-011-1156-4.

N. Malik, Y. Zou, N. Marwan, and J. Kurths. Spatial structures and directionalities in monsoonal precipitation over south asia. *Submitted to Europhys. Lett.*, 2011b.

H.B Mann. Nonparametric tests against trend. *Econometrica*, 13:245–259, 1945.

R. N. Mantegna and H. E. Stanley. *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, 2000.

D. Maraun and J. Kurths. Epochs of phase coherence between el niño southern oscillation and indian monsoon. *Geophys. Res. Lett.*, 32:L15709, 2005.

N. Marwan, J. F. Donges, Y. Zou, R. V. Donner, and J. Kurths. *Physics Letters A*, 373:4246, 2009.

W. May. Simulation of the variability and extremes of daily rainfall during the indian summer monsoon for present and future times in a global time-slice experiment. *Clim. Dyn.*, 22: 183–204, 2004a.

W. May. Simulation of the variability and extremes of daily rainfall during the indian summer monsoon for present and future times in a global time-slice experiment. *Clim. Dyn.*, 22: 183–204, 2004b.

I. I. Mokhov, D. A. Smirnov, P. I. Nakonechny, S. S. Kozlenko, E. P. Seleznev, and J. Kurths. Alternating mutual influence of el-niño/southern oscillation and indian monsoon. *Geophys. Res. Lett.*, 38:L00F04, 2011.

D. A. Mooley and J. Shukla. Main features of the westwerd moving low pressure systems which form over the indian region which form over the indian region during the monsoon season and their relationship with monsoonal rainfall. *Mausam*, 40:137–152, 1989.

U. Neff, S. J. Burns, A. Mangini, M. Mudelsee, D. Fleitmann, and A. Matter. Strong coherence between solar variability and the monsoon in oman between 9 and 6 kyr ago. *Nature*, 411: 290–293, 2001.

M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45(2):167–256, 2003.

E. M. Oblow. Supertracks, supertrack functions and chaos in the quadratic map. *Phys. Lett. A*, 128(8):406 – 412, 1988.

E. Ott. *Chaos in Dynamical Systems*. Cambridge University Press, Cambridge, second edition, 2002.

M. Palus. *Phys. Rev. Lett.*, 101:134101, 2008.

T. S Parker and L. Chua. *Practical Numerical Algorithms for Chaotic Systems*. Springer-Verlag, 1989.

B. Parthasarathy, A. A. Munot, and D. R. Kothawale. Monthly and seasonal rainfall series for all-india homogeneous regions and meteorological subdivisions: 1871-1994. *Contributions from Indian Institute of Tropical Meteorolgy, Research Report RR-065, Aug. 1995, Pune 411 008 INDIA.*, 1995.

Francesco S. R. Pausata, David S. Battisti, Kerim H. Nisancioglu, and Cecilia M. Bitz. Chinese stalagmite $\delta^{18}o$ controlled by changes in the indian monsoon during a simulated heinrich event. *Nature Geosci.*, 4:474–480, 2011. doi: 10.1038/ngeo1169.

K. W Pettis, T. A Bailey, A. K Jain, and R. C Dubes. An intrinsic dimensionality estimator from near-neighbor information. *IEEE Transactions on Pattern Analysis and Machine intelligence*, PAM-1(1):25–37, 1979.

A. Prasad and R. Ramaswamy. Characteristic distribution of finite-time lyapunov exponent. *Phys. Rev. E*, 60(3):2761, 1999.

W.L. Prell and J.E Kutzbach. Sensitivity of the indian monsoon to forcing parameters and implications for its evolution. *Nature*, 360:647–653, 1992.

William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C : The Art of Scientific Computing*. Cambridge University Press, 1992.

J. Quade, T. E. Cerling, and J. E. Bowman. Development of asian monsoon revealed by marked ecological shift during the latest miocene in northern pakistan. *Nature*, 342:163–166, 1989.

R. Quian Quiroga, T. Kreuz, and P. Grassberger. Event synchronization: A simple and fast method to measure synchronity and time delay patterns. *Phys. Rev. E*, 66:041904–1, 2002.

S. Rahmstorf. Timing of abrupt climate change: A precise clock. *Geophys. Res. Lett.*, 30: 17–1–17–4, 2003.

M. Rajeevan, J. Bhate, J.D Kale, and B. Lal. Development of a high resolution daily gridded rainfall data for the indian region. *Met. Monograph Climatology No. 22/2005*, 2005. Indian Meteorological Department, Pune, India.

M. Rajeevan, J. Bhate, J.D Kale, and B. Lal. High resolution daily gridded data for indian region: Analysis of break and active monsoon spells. *Curr. Sci.*, 9:296–306, 2006.

M. Rajeevan, J. Bhate, and A. K. Jaswal. Analysis of variability and trends of extreme rainfall events over india using 104 years of gridded daily rainfall data. *Geo. Res. Lett.*, 35:L18707, 2008.

V. Ramanathan, C. Chung, D. Kim, T. Bettge, L. Buja, J. T. Kiehl, W. M. Washington, Q. Fu, D. R. Sikka, and M. Wild. Atmospheric brown clouds: Impact on south asian climate and hydrologic cycle. *PNAS*, 102:5326–5333, 2005.

G. Ramstein, F. Fluteau, J. Besse, and S. Joussaume. Effect of orogeny, plate motion and land-sea distribution on eurasian climate change over the past 30 million years. *Nature*, 386: 788–795, 1997.

A. C. Ravelo. Walker circulation and global warming. *Oceanography*, 19(4):114–122, 2006.

M.E Raymo. The initiation of norther hemisphere glaciation. *Annu. Rev. Earth and Plant. Sci.*, 22:353–383, 1994.

Christoph Rieke, Karsten Sternickel, Ralph G. Andrzejak, Christian E. Elger, Peter David, and Klaus Lehnertz. Measuring nonstationarity by analyzing the loss of recurrence in dynamical systems. *Phys. Rev. Lett.*, 88(24):244102, May 2002.

M. J. Rodwell and B. J. Hoskins. Subtropical anticyclones and summer monsoons. *J. Climate*, 14:3192–3122, 2001.

G. H. Roe. Orographic precipitation. *Annu. Rev. Earth Planet. Sci.*, 33:645–71, 2005.

W. F. Ruddiam. *Earth's Climate Past and Future*. W. H. Freeman Company, 2008.

Marten Scheffer. *Critical Transitions in Nature and Society*. Princeton University Press, 2009.

T. Schreiber. Detecting and analyzing nonstationarity in a time series using nonlinear cross predictions. *Phys. Rev. Lett.*, 78(5):843, 1997.

H. Schulz, U. von Rad, and H. Erlenkeuser. Correlation between arabian sea and greenland climate oscillations of the past 110,000 years. *Nature*, 393:54–57, 1998.

C. Schumacher and R. A. Houze. Stratiform rain in the tropics as seen by the trmm precipitation radar. *J. Climate*, 16:1739–1756, 2003.

M. A Sepúlveda, R. Badii, and E. Pollak. Spectral analysis of conservative dynamical systems. *Phys. Rev. Lett.*, 63(12):1226–1229, 1989.

D. R. Sikka. Some aspects of the life history, structure and movement of monsoon depressions. *Pure Appl. Geophys.*, 115:1501–1529, 1977.

D. R. Sikka and S. Gadgil. On the maximum cloud zone and itcz over indian longitudes during the southwest monsoon. *Mon. Wea. Rev*, 108:1840–1853, 1980.

S. V. Singh, R. H. Kirpalani, and D. R. Sikka. Interannual variability of the madden-julian oscillations in indian summer monsoon rainfall. *J. Clim.*, 5:973–979, 1992.

M. K. Soman and J. Slingo. Sensitivity of asian summer monsoon to aspects of sea surface temperature anomalies in the tropical pacific ocean. *Quart. J. Roy. Meteorol. Soc.*, 123: 309–336, 1997.

M. Steinbach, P. N. Tan, and V. Kumar et al. Discovery of climate indices using clustering. *ACM SIGKDD*, pages 446–455, 2003.

K. Steinhaeuser, N. V Chawla, and A. R. Ganguly. An exploration of climate data using complex networks. *SIGKDD Explorations*, pages 25–32, 2010.

D. B. Stephenson, K. Rupa Kumar, F. J. Doblas-Reyes, J. F. Royer, and F. Chavin. Extreme rainfall events and their impact on ensemble forecasts of indian monsoon. *Mon. Weather Rev.*, 127:1954–1966, 1999.

K. L. Swanson and A. A. Tsonis. Has the climate recently shifted ? *Geophys. Res. Lett.*, 36: L06711, 2009.

R. Tiedemann, M. Sarnthein, and N. J. Shackleton. Astronomic timescale for the pliocene atlantic d18o and dust flux records of odp site 659. *Paleoceanography*, 9:619–638, 1994.

C. Torrence and P. J. Webster. Interdecadal changes in the enso-monsoon system. *J. Climate*, 12:2679–2690, 1999.

Martin H. Trauth, Juan C. Larrsoana, and Manfred Mudelsee. Trends, rhythms and events in plio-peleistocene african climate. *Quat. Sci. Rev.*, 28:394–411, 2009.

Kevin E. Trenberth, Aiguo Dai, Roy M. Rasmussen, and David B. Parsons. The changing character of precipitation. *Bulletin of the American Meteorological Society*, 84:1205–1217, 2003.

A. A. Tsonis and K. L. Swanson. Topology and predictability of el nino and la niña networks. *Phys. Rev. Lett.*, 100(22):228502, 2008.

A. A. Tsonis, K. L. Swanson, and P. J. Roebber. What do networks have to do with climate? *Bull. Am. Meteorol. Society*, 87(5):585+, 2006.

A. A. Tsonis, G. Wang, K. L. Swanson, F. Rodrigues, and L. Costa. Community structure and dynamics in climate networks. *Clim. Dyn.*, 2010. doi: 10.1007/s00382-010-0874-3.

D. E. Waliser. Interseasonal variability. In B. Wang, editor, *The Asian Monsoon*. Springer, 2006.

G.T. Walker. Correlation in seasonal variations of weather. *Mem. Indian Meteorol. Dept.*, 24: 275–332, 1924.

G.T. Walker. World weather. part iii. *Mem. R. Meteorol. Soc.*, 4:97–106, 1928.

B. Wang, editor. *The Asian Monsoon*. Springer, 2006.

P. Wang, S. Clemens, L. Beaufort, P. Braconnot, G. Ganssen, Z. Jian, P. Kershaw, and M. Sarnthein. Evolution and variability of the asian monsoon system: state of the art and outstanding issues. *Quaternary Science Reviews*, 24:592–629, 2005.

Y. J. Wang, H. Cheng, R.L. Edwards, Z.S An, J.Y. Wu, C.-C. Shen, and J. A. Dorale. A high-resolution absolute-dated late pleistocene monsoon record from hulu cave, china. *Science*, 294:2345–2348, 2001.

D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393 (6684):440–442, 1998.

P. J. Webster. The elementary monsoon. In J. S. Fein and P. L. Stephens, editors, *Monsoons*. John Wiley & Sons, 1987.

P. J. Webster. The coupled monsoon system. In B. Wang, editor, *The Asian Monsoon*. Springer, 2006.

P. J. Webster, V. O. Magana, T. Palmer, J. Shukla, R. A. Tomas, M. Yani, and T. Yasunari. Monsoons: Processes, predictability, and the prospects for prediction. *J. Geophys. Res.*, 103: 14,451–14,510, 1998.

W.May. Variability and extremes of daily rainfall during the indian summer monsoon in the period 1901-1989. *Global and Planetary Change*, 44:83–105, 2004.

A. Wolf, J. B Swift, H. L Swinney, and J. A Vastano. Determining lyapunov exponents from a time series. *Physica D*, 16D:285–317, 1985.

H. Wulf, B. Bookhagen, and D. Scherler. Seasonal precipitation gradients and their impact on fluvial sediment flux in the northwest himalaya. *Geomorphology*, 118:13–21, 2010.

Tolga Yalçinkaya and Ying-Cheng Lai. Bifurcation to strange nonchaotic attractor. *Phys. Rev. E*, 56(2):1623–1630, 1997.

K. Yamasaki, A. Gozolchiani, and S. Havlin. Climate networks around the globe are significantly affected by el niño. *Phys. Rev. Lett.*, 100(22):22850, 2008.

M. Yanai and G. Wu. Effects of the tibetan plateau. In B. Wang, editor, *The Asian Monsoon*. Springer, 2006.

A. Yatagai, O. Arakawa, K. Kamiguchi, and H. Kawamoto. A 44 year daily gridded precipitation dataset for asia based on dense network of rain gauges. *SOLA*, 5:137–140, 2009.

K. Zickfeld, B. Knopf, V. Petoukhov, and H J Schellnhuber. Is the indian summer monsoon stable against global change? *Geophys. Res. Lett.*, 32:L15707, 2005.