INDIAN NATIONAL COMMITTEE ON SURFACE WATER (INCSW-CWC)

UID	MH-2011-106
Type (State whether final or draft report)	final
Name of R&D Scheme	Hydrological Impacts of Global and Local Changes in a Metro City
Name of PI & Co-PI	S. Ghosh & V. Jothiprakash
Institute Address	IIT Bombay, Mumbai - 400076
Circulation (State whether Open for public or not)	Open
Month & Year of Report Submission	April 2016

Hydrological Impact of Global and Local changes in a Metro City

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

(April 2016)

Funded by

MINISTRY OF WATER RESOURCES

ABSTRACT

The rapid urbanization all around the world is a matter of concern to the scientific community. Though urbanization is a worldwide phenomenon, it is especially prevalent in India, where urban areas have experienced an unprecedented rate of growth over the last 30 years. Presently, India has the second largest urban population of the world and is projected to have the highest urban growth rate in the next 30 years.

The urban regions are reported to have a warmer temperature in comparison to its surrounding rural areas; this phenomenon is recognized as the Urban Heat Island (UHI) effect. The formation of UHI has been shown to have a seasonal and diurnal variation. We observe the characteristics of Surface Urban Heat Island (SUHI) over 84 large urban centers of India. The novel SUHI seasonality observed in terms of cool SUHI over summer season is largely explained with the regional meteorology, vegetative condition and BC emission. The investigation reveals a higher depletion of vegetative land cover over the surrounding non urban region than the urban region during the summer season. The key findings of the study lies in unique set of SUHI characteristics and a higher diurnal temperature range of the nearby non-urban areas.

We take up an observational study to understand the influence of urbanization on mesoscale circulations and resulting convection, thus the nature of precipitation around urban areas. We evaluate the statistically significant changes in selected rainfall statistics of the Indian Summer Monsoon and investigate the explicit changes around urban land use in context of 40 large Indian urban areas. We find that rainfall activities are enhanced around the urban areas across different climate zones of the country. An additional examination of urbanization influence on heavy rainfall climatology is carried out through a point scale experiment with statistical framework of quantile based regression for the most populated city of India Mumbai, in pair with a nearby non-urban area Alibaug. The resultant extreme rainfall regression quantile also point toward sensitivity of extreme rainfall events to the local land use under urbanization.

The densely populated urban centers require a reliable weather forecast system to ascertain effective management of facilities, especially under the extreme weather conditions. The currently available precipitation forecast with numerical weather prediction models is limited to averages over larger areas, moreover the hydrological processes that occur on finer scales,

typically the circulation pattern leading to the extreme events remains unresolved. To overcome this limitation a statistical methodology is proposed for extreme precipitation forecast over the urban region of Mumbai, India. The methodology is employed undertaking the forecast output from the numerical weather forecast from Global Ensemble Forecasting System (GEFS). The model testing and validation results reveal the sensitivity of forecasting skill to the selection of predictor variables. The proposed model yields reasonably well forecasting performance and may provide a promising alternative for forecasting extreme rainfall occurrences over the city of Mumbai.

Climate projections are very important to understand the behaviour of different atmospheric processed under the changing global climate. Precipitation downscaling improves the coarse resolution and poor representation of precipitation in global climate models, and helps end users to assess the likely hydrological impacts of climate change. The two downscaling techniques namely dynamic and statistical downscaling, are widely used for this purpose. A major limitation of dynamic downscaling is the requirement of very high computational efficiency, the statistical downscaling remains relatively non expensive but does not take care of the physical processes. Here we propose a novel approach integrating the outputs from the statistical and dynamic downscaling methodology with the dynamical downscaling outputs for projection of extreme precipitation events. We observe the spatiotemporal growth of the city of Mumbai from 1973 to 2010 and obtain the urban land use projections for 2050. The representative climate projections obtained limited dynamical downscaling runs taking the changes in urban land use into consideration. The model runs reveals the changes in precipitation climatology at higher quantiles. These results are integrated with the statistical downscaling outputs to obtain a more realistic projection of extreme events.

Overall, this study highlights the important role of land use land cover and urbanization for understanding the mesoscale rainfall changes as part of regional climate change. The Indian urban regions reveal diverse SUHII characteristics than that reported in other parts of the world. The intensified the extreme rainfall in Mumbai under the influence of urbanization. However, the impacts of urbanization over heavy rainfall climatology are observed to be non-uniform across the country mainly resulting from dominating zonal trends. An integrated of dynamic and statistical downscaling approach is found to be much valuable to progress in the field of objective extreme weather forecasting as well as projections.

The scope for future studies includes i) A hypothesis driven model simulations and experiments to identify the detailed mechanism governing the behaviour of SUHII as well as the local scale physical mechanism during extreme rainfall events ii) An investigation may be undertaken for quantifying the impact and estimation of uncertainties associated with influence of urbanization over changes in regional rainfall pattern for better assessment of urban planning, water resources management and urban flooding. iii) the role of satellite based cloud data may be examined to improve the urban extreme forecast iv) development of computationally inexpensive data driven models integrating statistical and dynamical downscaling outputs to address urban climate related research questions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 : INTRODUCTION

1.1 Urbanization

Since the evolution of civilized society, man has modified the natural landforms for catering different community needs. The legendary cities have been identities of various eras of civilization and became landmarks of advancement of the civil society at the times. For many decades, centuries in some cases, cities have been spreading (Anas et al. 1998). Extended from the ancient times urbanization today is the defining phenomenon of the 21st century.

Rapid urbanization all around the world is a matter of concern to the scientific community. The fast growing urban areas carries out huge anthropogenic activities that burdens natural environment and its resources like air-water quality and space, thus have different climatology to their rural surroundings. Earth's climate is rapidly changing which is hard to explain by its natural variability and realized incident largely by disruptive impacts of various human activities.

With the industrial revolution in 1830s and the subsequent increase in the world's population, rapid land use/land cover (LULC) change has been underway. Forests have been altered to support farmland, grassland, and other land uses; natural landscapes have been modified to support settlement causing urbanization to rise. As specified by the United Nations (UN) report (2003) projects that almost all global population growth in the next 30 years will be concentrated in urban areas. On the same line the World Bank Institute of urban development gives an overview that for the first time in history, more than half the world's population lives in cities, with 90 percent of urban growth taking place in the developing world.

Though urbanization is a worldwide phenomenon, it is especially prevalent in India, where urban areas have experienced an unprecedented rate of growth over the last 30 years. During the last 50 years the population of India has more than doubled, but the urban population has grown nearly five times. The number of Indian mega cities will increase from the current three (Mumbai, Delhi and Kolkatta) to six by the year 2021 (including Bangalore, Chennai and Hyderabad), when India will have the largest concentration of mega cities in the world (Chakrabati 2001). According to the World Bank's urban challenges studies for India, the estimated urban population will reach 500 Million by year 2017.

1.2 Climate change

Climate change refers to any systematic change in the long-term statistics of climate elements (such as temperature, pressure, or winds) sustained over several decades or longer time periods (American Meteorological Society, http://amsglossary.allenpress.com/glossary). Human activities contribute to climate change by causing changes in Earth's atmosphere in the amounts of greenhouse gases, aerosols and cloudiness. For thousands of years, the Earth's atmosphere has changed very little. Temperature and the balance of heat-trapping greenhouse gases have remained just right for humans, animals and plants to survive. But, the human-induced enhanced effects are causing environmental concern, with its potential to warm the planet at a rate that has never been experienced in human history. According to NASA's Goddard Institute for Space Studies, average temperatures have climbed 0.8^0C around the world since 1880, much of this in recent decades. Also a number of climate studies indicate that the 20^{th} century's last two decades were the hottest in 400 years and possibly the warmest for several millennia. The United Nations' Intergovernmental Panel on Climate Change (IPCC) reports that 11 of the past 12 years are among the dozen warmest years since 1850.

Almost all meteorological parameters have seen changes in the last century. In particular rainfall has been observed to increase and become more intense. Several studies indicate this could be a signature of global climate change/warming (Allan and Sodden, 2008; Goswami et al. 2006). Both observational and modeling studies suggest that with warming, higher evaporation and precipitation rates can occur, which could lead to an overall acceleration of the global water cycle (Webster et al. 2005; Trenberth 2005). The events of excessive amounts of rainfall have also increased over the past 50 years. The consequences of global warming are reflected in global as well as regional climate in terms of changes in key climatic variables such as precipitation and atmospheric moisture, snow cover, extent of land and sea ice, sea level and patterns in atmospheric and ocean circulation. Therefore, study of climate change is necessary to understand its impact on hydrological processes. Water resources are inextricably linked with climate so the prospect of global climate change has serious implications for water resources and regional development (IPCC, 2007).

1.3 Urban climate and urban heat island

The process of urbanization involves removal of local vegetation, draining of marshes and turning the natural soil into impervious paved surfaces such as roads and buildings. In addition to this other human activities taking place in urban areas encourage generation of greenhouse gases. The increasing number of industries and automobiles tend to alter chemical composition of the atmosphere, the thermal and hydrological properties of the earth's surface as well as the aerodynamic roughness parameters. This effect is coined in terms of Urban Heat Islands (UHI) that involve temperature differences measured over space i.e. urban to rural with the surface heating as a main driving factor. Urban areas are warmer, like an "island" of heat surrounded by cooler rural areas. Large horizontal temperature gradient exist at the urban/rural boundary, could be as large as 4°C/km Urban areas appear like a "plateau" with weaker increasing gradient. The impacts of urban heat islands can be best understood by studying the nature and characteristics of urban climate.

With the same amount solar radiation received by an acre of forest/cropland and the densely built up urban area the green space remains cooler than the built up because of transpiration and shading of the ground, whereas the urban surfaces get much hotter than vegetated surfaces. The built up areas also release this energy at night, creating a dome of warmer air at night (Olanrewaju, 2009). Almost every city in the world is between 1-4^{0}C hotter than its surrounding areas.

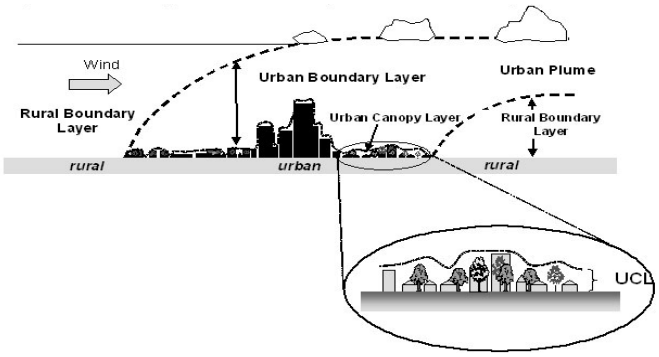


Figure 1.1Schematic depiction of the main components of the urban atmosphere (http://www.actionbioscience.org)

1.4 Influence of urban heat island on precipitation

Urban microclimate is the climate develops over a city and modified by variation in aspect, shape and form of the ground, soil moisture and surface vegetation (Oke, 1988). The urban

microclimate affects the heat and water balance, water runoff and causes differences in temperature, precipitation and cloud cover of the area. Furthermore, it impacts the local meteorology by altering local wind patterns, forming cloud and fog, increasing humidity, and changing the precipitation rate.

Observational and climatological studies have theorized that the UHI can have a significant influence on mesoscale circulations and resulting convection. Thus due to the influence of UHI the nature of precipitation in and around urban areas has become unpredictable using the conventional methods and existing models. Metropolitan Meteorological Experiment (METROMEX) studies have shown that urban effects lead to increased precipitation during the summer months and this increased precipitation is typically observed within, and 50–75 km downwind of the city.

1.5 Identification of the research gap

According to the 2011 Census, the urban population growth of India is reported at 2.76 % per annum during 2001-2011 with the total urban population of 377 million. The level of urbanization in the country has increased from 27.7% in 2001 to 31.1% in 2011. The existing urban centers have visibly expanded also a huge number of new towns emerged during the last decade, contributing significantly to the speeding up of urbanization.

Functioning as centre of economic activities the urban areas are reported to have a different climatology. The causative factors of the UHI effect given by Oke (1982) have been confirmed and further broadened through a variety of studies around the world. Compared to non-built surroundings, built-up areas of cities differ considerably in albedo, thermal capacity, roughness, etc. which can significantly modify the surface energy budget (Arnfield, 2003). A number of studies suggest that the intensity of UHI could be increased by anthropogenic heating (including contributions from vehicles, building sector, and human metabolism) (Sailor and Lu, 2004) as well as CO2 and pollutants emissions (McCarthy et al.,2010; Taha, 1997). The UHI of the Indian cities is mostly studied in isolation dealing with individual cities. These studies largely vary in terms of methodology, and empirical approaches, based on either air temperature or LST, attempt to reveal the linkage between the UHI intensity and various descriptive indicators of cities. Therefore suffering from inconsistency and instability in regards to the urban-rural definition, hindering the inter-comparison between results. A

comprehensive analysis to assess the UHI intensity of all the Indian cities with varied area, population and geographical location by means of remotely sensed land surface temperature data does not exist.

With the increased urbanisation and industrialization (UN Habitat, 2010b) the impact of local scale land surface processes on regional meteorology and extreme precipitation has become important (Zhang etal, 1996, Karl and Trenberth 2003). Number of studies analysed impact of urbanization on mesoscale convection and precipitation found that the urbanization has an important feature in regional meteorology (Oke, 1988, Shepherd et al., 2002, Rozoff et al., 2003, Gero and Pitman 2006, Pyle et al., 2008). A significant number of studies reports intensification of rainfall under the influence of urbanization over the world. On the other hand there exists considerable debate on influences of urbanization versus large scale forcing on changes in extremes, and also on the pattern of changes of extremes in urban areas over India. The hypothesis that urbanization affects the association between extreme rainfall and synoptic scale weather patterns has also not yet been tested, except with a few Weather Research and Forecasting (WRF) model runs (Litta et al. 2010,Srinivas et al. 2013). A comprehensive understanding of impact of urbanization over the heavy rainfall climatology is very important for regional water resources planning and management.

Occurrence of rainfall extremes are expected to increase under the warming global climate (Diffenbaugh et al., 2005). A correct scientific understanding and modelling of occurrence the extremes events is important to minimize the impacts of rainfall extremes. Significant research has been carried out in the science of extremes in the last few decades (Knukel et al., 1999, Coles and Powell, 1996). However, there are considerable numbers of extreme events that remain unpredicted, and flooding related to severe storms claims many lives every year in different parts of the world. The impact of extreme precipitation goes worst when it occurs over densely populated urban areas functioning as centre of economic activities of the country. For example in India the extreme rainfall event with heavy downpour about 944 mm in 24 h occurred in July 2005 over the urban centre of Mumbai, caused nearly 500 fatalities and the economic loss of 2 billion US dollars(Ranger et al., 2011).

Numerical Weather Prediction (NWP) models based on dynamical weather equations are run to produce an ensemble forecast of mesoscale (20–200 km) precipitation, along with other meteorological variables, using synoptic scale (200–2,000 km) weather conditions commonly used to provide accurate and meaningful forecasts based on the weather conditions (Giorgi et al., 2001). However, the current NWP ensemble forecasts are found to be typically biased and

underestimate forecast uncertainty resulting mainly from model structural and approximations of subgrid-scale processes (Goddard et al., 2001; Maraun et al., 2010) and are highly ineffective in predicting heavy rainfall events (Březková et al., 2010; Hong and Lee 2009; Khaladkar et al., 2007; Selvam 1988). In India the National Centre for Medium Range Weather Forecasting (NCMRWF) produces rainfall forecasts at 50 and 35 km spatial resolution based on physics-based models. However, these models are poor in simulating extreme rainfall in India (Khaladkar et al., 2007). A skilful rainfall forecasts and early prediction of the extreme events may help in broadcasting alerts to the population to safeguard their lives and properties before advancement of the flood. A warning system can also be very useful in effectively operating evacuation, flow diversion, preparedness of the disaster mitigation team of the existing flood control systems. Despite the fact that myriad previous studies have addressed the precipitation forecasting, as per the author's knowledge, none of the research has addressed the one important attribute of the rainfall i.e. the day when certain amount of rain is received; and its relation to the climatic circulation pattern.

Under the changing global and local climate it is very important to obtain the precipitation projections for an urban region. The regional simulation of coupled Weather Research Forecast (WRF) Model with urban canopy models with changes on LULC provides a precise projection of rainfall for future in urban regions. However the coupled model simulation demands a huge computational efficiency for long term simulations. For example in order to obtain climate projections of an individual urban region need 30 years of historical (past) and 30 for future simulations. Statistical downscaling is computationally inexpensive in this regard but cannot consider urban feedback. A methodology integrating statistical downscaling to produce a long term downscaling of GCMs with WRF runs over limited time period coupled with urban canopy model (UCM) to consider future urbanization impacts may be of much useful in this regard.

1.6 Objective of the study

The objectives of the present study can be summarized as follows:

i) To study the characteristics of SUHII over Indian cities with its diurnal and seasonal variations.

ii) To understand the factors a dependability of SUHII development over Indian cities

iii) To analyse the changes of Indian rainfall extremes at different urban regions, and nearby non–urban regions,

iv) To test, if the changes are solely due to urbanization or affected by regional synoptic scale changes.

v) To identify the observed changes in association between the synoptic scale weather pattern and development of urbanization.

vi) To observe the development of Urban Hear Island of the major urban centres of the country

vii) To understand the seasonal and diurnal variation of Urban Hear Island of the major urban centres of the country.

viii) To provide a computationally inexpensive data driven methodology to forecast the occurrence of extreme rainfall event for an urban region.

ix) To instigate the projection of extreme rainfall considering the future urban expansion.

Chapter 2 : URBAN HEAT ISLANDS IN INDIA

2.1 Introduction

Urbanisation process involves change in Land Use Land Cover (LULC) the impact of change of LULC influences the regional climate (Su et. al., 2010; Su and Yang, 2007). Various studies have shown that land uses in a region influenced the urban temperature (Su et. al., 2010). A warmer surface temperature exists in the urban area due to change in land cover and population density. The difference between urban and surrounding rural surface temperatures is indicated as Urban Heat Island (UHI) (Oke, 1987). The factors such as the thermodynamic capacities of materials, structural geometry, and heat generating activities cause increased storage and re-radiation of heat to the atmosphere (Arnfield, 2003). UHI is one of the most significant impacts of rapid urbanization and urban sprawl. The phenomenon of urban heat islands was first investigated by Luke Howard (1810), through the influence of urban areas on local climate. Oke (1988), followed by Arnfield (2003), analyzed the drivers of urban heat island by the contrast of surface energy exchange between urban and suburban areas. Downward net solar radiation and anthropogenic heat flux produced by appliances, building heating and light, humans, combustion engines, and transportation constitute the two major sources of energy available to cause urban heat island. These sources of energy are converted into sensible heat fluxes, latent heat fluxes, surface heat storage, and net heat advection (Arnfield 2003). During the day, generally, sensible heat fluxes and latent heat fluxes mainly derived from net solar radiation are the largest upward heat fluxes (Voogt and Oke 2003,Oke 1988). The incoming solar energy partitioning between latent heat flux and sensible heat flux is modulated by vegetation fractional coverage and its ability to transpire soil−water per unit of vegetated area. In an urban area the precipitation is channelled into the storm sewers; where runoff is treated or discharged rather than remaining available for evaporation, at the same time the reduction of vegetation decreases the amount of water available for evapotranspiration. Hence, the replacement of natural soil or vegetation by materials used in cities like concrete or asphalt reduces the ability to decrease the ambient temperature through evaporation and plant transpiration. This largely affects Latent Heat Flux (LHF) referred as the heat released through the process of evapotranspiration of water. Sensible Heat Flux (SHF) refers to the heat transferred by conduction within different surface materials and by dry convection (vertical) and advection (horizontal) wind. The urban areas have an increased surface roughness because

of which the surface winds slows down, hence inhibiting the sensible heat loss through convection. The thermal properties of materials used in building urban structures also increases the stored amount of sensible heat within the city during daytime and it is released into the urban atmosphere after sunset. Thus, it can be seen that land surface temperature change and formation of UHIs are directly related to the LULC changes. UHI phenomena is generally seen as being caused by a reduction in latent heat flux and an increase in sensible heat in urban areas as vegetated and evaporating soil surfaces are replaced by relatively impervious low albedo paving and building materials. This creates a difference in temperature between urban and surrounding non-urban areas (Imhoff et al., 2010).

The Land Surface Temperature (LST) refers to the radiation properties of the earth surface and determines the intensity of the radiation of long waves emitted by it, which is detected by aircrafts or satellite based remote sensing platforms (Jin and Dickinson, 2010; Urban et. al., 2013). The temperature recorded by the synoptic measurements in weather stations is known as the air temperature (T_{air}) measured at a height of 1.5-2 m height. As the spatial distribution of the weather stations is limited and the dissemination of temperature data is variable, their use for real-time applications is limited. Compensation for this paucity of information can be obtained by using satellite-based methods. The land surface emits radiance differently across the thermal spectrum, and the emitted radiance is affected by the composition of the surface constituents, particularly the spectral emissivity.

2.2 Motivation and Objectives

India has experienced an unprecedented growth of urbanization over recent decades with urban population of 3.8 million(Census of India 2011). However the UHI characteristics of Indian urban centres are highly overlooked. The UHI studies over India are undertaken mostly with individual cities using varied methodology, without any specific analysis on understanding the characteristics of UHI, its seasonal, diurnal and spatial variability. This study is perceived to address this gap in terms to provide a comprehensive observation of the UHI characteristics of all major urban centres of India. We attempt to analyse the SUHII based on the surface temperature data obtained from MODIS satellite data for 84 large urban locations in India to reveal the factors governing the characteristic of SUHII. To explore the drivers of surface urban heat islands, we observe combine satellite observations of vegetation index, and albedo, large scale climatic circulation, and population as a proxy of socio-economic activities.

2.3 Data used

The population of each selected urban centre is extracted from the Census India 2011, urban population database. The study utilizes three major data sets namely Resolution Imaging Spectroradiometer (MODIS) data, large scale climate circulation from ERA Interim reanalysis and Black Carbon emission inventory data.

2.3.1 MODIS satellite data

Here we use three data sets from Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data product: (i) Land Cover (LC) information (ii) Land Surface Temperature (LST) and (iii) vegetation condition (NDVI). Identification of the urban clusters is based on the MODIS LC type product (MCD12Q1, 500 m, annual) of the year 2008. Here we use the LC classification types from International Geosphere Biosphere Programme (IGBP) (Belward et al., 1999; Scepan, 1999) with 17 LC classes.

We estimate the SUHII with MODIS-Aqua LST dataset (MYD11A2, 1000 m, Version 5) at eight-day interval. This data product is validated over a widely distributed set of locations and time periods via several ground-truth and validation efforts (Glynn, C. H., Simon, 2009, Cesar C. et al, 2009, Wan, Z. and Li, Z. 2008) and is frequently used for surface UHI analysis (i.e. Jin et al., 2005; Hung et al., 2006; Imhoff et al., 2010; Rajasekar and Weng, 2009; Peng et al., 2010; Zhang et al., 2010; Zhou et al., 2010). The MODIS LST data is derived from two thermal infrared band channels, 31 (10.78–11.28 μm) and 32 (11.77–12.27 μm), using the split-window algorithm (Wan et al., 2002). This algorithm corrects for atmospheric effects and emissivity using a look-up table, based on global land surface emissivity in the thermal infrared (Snyder et al., 1998). The dataset is comprised of daytime (~13:30) and nighttime (~01:30) LSTs, quality control (QC), observation times, view angles, bits of clear sky days and nights, and emissivity estimates. The QC values provide very important information for filtering of low-quality pixels due to clouds or other processing failures. For the current study the QC Scientific Data Set (SDS) for LST are extracted by reading the bits in the 8-bit unsigned integer (Wan, 2007). The data pixels where, the error in the computed LST is less than 3^0K, are considered for this analysis.

The vegetation indices are obtained from MODIS-Aqua product (MYD13A, 1000 m) at temporal resolution of 16 days. The QC information available with the data product is utilised to filter out good quality data to be utilised for the analysis.

All three MODIS data products for the present study are obtained from climate data archive at from the Earth Observing System Data and Information System (EOSDIS) service tool reverb: http://reverb.echo.nasa.gov. The LST and NDVI datasets were selected during the period between 2003 to 2013-14 summer (March to May) and winter (December to February).

2.3.2 Climate Reanalysis data

The large scale climate circulation pattern is derived from ERA-Interim (Simmons et al. 2007) reanalysis data produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) at 1° latitude by longitude resolution. Here we use air temperature, Relative Humidity (RH), and the northward and eastward wind field. The circulation pattern observed at the surface level over Indian subcontinent region encompassing 5°-40° N and 60°-120° E, is collected for the present analysis during 2003-2013. The ERA-Interim re-analysis project output of 6-hourly surface analysis fields, at 00-hr and 12-hr are collected. This selected time period of observation of the reanalysis data differs only by 1.5 hours from the nighttime (~01:30) and daytime (~13:30) LST observation. Hence, it is considered appropriate as the background climatic condition over development of the SUHII.

2.3.3 Emission inventory data

Emission inventories are important research and regulatory tools, for understanding atmospheric responses to changing emissions and for formulation of mitigation policies at national and inter-governmental levels. They are inputs to atmospheric models operating on scales from a few kilometers, for studies of urban to regional air quality (Guttikunda et al., 2005; Carmichael et al., 2009), to hundreds of kilometers, for studies of inter-hemispherical transport of pollutants and global climate change (Reddy and Boucher, 2007; Schultz et al., 2006). To examine the possible role of Black Carbon (BC) aerosols on SUHII development, we undertake the BC emission, at 0.25^0 latitude spatial resolution, following earlier studies (Sadavarte and Venkataraman, 2014, Pandey etal., 2014). The emissions inventory data includes emissions of BC from residential cooking and heating with biomass fuels, lighting with kerosene lamps, on-road diesel vehicles, and agricultural residue burning in fields, diesel use in agricultural tractors, pumps and brick production in traditional kilns.

2.4 Methodology

The 84 big cities of India with a population larger than 1million are identified according to population data for year 2011 (Census India 2011). A geographical database is prepared with

the root location of the city. Since administrative city boundaries differ from the actual extent, the cities are defined as clusters of urban land cover. The Surface Urban Heat Island Intensity (SUHII) is defined as the difference of land surface temperature between urban area and suburban area. We used City Clustering Algorithm (Rozenfeld et. al. 2008) to determine the urban area for each big city. We defined the suburban area as the buffer zone that is a percentage (50-150%) of the urban area around the urban area. A similar UHI intensity calculation has been conducted by Peng et al. (2012), suggesting minor influence of the boundary size, i.e., 50%, and 100%, 150% of the cluster size. The urban pixels are defined as land covered by buildings and other man-made structures. The detailed algorithm for the urban and suburban area determination is as follows:

1. Enqueue the pixel of the location for each big city, determined from a geographical database as the root node into the search queue.
2. Dequeue a node and examine the eight neighbours around it. If the land cover type of the neighbouring pixel is urban land cover, add the neighbouring pixel into the queue, and assign attribute of the neighbouring pixel as an urban pixel, otherwise assign attribute of the neighbouring pixel as a non-urban pixel.
3. If the queue is empty, every pixel in the queue has been examined – quit the search and return the urban map, otherwise repeat from step 2.
4. After the urban map is returned, suburban area are defined as the buffer zone, which is a ring zone around urban area that consists the nonurban pixels excluding water pixels covering the 50-150% of land as urban area.

Figure 2.1 shows the urban centers Surat and Vadodara of Western India land cover map. The urban area maps with selected suburban regions for areas are 60% and 71% of the respective urban areas are overlaid on the MODIS land use and LST data. As it can be observed form figure 2.1(b); the SUHI is apparently visible with the LST data. The land use around the two cities is mainly croplands and shrub lands.

Table 2-1 Selected Urban centers of India

Cluster no	Urban cluster	Cluster no	Urban cluster	Cluster no	Urban cluster
1	Nellore	29	Jamnagar	57	Srinagar
2	Bhagalpur	30	Aligarh	58	Meerut
3	Shahjahanpur	31	Dehradun	59	Jamshedpur
4	Malegaon	32	Mysore	60	Durg-bhillai
5	Raipur	33	Tiruchirappalli	61	Jammu
6	Kolhapur	34	Bharuch	62	Allahabad
7	Tirunelveli	35	Rourkela	63	Coimbatore
8	Korba	36	Kota	64	Bhopal
9	Udaipur	37	Gorakhpur	65	Varanasi
10	Tiruppur	38	Bokaro	66	Visakhapatnam
11	Bhatinda	39	Amravati	67	Nagpur
12	Mathura	40	Nanded	68	Vadodara
13	Moradabad	41	Dhule	69	Patna
14	Siliguri	42	Kalyan	70	Ludhiana
15	Bareily	43	Solapur	71	Chandigarh
16	Salem	44	Jalandhar	72	Agra
17	Nashik	45	Ankleshwar	73	Indore
18	Jalgaon	46	Guwahati	74	Lucknow
19	Guntur	47	Vijayawada	75	Kanpur
20	Ambala	48	Jodhpur	76	Surat
21	Bhubaneshwar	49	Warangal	77	Jaipur
22	Aurangabad	50	Gwalior	78	Pune
23	Gulbarga	51	Amritsar	79	Indore
24	Durgapur	52	Madurai	80	Ahmedabad
25	Akola	53	Bhavnagar	81	Bangalore
26	Patiala	54	Jabalpur	82	Chennai
27	Cuttack	55	Rajkot	83	Hyderabad
28	Ahmednagar	56	New delhi	84	Mumbai
				85	Kolkata

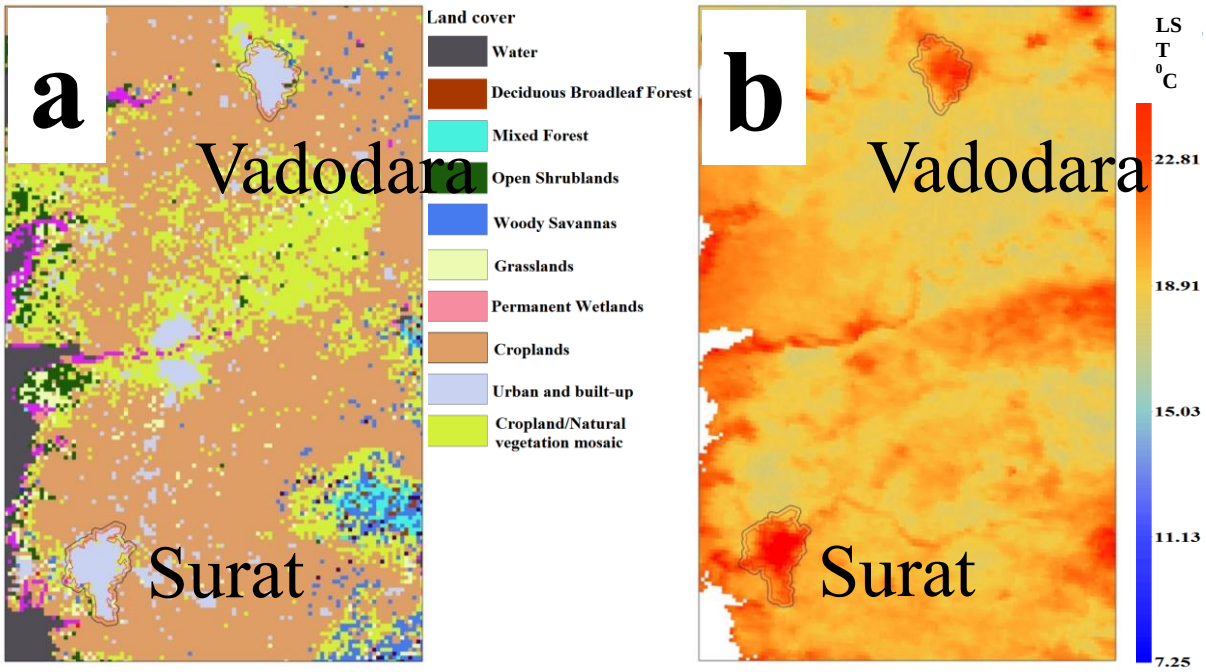


Figure 2.1 Map of cities Surat and Vadodara

a. MODIS data derived land cover/use, b. Mean nighttime LST

Since LST data are based on clear-sky conditions, a coverage threshold is defined as the UHI intensity is regarded as valid only if the LST values are available for at least 50% of the cluster and boundary cells. The quality control data are supplied with each MODIS pixel, classified mean LST error into four levels, $\leq 1^0$C, $\leq 2^0$C, $\leq 3^0$C and $> 3^0$C. While calculating mean temperatures of clusters and boundaries, the pixels with a mean LST error $> 3^0$C are first filtered out. For each city, we calculate the difference of urban LST minus suburban LST. Daytime and nighttime SUHII were calculated separately from EOS-Aqua-MODIS LST in the early afternoon (~13:30) and at night (~01:30), respectively. The seasonal mean daytime and nighttime SUHII are computed for the Indian Summer months (March-April-May) and Winter months (December-January-February). The spatial differences of SUHII among big cities are studied during the period 2003−2013.

In parallel with daytime and nighttime SUHII, we define a vegetation parameter, an albedo parameters, three climate parameters, and population parameters for each big city. The vegetation parameters are calculated as the NDVI difference between urban and suburban pixels. The albedo parameters are estimated by the difference between urban and suburban pixels in the black sky albedo. Two more parameters WSA-linear with BSA and EVI- linear with NDVI are also checked, they shows similar results to BSA, so only BSA, NDVI is shown

here. The climate parameters over the Indian sub-continental region as well as each big city including mean air temperature, mean Humidity, and mean resultant wind velocities are extracted from ERA Interim data sets.

2.5 Results and Discussion

Among the 84 urban locations in India (Figure 2.2 a), 9 have population more than 5 million, and 34 have in the range of 1-5 million. We compute the SUHII for all these locations separately for Indian pre-monsoon summer (March-May) day time, summer night time, winter (December-February) day time and winter night time.

2.5.1 Seasonal and Diurnal SUHII characteristics

The figure 2.2 presents computed SUHII for Indian pre-monsoon summer (March-May) day time (Figure 2.2 (b)), summer night time (Figure 2.2 (c)), winter (December – February) day time (Figure 2.2 (d)) and winter night time (Figure 2.2 (e)). We observe majority of the urban locations in India, specifically in Central India and Gangetic Basin have statistically significant negative SUHII during summer day, which is exactly opposite to that presumed as the profound impact of urbanization. During summer night, the negative SUHII change their sign to positive, with statistical significance, at almost all locations and the UHI is prominent over the entire Central India. This points to strong diurnal characteristics of SUHII with mostly positive differences between summer night and summer day LST in the interior India; however, the differences are negative in many coastal cities.

The night time winter SUHII is positive in all locations except one, but with no statistical significance for majority of locations. The differences between winter day and night time SUHII show almost similar pattern to that observed during summer.

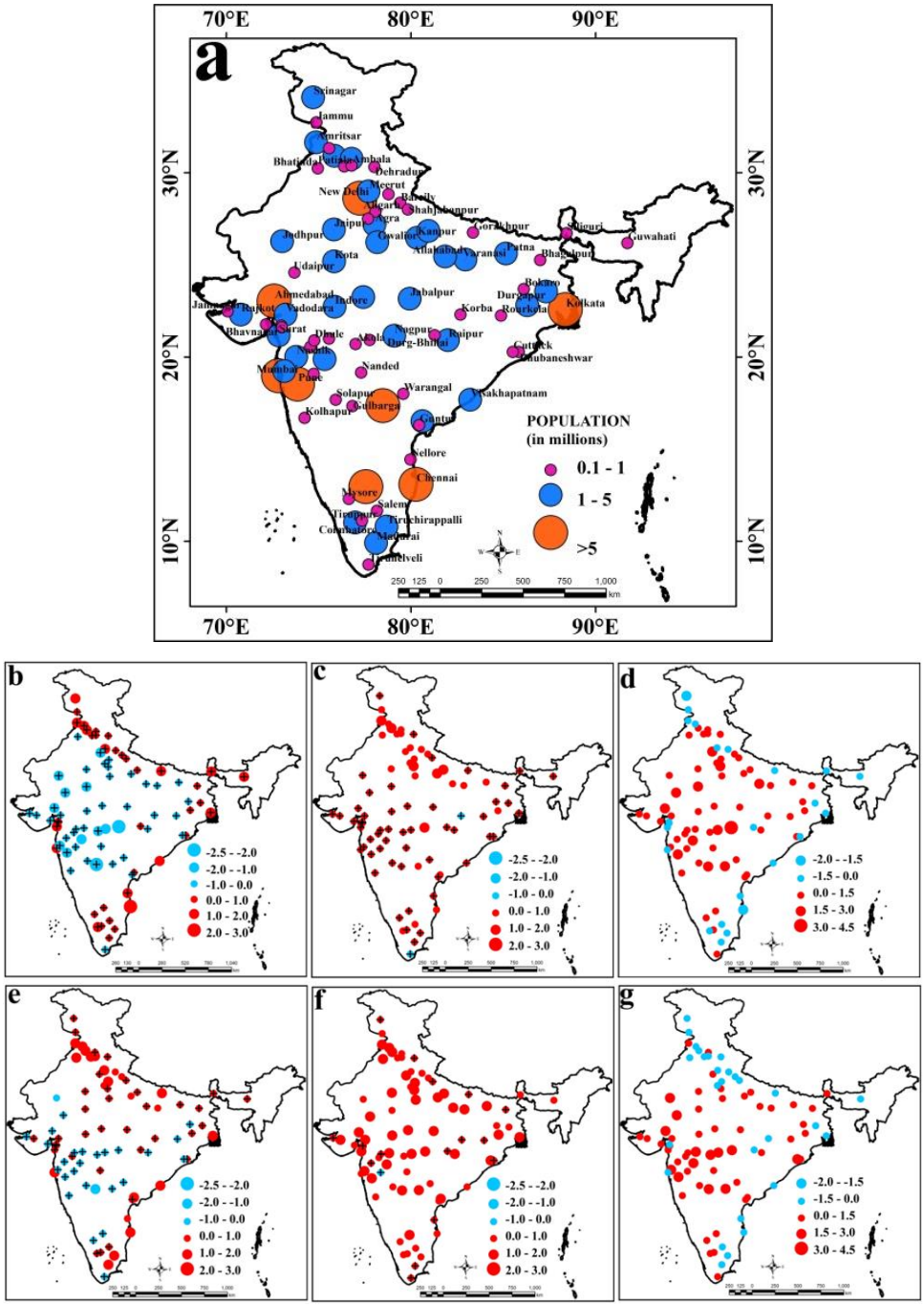


Figure 2.2 Location and population of urban centers.. SUHI of Indian urban centers.

a.Location and population of major urban centers **b.** Summer mean daytime SUHII **c.** Summer mean nighttime SUHII **d.** Difference summer night-day SUHII **e.** Winter mean daytime SUHII **f.** Winter mean nighttime SUHII **g.** Difference Winter night-day SUHII. The + sign indicates field significant SUHII (at 5% significance level) estimated using t-test statistics.

2.5.2 Attribution of SUHII to background climate condition

Comparison of wind velocities between summer day and night (Figure 2.3 (a) and (b)) in coastal region reveals higher sea breeze during summer night resulting reduction in SUHII. In the summer time the stronger coastal wind built up is observed along the western coast revealing the greater air circulation and cooler night time SUHI over the cities of Kolkata, Vishkhapattnam, Guntur, Bhubneshwar, Nellor and Chennai. Coastal cities cool down more rapidly than cities lying in a valley where the influence of cooling air from surrounding rural areas is limited.

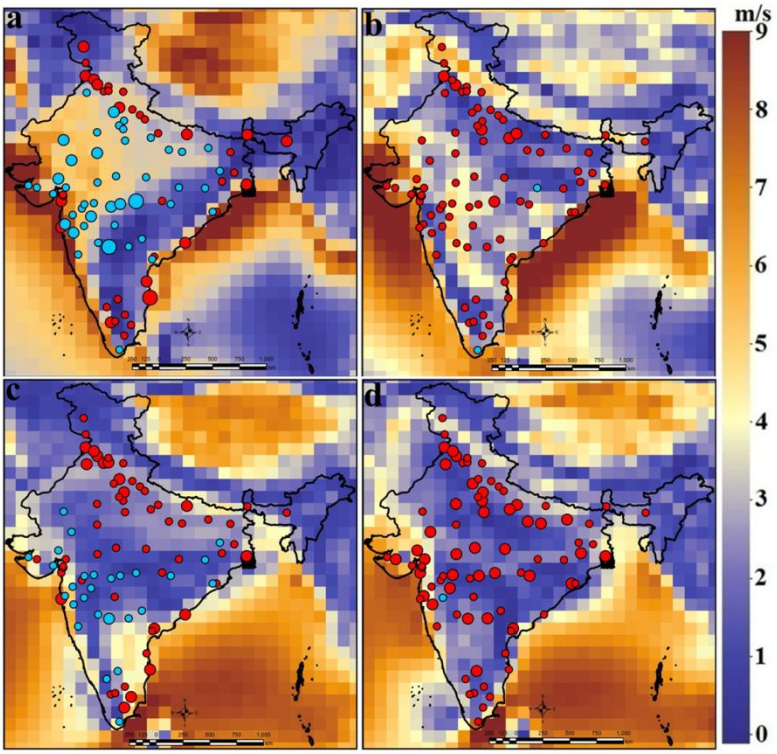


Figure 2.3 Attribution of SUHII to the surface level Wind indicated by large scale circulation pattern derived from the ERA Interim reanalysis data

a. Summer daytime wind overlaid with the corresponding SUHII b. Summer nighttime wind overlaid with the corresponding SUHII c. Winter daytime wind overlaid with the corresponding SUHII d. Winter nighttime wind overlaid with the corresponding SUHII

The night time SUHII for both summer and winter are significantly correlated with population in the urban regions excluding mega-cities, probably because the patterns of population and built up areas are different in mega-cities compared to other urban areas. The day time SUHII does not have any statistically significant correlation with population (Figure 2.4).

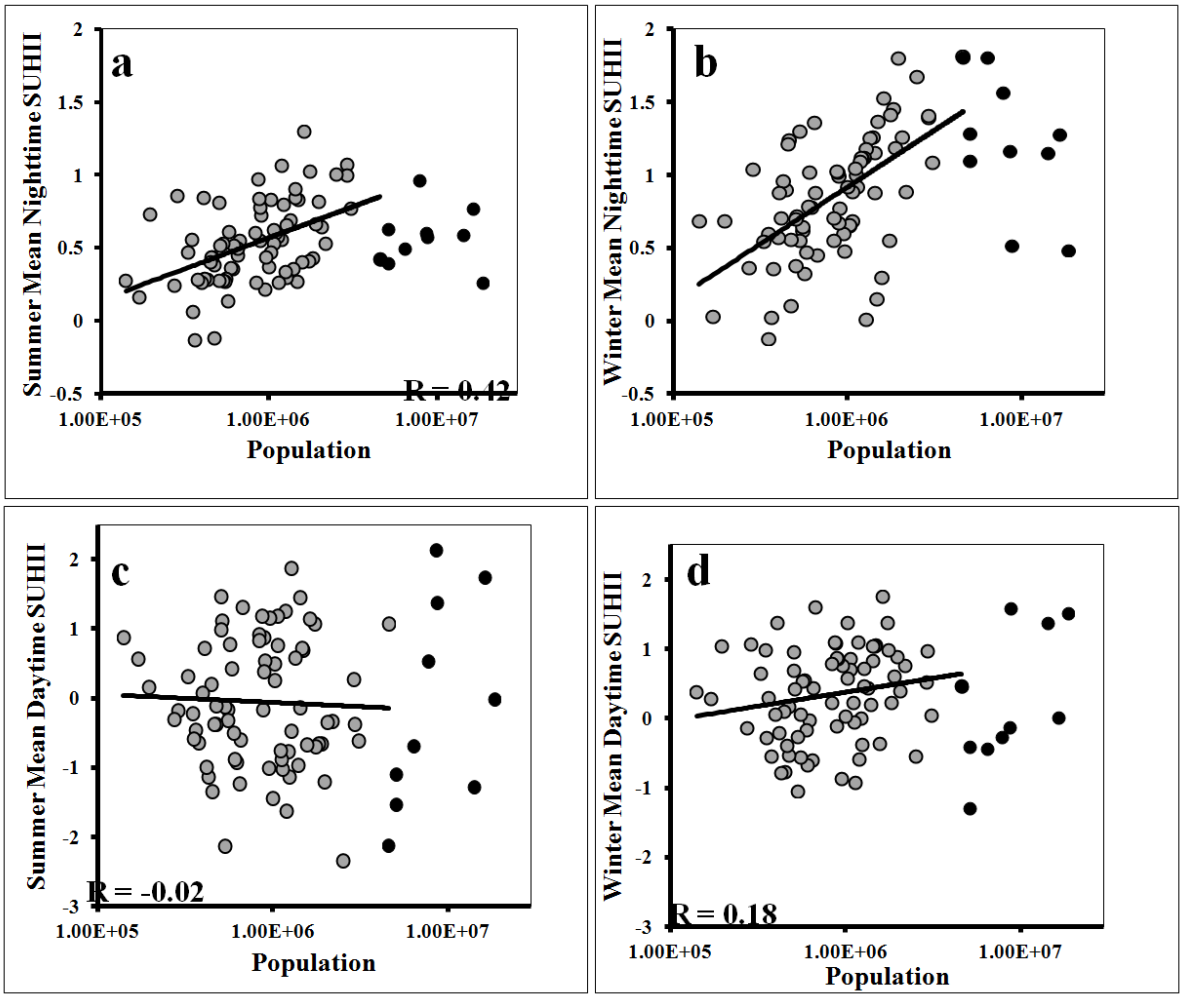


Figure 2.4 Sensitivity of day time SUHII to the population, lines are linear regression fits to the data, megacities are indicated with solid points.

a. Dependence of Summer nighttime SUHII on population b. Dependence of Winter nighttime SUHII on population c. Dependence of Summer nighttime SUHII on population d. Dependence of Winter nighttime SUHII on population

Here we try to bring out the reasons behind the unexpected negative summer day SUHII. We find that low vegetation cover during pre-monsoon summer in non-urban regions is responsible for such unusual characteristics of summer day SUHII. Figure 2.5(a) and (b) present the NDVI over Indian landmass which show very low vegetation cover during the pre-monsoon dry summer compared to post monsoon winter season. The differences in the NDVI between urban and nearby non-urban regions are presented in Figures 2.5(c) and (d) for summer and winter respectively. Positive to very low negative differences in the summer NDVI between urban and nearby non-urban regions attribute to the low vegetation cover over non-urban region during the same season, which is not the case for winter. Both summer and winter SUHII are

negatively correlated with the difference in NDVI between urban and non-urban regions and this proves our hypothesis that low vegetation in non-urban regions result into negative SUHII. Low vegetation during pre-monsoon summer season results into barren land surface which has lower albedo as compared to the urban build up areas[24]. Figure 2.5 (g) and (h) further show that the variations of difference NDVI for summer and winter. They are highly negative in winter resulting positive SUHII during the same season. Further analysis shows that majority of the non-urban regions have the land use as cropland. During pre-monsoon dry period they turn into barren land resulting high LST. The night time SUHII does not depend on NDVI as albedo has minimum role to play in absence of sunlight, and we find positive SUHII.

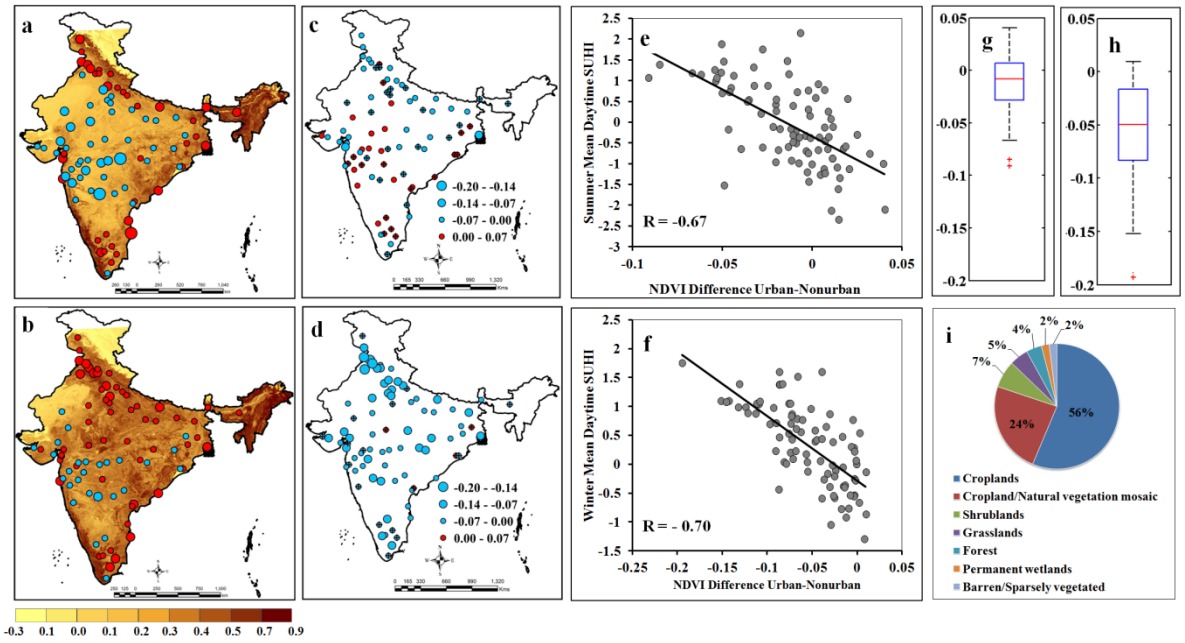


Figure 2.5 Association of SUHII in India with change in vegetation cover.

The SUHII for summer day (a) winter day and winter day (b) are overlaid on the vegetation cover. The difference between vegetation cover over urban and nearby non urban region is estimated for the Summer (c) and Winter (d) season. The Summer and Winter daytime SUHII is negatively associated with difference between vegetation cover over urban and nearby non urban region ((e) and (f) respectively). The overall variability of difference between vegetation cover over urban and nearby non urban region for Summer and Winter season is estimated ((g) and (h) respectively).

We find that the diametrically opposite seasonal patterns of SUHII between summer and winter day time exist in the North and Central India (Figure 2.2). Daytime land-surface temperature could also be influenced by atmospheric abundance of radiation absorbing constituents, like BC aerosols, pollution particles emitted from incomplete combustion of fuels, which strongly

absorb radiation over the entire solar spectrum[25]. Radiation absorption from BC can lead to heating of atmospheric layers in which these particles are abundant, at instantaneous rates up to several degrees °K per day[26]. There is strong haze from aerosol pollution over north and central India[27], particularly in winter months. Such aerosol induced haze, including a significant fraction of BC, could lead to surface radiation balance changes which affect land surface temperature, and consequently SUHII. To examine the possible role of BC aerosols the spatial plot of SUHII is overlaid on BC emission fluxes, spatially distributed on a 25 km grid, calculated in a recent emissions inventory for India[28-29]. The BC spatial distribution in winter reveals larger emissions in the Indo-Gangetic plain, central India and some clusters in the west coast (Ahmedabad-Mumbai belt), east coast and south India, corresponding to the density of population of users of biomass fuels.

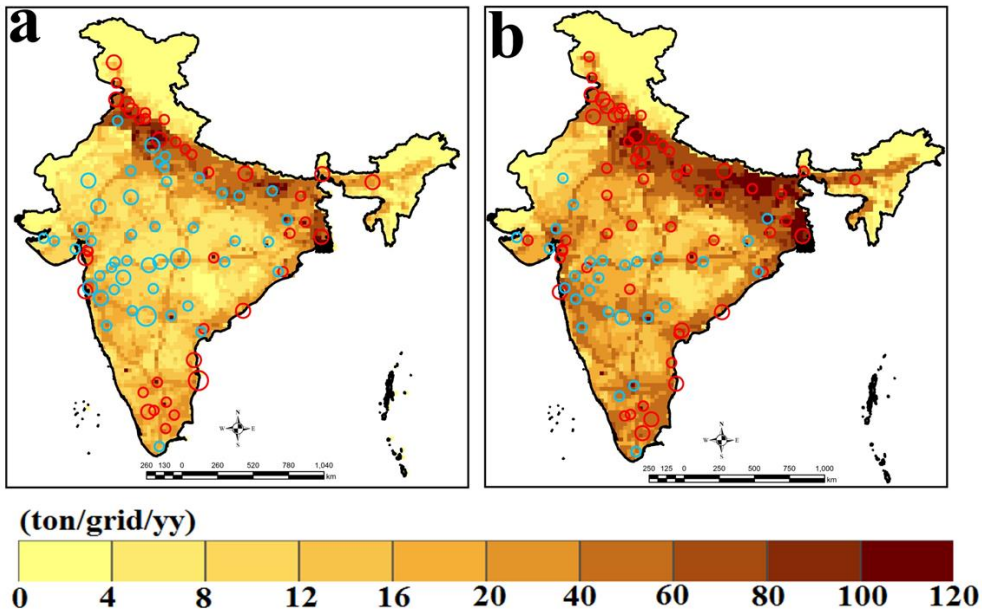


Figure 2.6 The effect of Black Carbon (BC) emission on variability of SUHII.

The SUHII for Summer and Winter day are overlaid on the BC emission colour map ((a) and (b) respectively).

The BC mediated mechanism of radiation flux changes, would only manifest during daytime. Thus, comparing daytime SUHI values of winter with those in summer, it is observed that SUHI at over 20 sites in regions with higher wintertime BC emissions is positive (red) but is negative (blue) at the same sites in summer. This change is co-located with increases in BC emissions in winter over summer, indicating urban land-surface temperatures exceeding those in adjoining areas in the winter months, while being reduced in summer months. Further, at about 10 more sites, in northwest and central India, the negative SUHI value is seen to be

reduced (smaller blue circle) in winter, compared to that in summer, once again indicating increase in urban land-surface temperatures compared to those in adjoining areas in the winter months. Correlation analysis of point SUHII with BC emission density (Gg/mon-grid), however, failed to show statistically significant association (Figure 2.7).

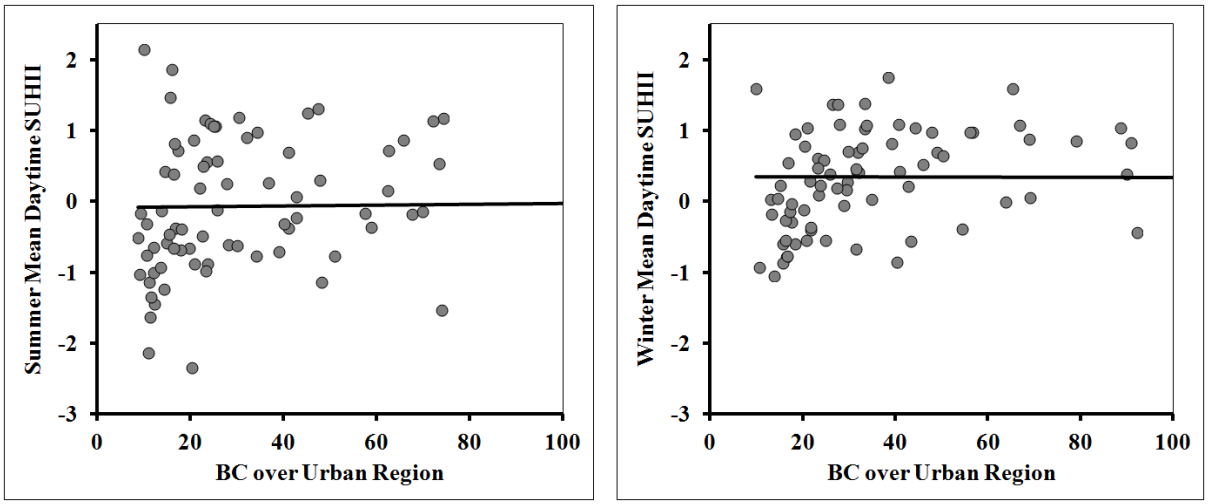


Figure 2.7 Sensitivity of the daytime SUHII to Black Carbon (BC) emission, lines are linear regression fits to the data

a. Dependence of Summer daytime SUHII on BC b. Dependence of Winter daytime SUHII on BC

This is not unexpected, since possible change in the radiation balance would be linked to the atmospheric concentration of BC in the surface layer, and its vertical distribution, rather than to emissions. In winter months, the prevailing meteorology in north India leads to low mixed layer heights and poor ventilation rates[30], which concentrates pollutants close to the surface. Thus emission distributions, like those analysed in this work, reveal a possible effect of BC emissions influencing SUHI. Further analysis is needed, utilising such emissions in chemical transport models to calculate columnar concentrations of BC, and subsequent radiative effects, which could influence the surface radiation balance.

We also compute the correlation between SUHII and the overall surface air temperature in the regions to which the urban regions belong to. For India, we find it negative for both summer and winter (Figure 2.8(a)-(d) and (i)-(j)). It indicates that during high temperature spells, the SUHII will be less and hence, the urban heat wave characteristics of India are different from other regions around Globe[23](Mishra et al., 2014, ERL).

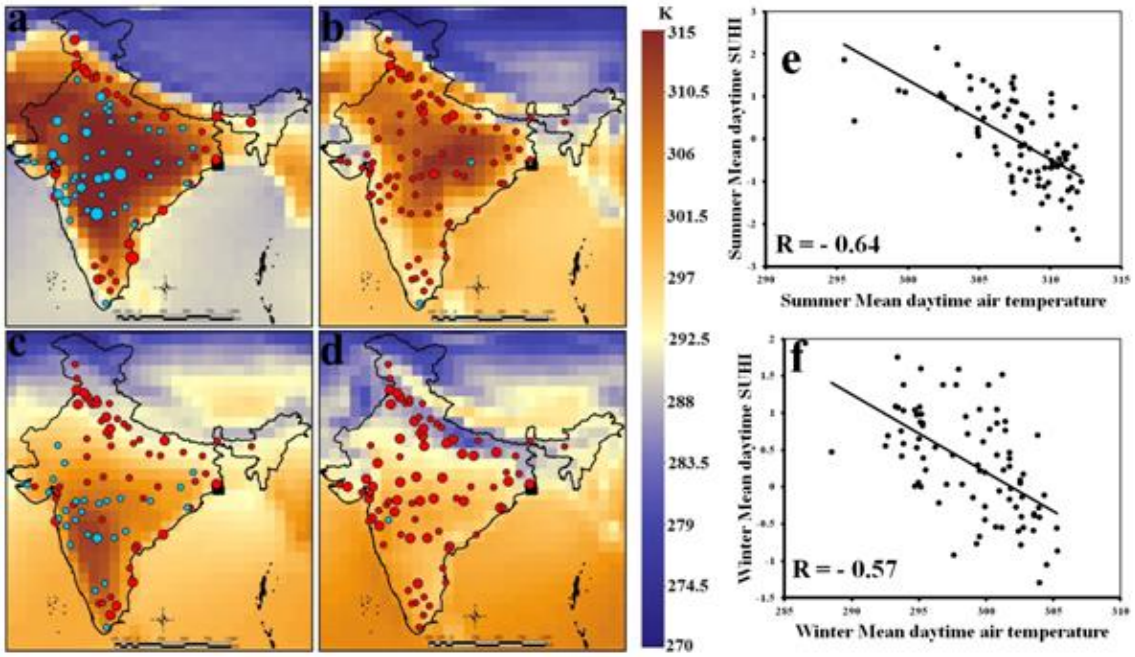


Figure 2.8 Attribution of SUHI to the surface level air temperature (Tair) indicated by large scale circulation pattern derived from the ERA Interim reanalysis data

a. Summer daytime T_{air} overlaid with the corresponding SUHI **b.** Summer nighttime T_{air} overlaid with the corresponding SUHI **c.** Winter daytime T_{air} overlaid with the corresponding SUHI **d.** Winter nighttime T_{air} overlaid with the corresponding SUHI e**.** Dependence of Summer daytime SUHI on T_{air} f**.** Dependence of Winter daytime SUHI on T_{air}

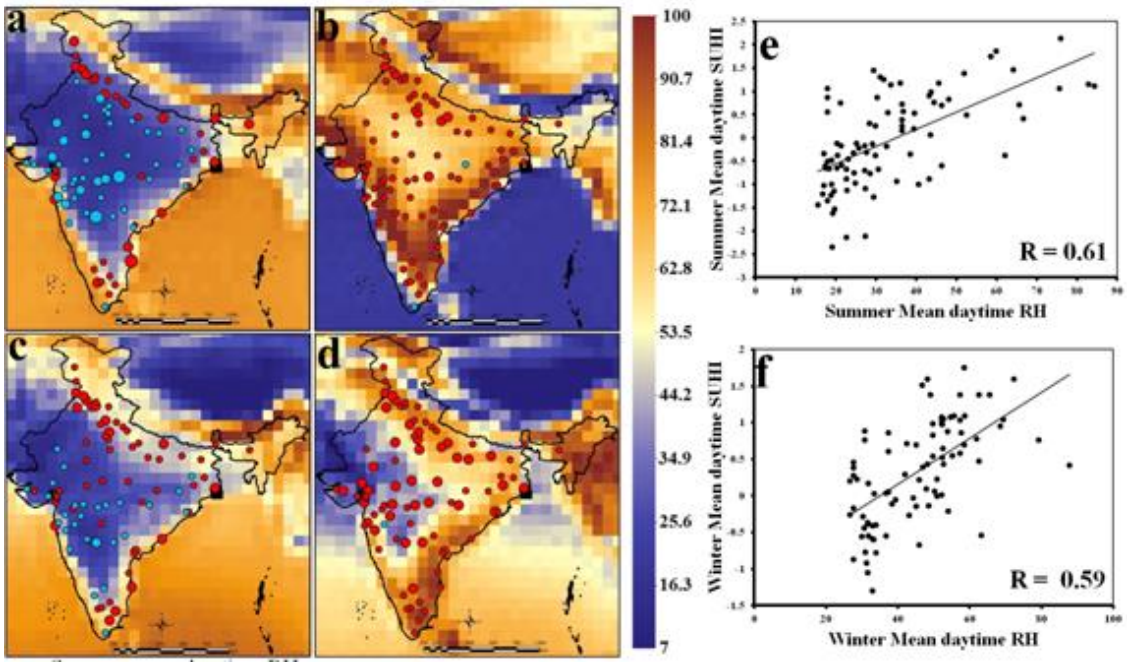


Figure 2.9 Attribution of SUHI to the surface level Relative Humidity (RH) indicated by large scale circulation pattern derived from the ERA Interim reanalysis data

a. Summer daytime RH overlaid with the corresponding SUHI **b.** Summer nighttime RH overlaid with the corresponding SUHI **c.** Winter daytime RH overlaid with the corresponding SUHI **d.** Winter nighttime RH overlaid with the corresponding SUHI e**.** Dependence of Summer daytime SUHI on RH f**.** Dependence of Winter daytime SUHI on RH

We again observe that the hot daytime climate also corresponds to the drier atmosphere with the low relative humidity. In the night time the coastal winds increases the humidity over the coastal belt of southern India over both eastern and western coast. The increase in humidity is observed to relate with the night time lower SUHII of the coastal cities. We find an overall correlation of 0.61 and 0.59 for the daytime SUHII and RH for the summer and winter season respectively.

We further plot the differences in LST between the urban and non-urban regions, when the temperature attains seasonal maxima for both summer and winter, separately (Figure. 2.10). Each of the box plots presents each city for the corresponding season. We observe that during summer, more than 50% of urban regions, the SUHII is negative when temperature is maximum for that season.

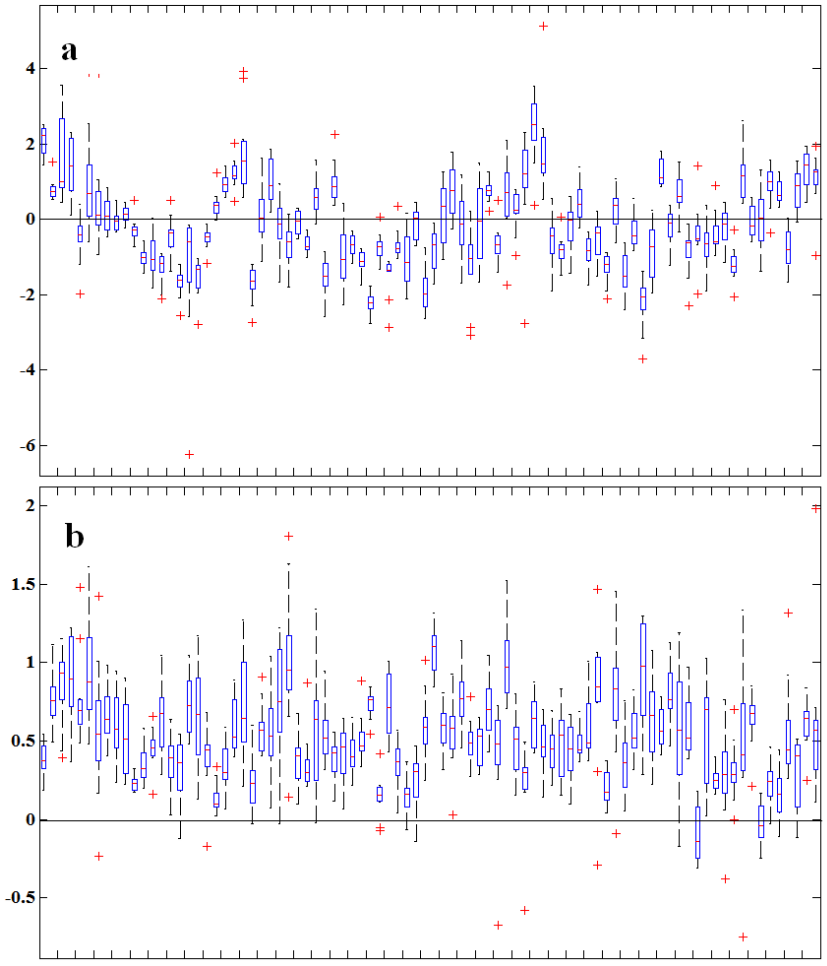


Figure 2.10 The estimated difference of daytime annual maximum temperature over urban and nearby non urban region

For the Summer (a) and Winter (b) season reveals a higher extreme temperature occurrences over the nearby non-urban region in the Summer season.

This brings out an important conclusion that the intensities of heatwave are less for majority of urban regions in India, compared to nearby non-urban regions. This is not in agreement with our general understanding of urban climate and this is primarily due to low vegetation cover in non-urban regions in pre-monsoon dry summer. In winter, the urban LST is higher than non-urban regions.

2.6 Summary

Urban areas are reported to have a warmer temperaturein comparison to its surrounding rural areas; this phenomenon is recognized as the UHI effect. The formation of UHI has been shown to have a seasonal and diurnal variation, for example UHI is stronger at the nighttime than during the daytime[1,2] and also stronger over the summer season than the winter season[2,3].UHI is caused by an increase in population density which results in change in the physical characteristics of the surfaces, anthropogenic heat, and high level of pollutants that alter the radiative nature of the atmosphere[4,5]. Presently, India has the second largest urban population of the world and is projected to have the highest urban growth rate in the next 30 years[6].Here we report results from the first assessment of seasonal and diurnal characteristics of SUHI over 84 large urban centers of India. The mean daytime SUHI over the winter season (0.3 ± 0.7 °K) is observed higher than the daytime SUHI over the summer season(-0.06 ± 1.0 °K). A similar observation is made for the nighttime SUHI measuring (0.9 ± 0.4 °K) over the winter season that remains higher than summer season (0.6 ± 0.3 °K). The distribution of nighttime SUHI positively correlates with the population strength of the city however the same is not observed with the daytime SUHI. This suggests that factors other than anthropogenic forcing dominate the driving of day time SUHI. The novel SUHI seasonality observed in terms of cool SUHI over summer season is largely explained with the regional meteorology, vegetative condition and BC emission.The distribution of daytime SUHI across cities negatively correlates with the

air temperature, while it positively correlates with the humidity; observed by the large scale circulation pattern. We also find a positive correlation between daytime SUHI and vegetative land cover over surrounding non urban region. The investigation reveals a higher depletion of vegetative land cover over the surrounding non urban region than the urban region during the summer season. The possible change in the radiation balance is also linked to the increased atmospheric concentration of BC in the surface layer over the winter season. The key findings of the study lies in unique set of SUHI characteristics and a higher diurnal temperature range of the Indian urban areas. Heat-waves that are caused by the UHI effect have profound impact on the quality of life of the large volume of urban population of the country[7]. Higher LST in non-urban regions during summer day results into higher intensities of heatwaves compared to those in cities as opposed to the presumptions made in literature. These observations highlight the need for re-evaluation of SUHII in India before adaptation planning to heat waves or usual practicing to link them with intensified urban precipitation.

Chapter 3 : URBAN PRECIPITATION IN INDIA

3.1 Introduction

Numerous studies carried out over different parts of the globe show changes in the daily precipitation characteristics as well as the frequency of tropical cyclones under warming condition (Hennessy et al. 1997; Trenberth et al. 2003). Precipitation is also reported (G. A. Meehl 1994, Xue et al. 2004) to be affected by the feedback from land surface processes. Observational studies (Huff and Changnon 1972, Changnon and Westcott 2002) reveal that increase in urbanization also has significant impacts on extremes. Increasing observational evidence of significant impacts of urban land cover on precipitation is available in literature. Braham (1981) showed that a higher thermal instability can exist over urban regions compared to nearby rural areas which can affect thunderstorms Impacts of urbanization on precipitation were also observed in Mexico (Jauregui and Romales 1996), Atlanta (Rose et al. 2008), Houston (Burian and Shepherd 2005), Guangzhou (Meng et al. 2007) and Taiwan (Lin et al. 2007). Mishra et al. (2012) analysed 100 most populous urban regions in pair with surrounding non–urban areas and observed a spatially mixed picture for precipitation related changes across the U.S., affected by urbanization. Ganeshan et al. (2013) has analysed spatial rainfall anomalies for several US cities; the study has observed a dominant influence of UHI on precipitation for both inland and coastal cities.

While a number of studies provide observational and numerical model based evidences of the impacts of urbanization on the regional urban precipitation, indifferent regions around globe very few case studies are available on the same for India. Goswami et al. (2006) observed increase of extreme rainfall events over Central India, in the last 50 years. However, such spatially aggregate increases were not supported by field significance tests performed by Ghosh et al. (2009) and Krishnamurthy et al. (2009). Increasing spatial variability of extremes, as observed by Ghosh et al. (2012), pointed to the need of systematic examination of global versus regional drivers of trends in rainfall extremes over India. Vittal et al. (2013) has pointed out that the changes in rainfall extremes over majority of the urbanized regions in India have occurred after 1975, when urbanization has started intensifying.

3.2 Motivation and Objectives

Though a significant number of studies are available on relating urbanization to intensification of rainfall, still there exists considerable debate on influences of urbanization versus large scale

forcing on changes in extremes, and also on the pattern of changes of extremes in urban areas over India. This study is conceived as to analyze the changes of Indian rainfall extremes at different urban locations of the country and nearby non–urban regions, to test, if the changes are solely due to urbanization or affected by regional synoptic scale changes. We also develop a statistical experimental technique that identifies the changes in association between the synoptic scale weather pattern and rainfall at different quantile levels, due to urbanization. Two main objective of the current study is

i) To check the trends of extremes of summer monsoon rainfall in urbanized India.
ii) Using a data driven model for understanding the changes in rainfall pattern in urban and nonurban areas using quantile based regression.

3.3 Data Used

The study utilizes four different datasets namely Urbanization data, APHRODITE gridded precipitation data, Climate reanalysis data and Station based observed precipitation data. A brief description and specification of these datasets is as below.

3.3.1 Urbanization dataset

To identify the urban areas in India, census 2001 and 2011 urban population data from official census reports, are used (http://www.censusindia.gov.in). The boundaries of selected urban areas are extracted from Moderate Resolution Imaging Spectroradiometer (MODIS) data product MCD12C1 at spatial resolution of 500 m with the 2010 classified data (Friedl et al. 2010).

3.3.2 APHRODITE gridded precipitation Data

Asian Precipitation Highly Resolved Observational Data Integration towards Evaluation of Water Resources (APHRODITE)), Japan is a daily gridded precipitation data set created from 1961-2004. The rainfall product is based on data collected from dense network of daily rain gauge data from all across Asia including the data from sparse areas like Himalayas and Mountainous areas of Middle East. The state of art of daily precipitation data is available at $0.5^{o}X0.5^{o}$ and $0.25^{o}X0.25^{o}$ resolution. They are based on (i) GTS (Global Telecommunication system) data (ii) data compiled by the organizations from the respective countries like, India Meteorological Department (IMD) for India, and (iii) APHRODITE's own data collection system, after proper quality control (Yatagai et al., 2012). The Aphrodite body has used an improved interpolation scheme which gives proper weightage to local topographical features to improve the orographic precipitation (Yatagai et al 2009). The APHRODITE data is derived

not by direct interpolation of station data but by interpolating the ratio of the daily precipitation to the daily climatology (Yatagai et al 2012). The gridded ratio is then multiplied by gridded climatology for each day to obtain gridded daily precipitation. The website address http://chikyu.ac.jp/precip. The data set is over 40 years so that it can be used for evaluating the long term water resources of Asian region. For the present analysis, we consider, the gridded daily precipitation data, compiled by the Asian Precipitation–Highly Resolved Observational Data Integration Towards Evaluation of the Water Resources (APHRODITE) version APHRO_V1003R1, for the period 1951- 2007, at a resolution of 0.25°×0.25° (Yatagai et al. 2009).

3.3.3 Observed Precipitation data.

Daily precipitation data of stations, Mumbai (Station name: Snatacruz) and Alibaug from January 1969 to December 2008 is obtained from the IMD. The meteorological stations Santacruz and Alibaug are situated at elevations of 14 m and 7 m respectively. The difference of elevation is 7 m, which is too small to have impacts on the differences in rainfall patterns between these stations. The quality control of the station data has been performed by India Meteorological Department.

3.3.4 Climate Reanalysis data

Reanalysis data is surrogate for observed data for any predictor variable. The NCEP-NCAR Reanalysis data set is a continually updating gridded data set representing the state of the Earth's atmosphere, incorporating observations and numerical weather prediction (NWP) model output dating back to 1948. It is a joint product from the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR), NOAA. The NCEP/NCAR reanalysis-I data (Kalnay et al., 1996) provide global atmospheric data which is a mixture of physical observations and model forecasts using different data assimilated systems such as global rawinsonde data, aircraft data, satellite data, and surface land synoptic data, advanced microwave surface wind speed data etc. to at 28 vertical sigma levels to calculate the reanalysis data products for various climate variables.

The atmospheric circulation patterns of western coast are obtained from the NCEP/NCAR reanalysis project. Gridded data for six predictors used in regression model formulation are obtained at 2.5°latitude×2.5°longitude spatial resolution for time period 1969 to 2008. The variables; namely atmospheric pressure, air temperature, relative humidity and vertical wind field, as well as 500 HPa pressure level eastward and northward wind field are obtained. The pressure level of 500 HPa is selected, considering it as the representative level for mean steering flow of convective storms (Hagemeyer 1991). The areal extent of parameters is delimited by10°–30°N in latitude and 60°–80°E in longitude. This region of predictors span is selected

to account for physical processes over large part of the Arabian Sea and Indian sub continental land, towards summer monsoon activity around selected metrological stations. Rainfall in both coastal stations Mumbai and Alibaug is essentially caused by the moisture flux coming from Arabian Sea. Hence, we have selected the predictor region slightly biased to the Arabian Sea.

3.4 Impacts of Urbanization on Regional Trends of ISMR

With its size and meteorological non–homogeneity, for analyzing ISMR characteristics, we consider seven meteorologically homogeneous zones of India (figure 3.1), as identified by the IMD (Parthasarathy et al. 1996). 42 urban areas, identified for the analysis are given in Table 3.1.

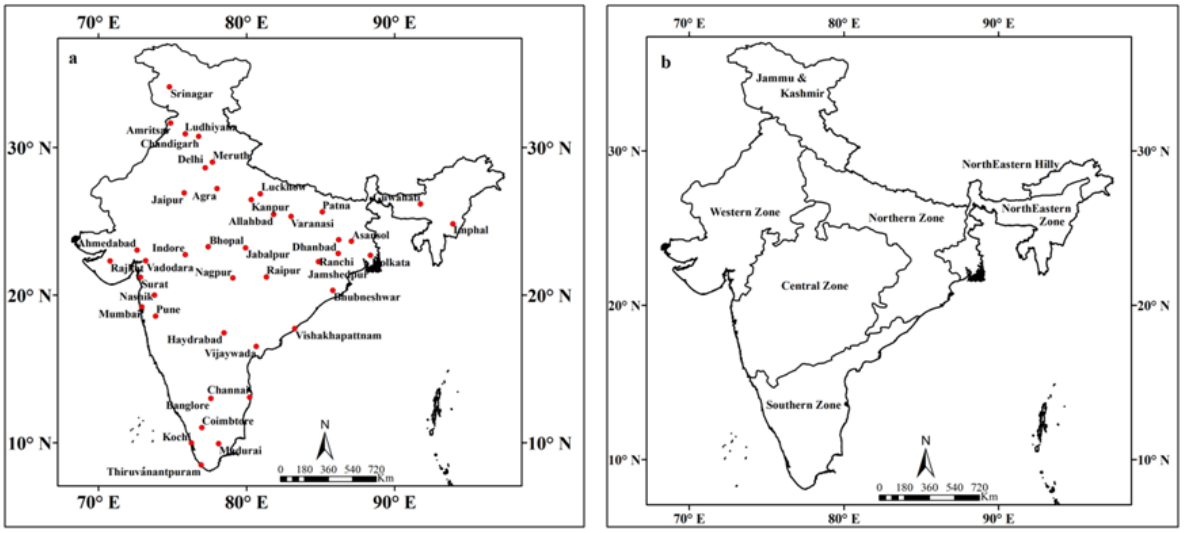


Figure 3.1 Urban locations (a) and meteorologically homogeneous regions (b) in India.

To analyse the characteristics of regional Indian Summer Monsoon Rainfall, the indices: mean rainfall, maximum 1–day precipitation, maximum 5–day precipitation, occurrences of heavy precipitation and ratio of heavy to non heavy precipitation (Caesar et al., 2006; Mishra et al., 2010; Hertig et al. 2013) are derived from gridded daily precipitation data. Annual time series of selected indices and mean precipitation are derived for all the grid points over entire India. Long term trends are estimated and tested at 5% field significance level of the indices are estimated with non parametric Mann Kendall rank statistics (Mann, 1945; Kendall, 1975). Magnitude of trends are estimated using Sen's slope estimator (Sen 1968). Figure 3.2 provides the spatial distribution of the trends for the above mentioned indices, where the null hypothesis of no–monotonic trend is rejected at the 5% significance level for the time period 1951-2007. To understand the influence of urbanization on the characteristics of precipitation, an urban

influence region is then selected by constructing buffer regions of 100–km radius around each urban area.

Table 3-1 List of selected urban areas of India

Sr. No.	Name of urban agglomeration	Sr. No.	Name of urban agglomeration
1	Agra	22	Kochi
2	Ahmedabad	23	Kolkata
3	Allahbad	24	Lucknow
4	Amritsar	25	Ludhiyana
5	Asansol	26	Madurai
6	Banglore	27	Meruth
7	Bhopal	28	Mumbai
8	Bhubneshwar	29	Nagpur
9	Chandigarh	30	Nashik
10	Channai	31	Patna
11	Coimbtore	32	Pune
12	Delhi	33	Raipur
13	Dhanbad	34	Rajkot
14	Guahati	35	Ranchi
15	Haydrabad	36	Srinagar
16	Imphal	37	Surat
17	Indore	38	Thiruvananthapuram
18	Jabalpur	39	Vadodara
19	Jaipur	40	Varanasi
20	Jamshedpur	41	Vijaywada
21	Kanpur	42	Vishakhapattnam

Figure 3.3 provides the summary of the findings in terms of percentage to investigate and present the urbanization impacts on rainfall at different climatologically homogeneous zones. Here, we compute the percentage of grid points that belong the urban buffer region in a zone (dotted line in each panel of figure 3.3), and the percentage of grid points showing increasing

and decreasing trend (bar diagram in figure 3.3). The results clearly indicate that the impacts of urbanization on precipitation are different in different regions and it largely depends on geographic locations.

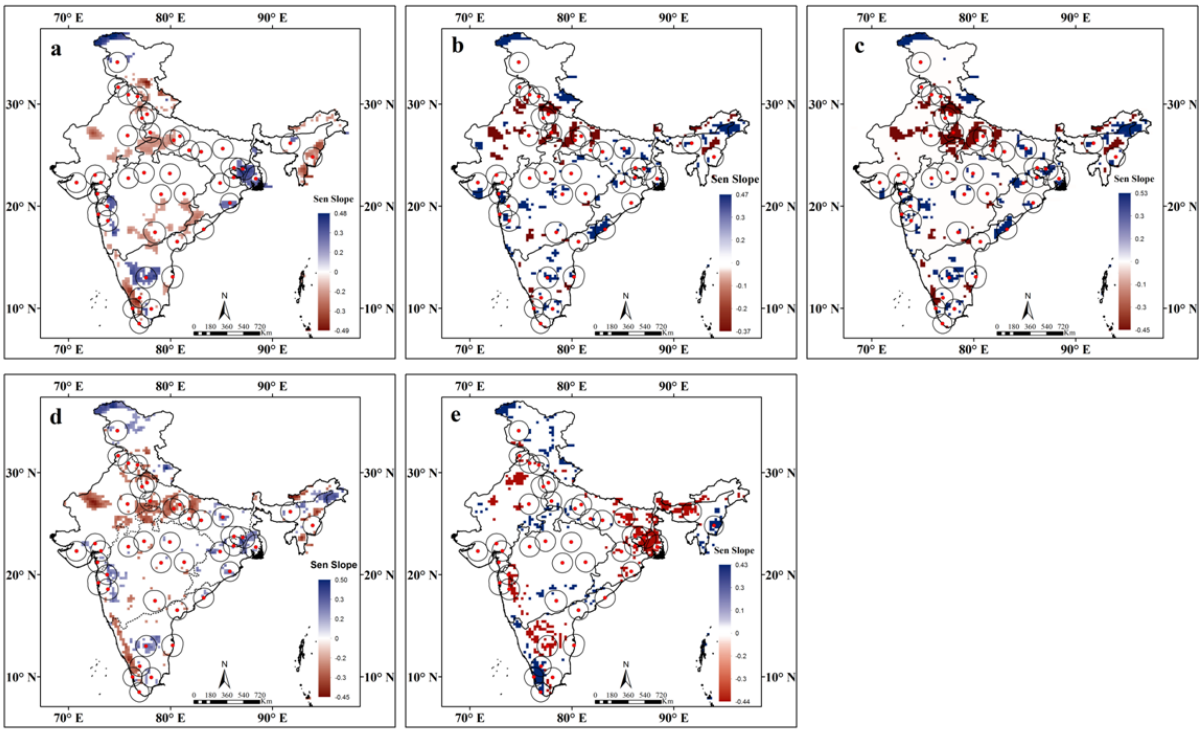


Figure 3.2 Trend of rainfall characteristics in India at 5% significance level.

The characteristics considered here are (a) mean monsoon (b) annual maxima (c) intensity of 5 most extreme events (d) occurrence over 95th percentile (e) heavy to non–heavy precipitation ratio. Locations of the centre of urban regions are marked with red dots and are bounded by a circle of radius 100 km, indicating the urban region. The interim lines indicate boundaries of climatically homogeneous zones of the country. Trend magnitudes estimated using the Senslope estimator.

The dotted line in each panel of Figure. 3.3, represent the urban fraction in a zone/ country. The bars in each plot show the urban percentage among the grid points having increasing and decreasing trend of rainfall characteristics. Urban signatures on intensification of extremes are prominent in western and central zone.

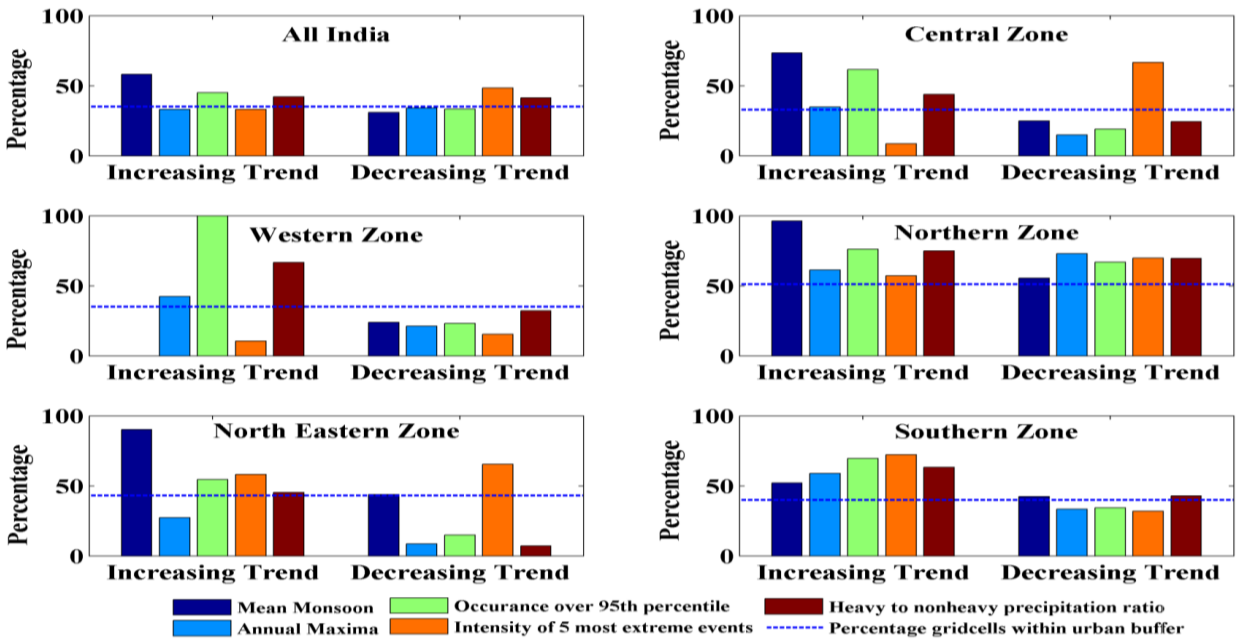


Figure 3.3 Urban percentage among the grid points having increasing and decreasing trend of rainfall characteristics.

Growth of urbanization has not been constant during post 1951 period and has intensified after 1975 (Vittal et al., 2014). To understand the association of trends of extremes with the growing rate of urbanization, we divide the period of data availability, into two halves. The second half, i.e., the period 1979-2007 is considered to have higher rate of urbanization as compared to years 1951-1978. Figure 3.4 provides the spatial distribution of the trends for the selected five indices during the two periods separately. Overall the trends of extreme indices are found to be higher during 1979-2007 than those during 1951-1978. Figure 3.5 provides the summary of the findings in terms of percentage to examine the impacts of urbanization growth on rainfall at different climatologically homogeneous zones, which is similar to figure 3.3.

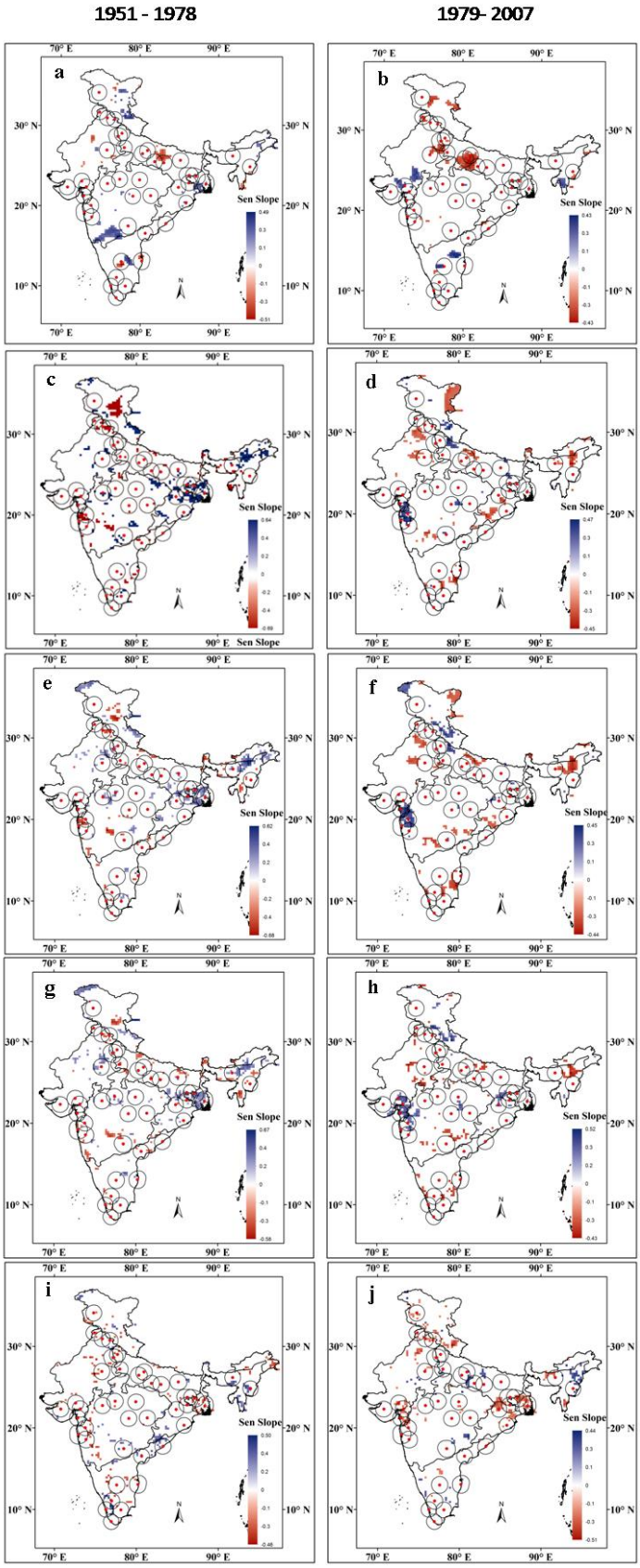


Figure 3.4 Trend of rainfall characteristics in India at 5% significance level for the time period 1951-1978 (Panel 1) and 1979 – 2007 (Panel 2).

(a)(b) mean monsoon (c)(d) annual maxima (e)(f) intensity of 5 most extreme events (g)(h) occurrence over 95th percentile (i)(j) heavy to non–heavy precipitation ratio. Same as figure 3.2.

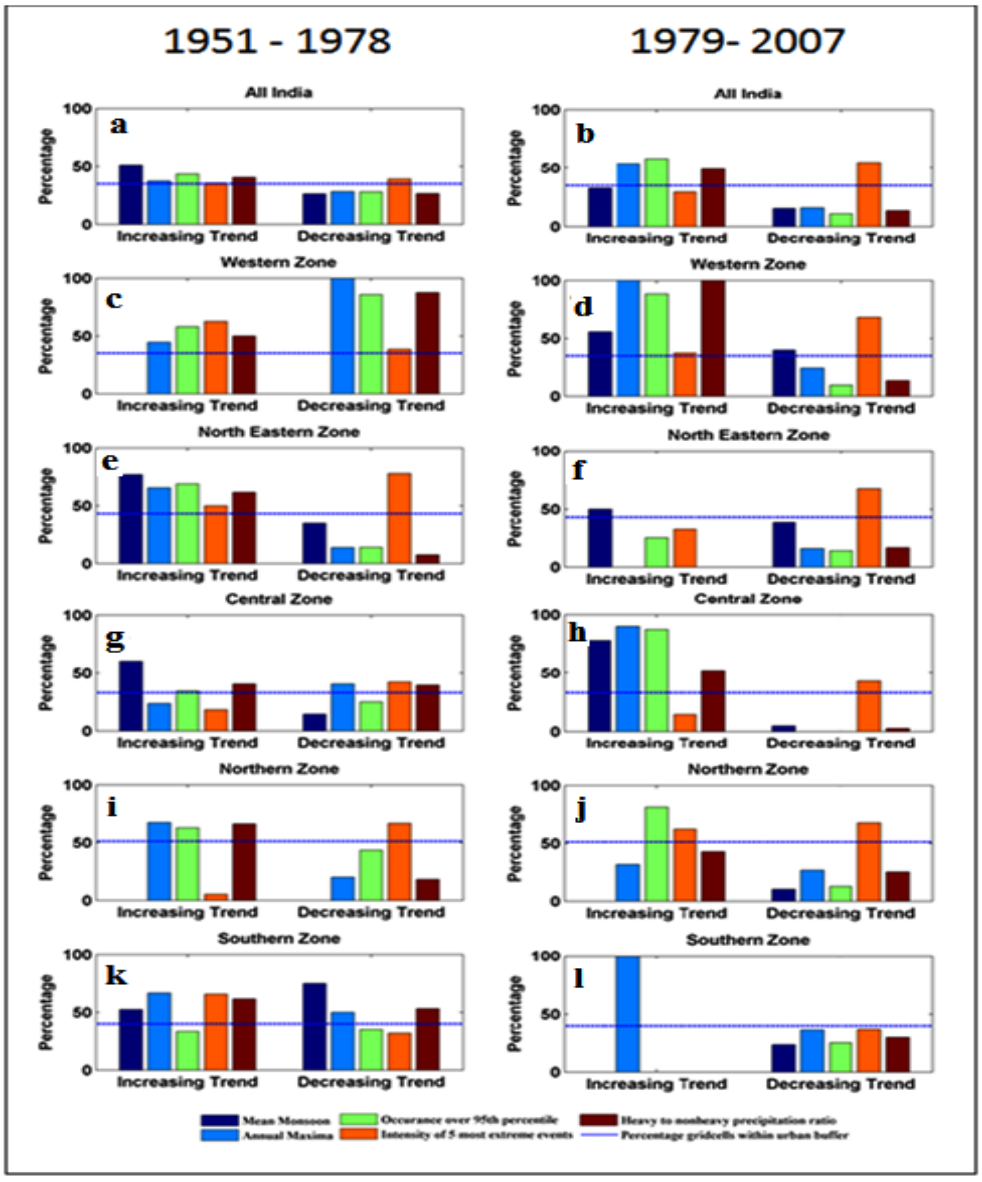


Figure 3.5 Urban percentage among the grid points having increasing and decreasing trend of rainfall characteristics for the time period 1951-1978 and 1979 – 2007.Same as figure 3.3

We observe that urbanization signature on increasing trend of rainfall extremes is not prominent during 1951-1978, but it is distinctly visible during the later period 1979-2007, when the growth of urbanization is high. Central and Western zone show maximum impacts of urbanization in terms of increasing trend of extremes for almost all the indices (figure 3.5 e, h). To understand the mechanism of intensification of extremes by urbanization, here we take a specific urban case study of most populated coastal city in India, Mumbai its surrounding non–urban area Alibaug located in Central zone and very close to Western zone, where the urban signature on intensified rainfall is prominent. The details are presented in the following section.

3.3 Urbanization Impacts on Rainfall Extremes over Mumbai

A statistical relationship is established between NCEP variables and precipitation with a specific emphasis to model extreme rainfall, i.e., the rainfall with higher quantile. We perform the regression analysis with 99, 95, 90, 85 and 80 quantiles of precipitation for both Mumbai and Alibaug. The methodology is presented in figure 3.6.

Six climate variables, at 81 grid points, result in high dimensionality, which is reduced with Principal Component Analysis (PCA). First we attempt to understand the large scale circulation types (classes), over western coast during monsoon using unsupervised classification technique, k–means clustering (McQueen 1967). Predictor data is divided into four groups with respective cluster ID. The climate variables corresponding to the four cluster centroids are presented in figure 3.7. Differences are observed between the cluster centroids for all the variables that show the importance of classification to understand the existence of multiple circulation patterns. The lower mean wind velocity over cluster1 (figure 3.9) is indicative of favourable weather condition for development of UHI (Szegedi and Kircsi 2003, Blazejczyk et al. 2006). At the same time high pressure region over land for cluster1 suggestive of a favourable condition for UHI impacts (Morris and Simmonds 2000).

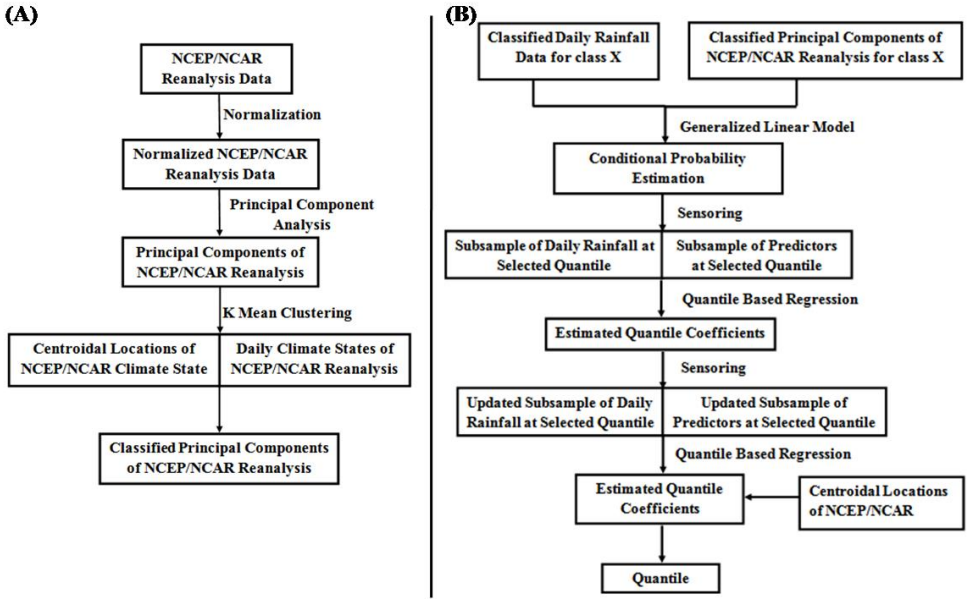


Figure 3.6 Methodology flow diagram

Data preparation (a) and Quantile regression approach (b).

The rainfall distributions of Mumbai and Alibaug, corresponding to the clusters of circulation patterns/types, are presented in figure 3.8. We divide the data for availability period, 1969–2008, in two equal halves. The second half, i.e., years 1989–2008 is considered to be more urbanized as compared to years 1969-1988. The distribution of circulation classes in both the periods show maximum changes in cluster 1, with a decrease of 4% from 1969–1988 to 1989–2008 (figure 3.7). However, most of the extreme metrics show increasing trend near Mumbai (figure 3.4), and such observation necessitates a detailed analysis to understand the changes in high quantile rainfalls corresponding to the same circulation pattern.

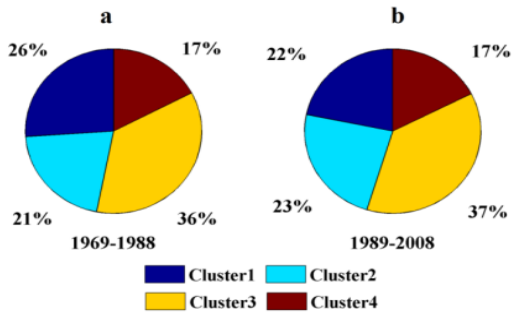


Figure 3.7 Distribution of circulation patterns during the periods 1969–88 and 1989–2008.

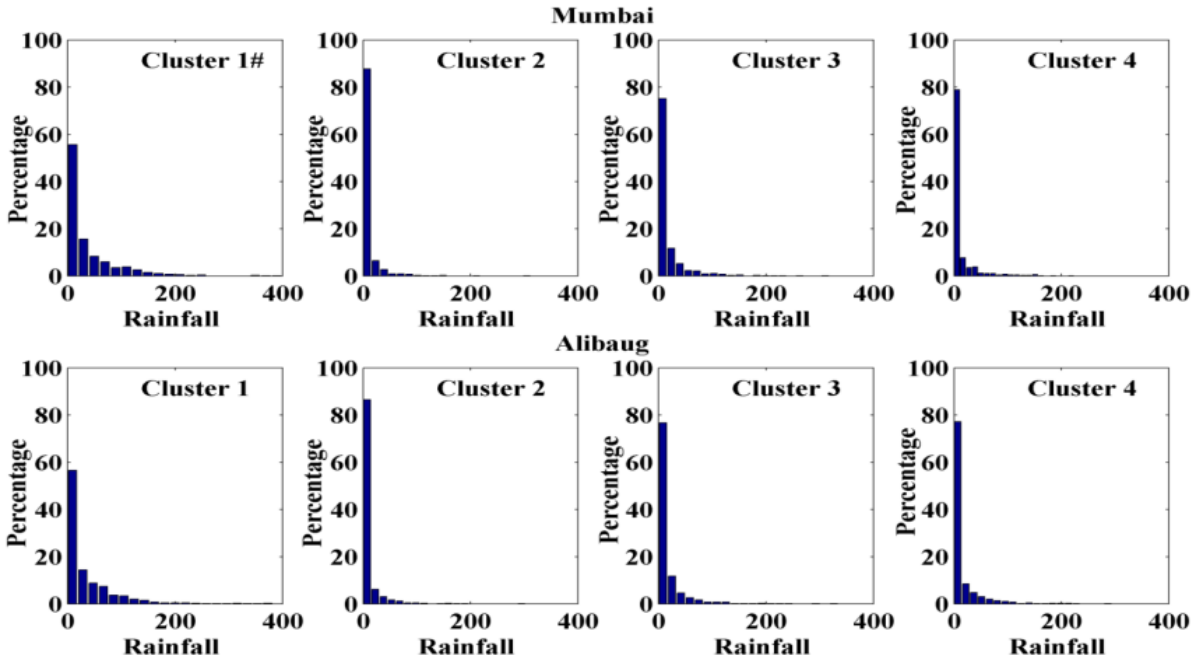


Figure 3.8 Rainfall distribution in Mumbai and Alibaug

As quartile regression is specifically designed for higher quantiles (which are supposed to have low sample size) and has been used in literature (Friederichs and Hense 2007) for extreme downscaling, we use the same for this purpose. Quantile Verification (QV) score, proposed by Gneiting and Raftery (2005) is used for validation of QR. Figure 3.10 provides skill scores of

proposed censored QR model ($CQVSS_{\tau}$) with, linear regression model outputs as reference. The results show an improved model performance for higher quantiles.

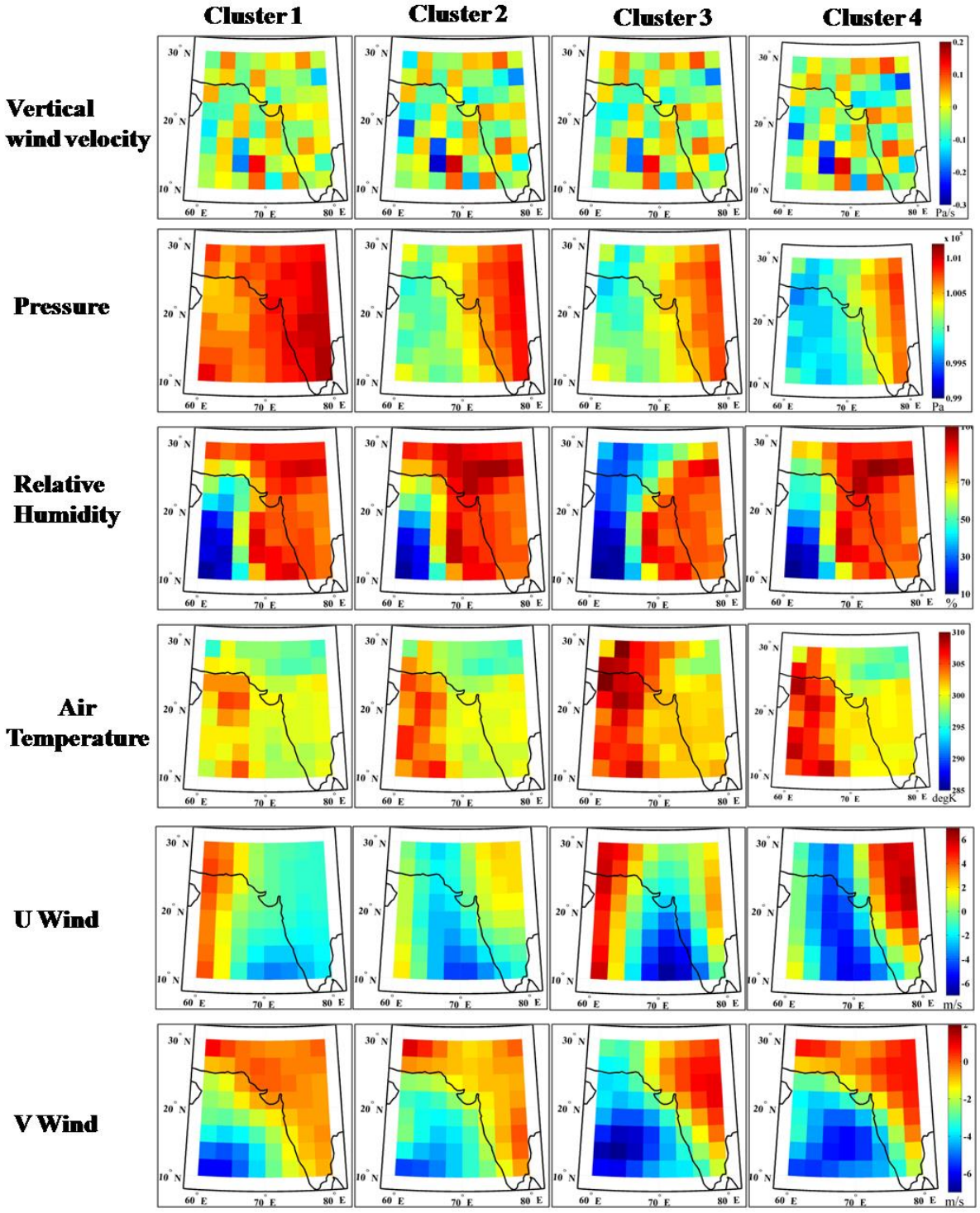


Figure 3.9 Centroids of each cluster/ circulation pattern for different climate variables.

The quantile values estimated for all the circulation patterns, and both time periods for Mumbai and Alibaug are given in figure 3.11.

We observe that for cluster 1 (which is associated with extremes) in Mumbai, the circulation pattern in post urbanization period results into high rainfall quantiles as compared to those of pre–urbanization period.

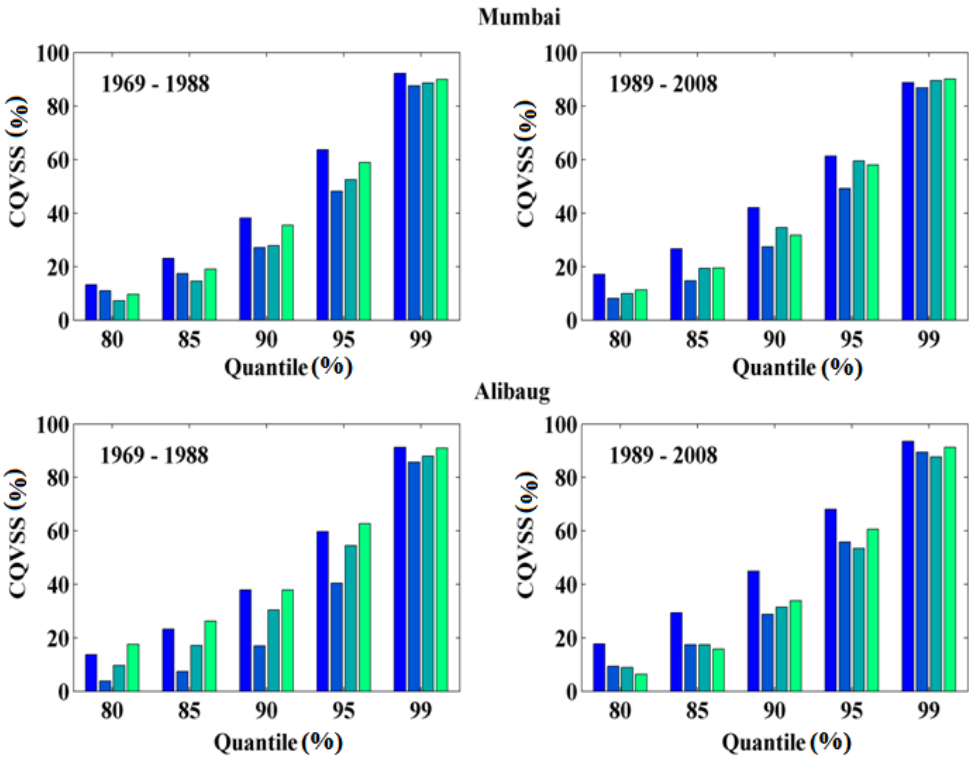


Figure 3.10 Quantile regression model performance

CQVSS for Mumbai and Alibaug, corresponding to different circulation patterns, during periods 1969–88 and 1989–2008.

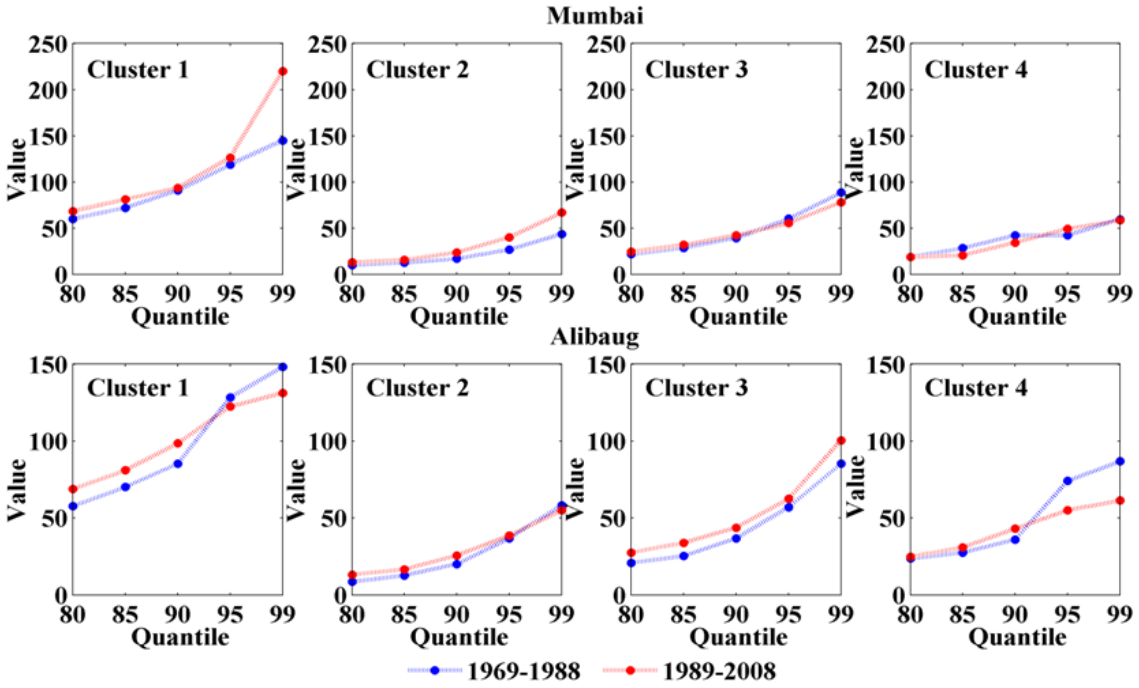


Figure 3.11 Rainfall quantiles in Mumbai and Alibaug,

The quantile values estimated corresponding to different circulation patterns, for periods 1969–88 and 1989–2008.

The same is not observed for non–urbanized area Alibaug. This indicates the possibility of changing relationship between the extreme precipitation of Mumbai and atmospheric circulation pattern, resulting higher rainfall quantiles in post urbanization period.

3.4 Summary

This work demonstrates an overall picture of the impacts of urbanization on Indian summer monsoon rainfall extremes. The key finding lies in understanding the non–uniformity of these impacts on extreme rainfall across the country. We observe such non–uniformity mainly result from the zonal trends, the dominances of which, over local changes, are different in different regions. To investigate more, with station level rain–gauge data, here we develop a new data driven experiment technique for understanding the impacts of urbanization on extreme rainfall. The novelty of the present approach is that this is computationally inexpensive, as well as, capable of understanding impacts of urbanization on extremes. The computational time required for statistical data driven modelling is less as compared to dynamic analysis with WRF

for all extreme events from 1969-2008. The results distinctly show that urbanization has intensified the extreme rainfall in Mumbai, which is not visible in Alibaug. Our results highlight that urbanization affects the relationship between the large scale circulation and rainfall extremes, and hence calibrating a rainfall model for urban area with past data may not guarantee good simulations for future. The limitation of the model is that, quantile regression is still a black box model and hence it cannot identify the physical mechanisms, which are affected by urbanization and are responsible for changing the patterns of rainfall extremes. This needs follow on research activities in understanding the fingerprints of extremes (Nayak and Ghosh, 2012) and then identifying the changes in the fingerprints at local scale during extremes for pre and post urbanization periods. This may be considered as the potential area of future research. The increasing trend of extreme rainfall and population at the same time over the metropolitan of Mumbai highlights that development and improvement of real-time flood forecasting systems is a very topical for the urban centre of Mumbai. The effectiveness of early warning systems for reducing the damages and casualties induced by floods is widely recognized.

Chapter 4 : EXTREME RAINFALL FORECASTS IN MUMBAI WITH WEATHER PATTERN RECOGNITION

4.1 Introduction

Extreme rainfall is one of the most difficult elements of the hydrologic cycle to forecast because of the high variability in space and time and the complex physical process. Occurrences of rainfall extremes are expected to increase in changing climate (Goswami et al. 2006; IPCC 2012), and hence, proper scientific understanding of extremes is crucial. Though there are significant research advancements in the last two decades in the science of extremes (Cavazos et al. 2008; IPCC 2012; Wheater 2002; Young 2002) to minimize the impacts, hazards, and losses. Early prediction-based alert broadcasting may help in operating the existing flood control systems with maximum efficiency, which minimizes losses, such as evacuation, flow diversion, alerting the population, preparedness of the disaster mitigation team, etc. A warning system may alert the population for making their own arrangements to safeguard their lives and their properties. People under attack will be relocated well before the advancement of the flood. In addition, early warning may enable the preparation of international assistance actions (Pappenberger et al. 2008). An effective rainfall forecasts have great potential to improve planning and management of water resources and agriculture (Everingham et al., 2008). It is of a paramount importance as it serves the role of key input in formulating modus operandi for immediate future. Short range rainfall forecasts influence a wide range of entities, right from agricultural industry to a common man. The disastrous flood of Mumbai, in 2005, is one such example (Rajendra et al. 2006). For an urban centre like Mumbai housing metropolitan population of 20.7Million (Census 2011) an accurate forecasts may help in minimizing the possible damage by implementing pre-decided plan of action.

At the same time weather forecast has always been a challenging research problem. Extreme rainfalls are expected to be associated with exceptional atmospheric conditions. They are of short duration and are the consequence of convective instabilities in moist air in small spatial location (Goswami et al. 2006). To obtain more effective forecasts of extreme rainfall, novel models with better ability are desired.

Many climate modeling centers around the world now routinely produce long lead global rainfall forecasts from coupled ocean-atmosphere general circulation models (GCMs). Numerical weather prediction uses mathematical models of the atmosphere and oceans processes to predict the weather for near future based on current weather conditions. Mathematical models can be used to generate weather forecasts (short-term) or climate predictions (long term). Numerical prediction efficiency is affected by density and quality of observation input and deficiency in the numerical model. To enumerate the large amount of inherent uncertainty remaining in numerical predictions, ensemble forecasts have been used to gauge the confidence in the forecast, and to obtain useful results (Wei and Toth, 2003). This approach analyzes multiple forecasts created with an individual forecast model or multiple models.

The Global Forecast System (GFS) is a weather forecast model produced by the National Centers for Environmental Prediction (NCEP). The atmospheric and land-soil variables available from GFS data range from temperatures, winds, and precipitation to soil moisture and atmospheric ozone concentration. The GFS model is a coupled model, composed of four separate models (an atmosphere model, an ocean model, a land/soil model, and a sea ice model), which work together to provide an accurate picture of weather conditions. The output from the GFS is also used to produce model output statistics.GFS dataset is available at a resolution of (0.5°x0.5°). This dataset is used to predict weather out to 16 days in the future (source:http://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs).

It is also necessary to gauge the quality of forecasts which might vary with the complexity of weather state and regional parameters associated with Indian Summer Monsoon Rainfall (ISMR) for the amount of damage it can cause (because of poor forecasts) to the economy. Among the most crucial scientific interests the assessment of the forecast uncertainty also plays a major role. The two usual sources of uncertainty in forecast models: (1) the errors introduced by the use of imperfect initial conditions, and (2) errors introduced because of imperfections in the model formulation, such as the approximate mathematical methods to solve the equations (Wei and Toth, 2003). Multiple model simulations are conducted to account for this uncertainty. Ideally, the verified future dynamical system state should fall within the predicted ensemble spread, and the amount of spread should be related to the uncertainty (error) of the forecast.

Downscaling is the method to convert coarse resolution data to a finer resolution data. It can be broadly divided into two categories: statistical and dynamic downscaling. In dynamic downscaling, output of the GCM is used to drive regional, numerical model in higher spatial resolution, which is able to simulate local condition in a better way. In statistical downscaling, a statistical relationship is established between large scale climate predictors like surface pressure, temperature and local scale predictand like rainfall. Extreme rainfall forecasting plays an essential role in flood disaster warning systems. To obtain more effective forecasts of extreme rainfall, the development of better models has always been recognized as an important task. The NCMRWF produced rainfall forecasts for the country of India at 50 and 35 km spatial resolution based on physics-based models are found to be poor in simulating extreme rainfall in India (Khaladkar et al., 2007). One possible reason may be because Indian regions are non-homogeneous with respect to the land–ocean interaction, terrain distribution, and prevailing weather systems. This aspect of non-homogeneity in Indian regions makes it important to take into account the mesoscale conditions of the atmosphere, along with synoptic scale circulations, for the prediction heavy rainfall events. An attractive alternative to the physically based models is statistical models that work on the basis of information processing system with great flexibility in modelling nonlinear processes.

The selection of predictors directly affects rainfall processes are used as input variables in statistical downscaling procedure. The selection of predictor variables is important for the accuracy of model performance. Wilby (1999) described three basic rules for selection of the predictor variable: (1) the data for the particular predictor should be available for the desired period, (2) the selected variable should be well simulated by the model, and (3) the predictor should show a good correlation with the predictand. Numerous studies across the literature studies demonstrate the dependence of predictor selection over the performance of model (Hewitson and Crane 1996, Charles et al., 1999). For the current study, we use reanalysis data of ten climatic variables that signify state of atmosphere in predicting extreme rainfall over the station. The climatic parameters namely geopotential height, relative humidity, air temperature, eastward wind field (UWind), northward wind field (VWind) at the pressure level 1000 hPa and 500 hPa as well as surface level atmospheric pressure are selected as the predictor variables. The 500 hPa level is chosen as a representative level for mean steering flow for convective storms (Hagemeyer 1991).

4.2 Motivation and Objective

The weather forecast obtained with the Global Ensemble Forecasting System (GEFS) 16 day provides weather forecast of precipitation at resolution of 1° x 1°. This is not sufficient for finer resolution precipitation and flood forecasting for an urban region. The statistical properties, mean, standard deviation, correlation coefficient and extreme parameters of GEFS gridded precipitation data corresponding to the Mumbai urban grid is checked with the observed daily rainfall of station Santacruz of Mumbai.

The present study aims working out a statistical downscaling based numerical weather prediction model for the urban centre of Mumbai to improve GEFS rainfall forecast. Despite the fact that myriad previous studies have addressed the precipitation forecasting, the research has addressing the one important attribute of the rainfall i.e. the day when certain amount of rain is received; and its relation to the climatic circulation pattern is not taken up especially in context of urbanization development.

4.3 Study area

Mumbai city was chosen as the study area considering its rapid urbanization in the past two decades. Mumbai is situated at the mouth of the Ulhas River on the western coast of India. The city has a tropical climate, with seven months of dryness and peak of rains in July. The cooler season from December to February is followed by the summer season from March to June. Mumbai is the largest metropolitan region in India. Since the economic reform in 1978, this city has been experienced significant economic and population growth, owing to the economic radiation, high technology layout and favorable investment policies. It is the financial and commercial capital of the country as it generates 6.16% of the total GDP. Concomitant with significant economic development, the city has witnessed an extensive urbanization process and significant amount of land has been displaced by commercial, industrial, residential areas and redevelop idle areas. The Mumbai metropolitan region covers an area extent of about 1,500 km^2.

4.4 Data used

The training and testing of forecast model is carried out with a ground based dataset of annual precipitation at daily time period for meteorological station of Santacruz (Mumbai). The precipitation data is obtained for the time period 1979-2007 from IMD. The quality controlled rainfall data is extracted for southwest summer monsoon months; June-July-August-September (JJAS) from the annual time series.

The two important datasets namely the ERA interim reanalysis data and GEFS data sets used for statistical model formulation forecast generation.

4.4.1 Climate reanalysis data

Era-interim reanalysis data is used for predictors and daily observed rainfall of the station Santacruz is used as predictant. The selected climatic variables are obtained from the ERA Interim reanalysis dataset provided by European Centre for Medium-range Weather Forecasts (ECMWF) (Dee et al., 2011). The areal extent of parameters is delimited by latitudes 10°-30°N and longitudes 60°-80°E to account for physical processes, over the Indian subcontinent and Arabian ocean towards Indian summer monsoon activity around selected metrological station. The data is available on a 1° latitude 1° longitude grid. The baseline period considered in this study is from 1979 to 2007 (29years) based on the availability of the rainfall and reanalysis data. This duration is sufficient to establish a reliable climatology and is used for training and validating the proposed statistical model. The predictors used for the present study are temperature (surface level and 500hPa), pressure at surface level, wind velocities (u-wind and v-wind at 500hPa and 1000hPa), specific humidity (500hPa and 1000hPa), and geo-potential height at 500hPa. Figure 5.1 shows spatial extent of ERA-Interim reanalysis grid points (total 441 points, 21 × 21) over the map of India along with location of stations Santacruz (Mumbai).

4.4.2 Numerical Weather Forecast Data

Ensemble forecasting is a numerical prediction method that is used to attempt to generate a representative sample of the possible future states of a dynamical system. The Global Ensemble Forecast System (GEFS) is a weather forecast model made up of 21 ensemble members (Hammil et al., 2013). GEFS is initiated by NCEP to address the nature of uncertainty in weather observations, which are used to initialize weather forecast models (source:

http://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-ensemble-forecast-system-gefs). The second generation reforecast dataset of GEFS data consists of an 11-member ensemble of forecast. This dataset is available from December 1984 to present day. The operational medium-range GEFS data is available every 3-hourly intervals from 0 to 72 h and every 6 h thereafter. The data is procured from the global forecast fields available at 1° latitude 1° longitude resolution for the selected climate variables and areal extent. This resolution is reasonable for global model, but it is not able to capture fine resolution processes associated with precipitation. Hence, it does not give accurate information related to precipitation. For fine resolution rainfall, statistical downscaling is performed using large scale predictors.

The second generation reforecast dataset of GEFS data consists of an 11 member ensemble with 10 perturbed forecast members and the one control forecast. The forecasting is achieved by applying the tested regression model to the output from the NWF NCEP/GEFS datasets. The data for 1984-2007 (23 years) is used to validate the forecast model.

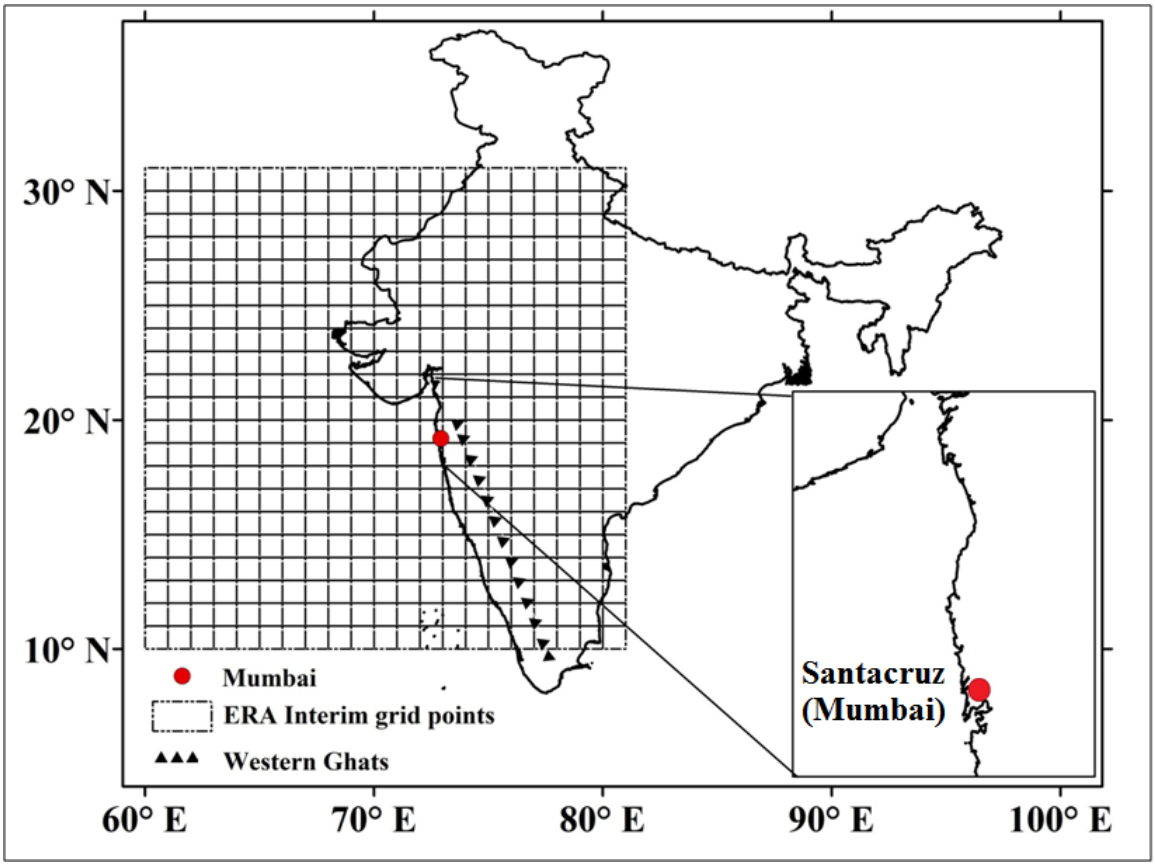


Figure 4.1 Location of Mumbai and spatial extent of predictors selected.

ERA interim grids overlaid on map of India.

4.5 Methodology

In this study, we refer to the extreme precipitation as those events observed at the tail of frequency distribution functions. Extreme event is considered when the daily precipitation exceeds 95th percentile of the time series. The overview of proposed methodological procedure used in this study for accessing quantile forecast values for extreme precipitation from a climatic circulation pattern is stepwise summarized in Figure 4.2. The three basic steps for formulation of regression data sets are as follows : (i) normalisation of the datasets; (ii) applying the dimensionality reduction technique (iii) bias correction. Quantile based regression is used to find relation between southwest monsoon precipitation over the station and climatic circulation pattern. The rainfall data available for 1979-2007 (29 years) is divided into two halves for model training and validation purpose. The first half of precipitation dataset 1979-1993 (15 years) is referred as the training period and the rear half of dataset 1994-2008(14 years) is referred as the validation period.

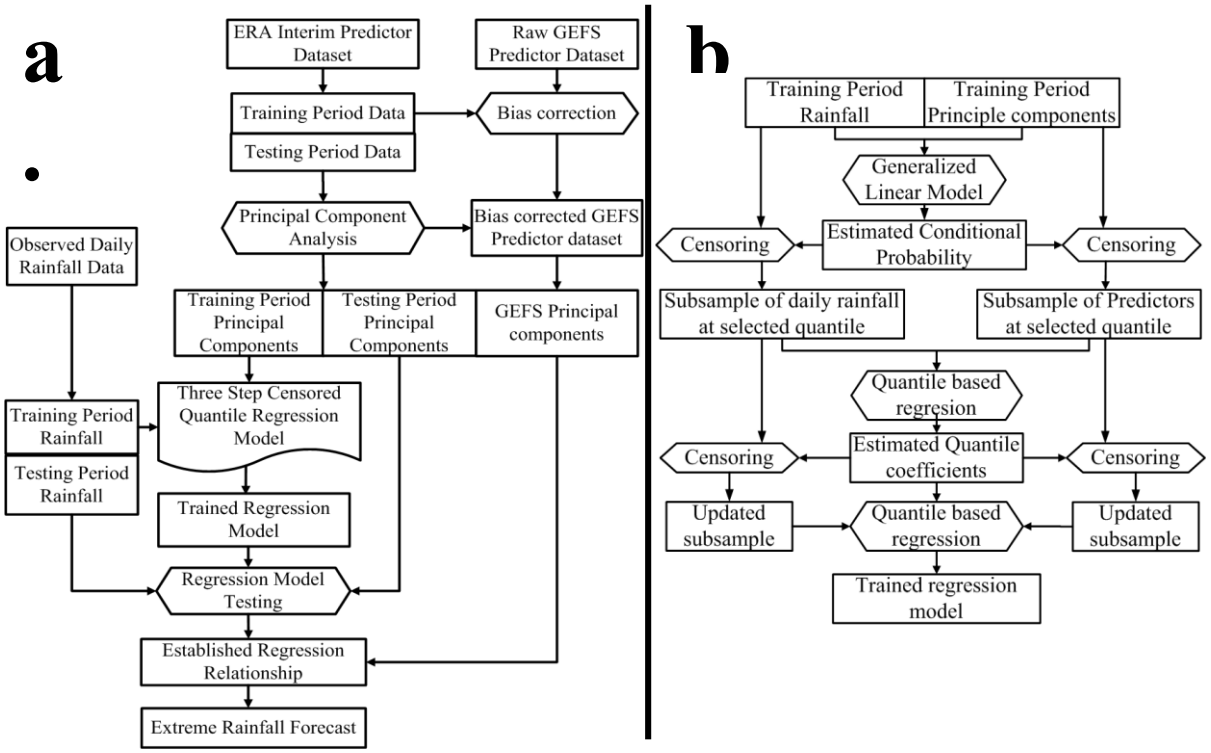


Figure 4.2 Statistical Downscaling Methodology

a. Flowchart for statistical downscaling model for extreme rainfall forecasting enlisting various mathematical operations that are performed on predictors (GEFS simulated climate variables) and predictand (rainfall) which take part in statistical downscaling as inputs.
b. Two step censored quantile regression model (Friederichs and Hense, 2006).

4.5.1 Forecast model formulation

The first step of the normalization of datasets takes care of adjusting values of selected predictor variables to a common scale. The normalization of data is essential, as the selected climatic variables are measured on significantly different characteristic scales and different dimensions. Selected 10 variables at 441 grid points around the station Santacruz(Mumbai) make a sum of 4410 variables to be utilised as predictors for regression analysis. However employing all these variables for regression analysis firstly poses the difficulty of multi-dimensionality. Secondly as the climatic variables are largely dependent on each other hence highly correlated pose the problem of multi-collinearity. Use of high-dimensional correlated data is computationally expensive and also leads to an increase in the degree of sparseness in data that may impact outputs where a statistical significance is required. On the other hand, if the dimensions are reduced without considering the internal data pattern and its variability, it hampers the accuracy of the model output. PCA is the powerful and most widely used multivariate statistical technique, for identifying patterns in multidimensional data set containing a large number of variables and reducing the number of dimensions to a data set containing fewer new variables representing a large fraction of the variability contained in the original data to the best possible extent. For the current study, PCA is used to obtain principal components (EOFs) acting as predictors to establish a statistical relationship of observed rainfall with the large scale circulation pattern. Preisendorfer (1988) investigated various principal component selection rules, with this study we have used Kaiser's rule adopted by Jolliffe (1972) to retain the principal components accounting for more than the average amount of the total variance present in the dataset. The standardized predictor containing 4410 variables (10 climate variables at 441 grid points) are reduced without discarding important information carried in the original data. It is also essential to preserve the eigen vectors obtained during PCA of the ERA-Interim data to obtain the principal components of validation time period as well as bias corrected GEFS data. We carried out different experiments retaining the predictors representing 98-60 percent of variability present in the set of complete predictor dataset. An additional experiment of conducing PCA over individual predictors is undertaken. We also apply an additional dimensionality reduction methodology Least Absolute Shrinkage and Selection Operator (LASSO). The method developed by Tibshirani (1996) selects variables on the basis of their effect on the response and reduces number of variables with a constraint of sum of absolute value of coefficients less than a constant. A detail description of LASSO is referred to the supplementary information. The list of experiments and the number of EOF retained with each experiment to be used as predictor variable is referred to Table 4.1.

4.5.2 Bias correction

An ideal conditions (zero bias) is GCM-simulated data be exactly the same as that of the reanalysis data. The ensemble mean from all the raw GEFS forecast data shows a systematic difference with the observed climate variables. This systematic deviation, known as bias, is firstly removed to obtain correct rainfall forecast. Standardization (Wilby et al., 2004) of GEFS predictor data is carried out prior to statistical downscaling to remove systematic biases in the means and variances of GEFS predictors relative to the observations or the ERA-Interim data. The procedure typically involves a subtraction of the mean and division by the standard deviation of the predictor variable over a predefined baseline period. The standardization of GEFS predictor data is carried out with the mean and standard deviation of ERA-Interim data predictor data of the same duration. The bias-corrected GEFS data is shown in Figure 4.3.

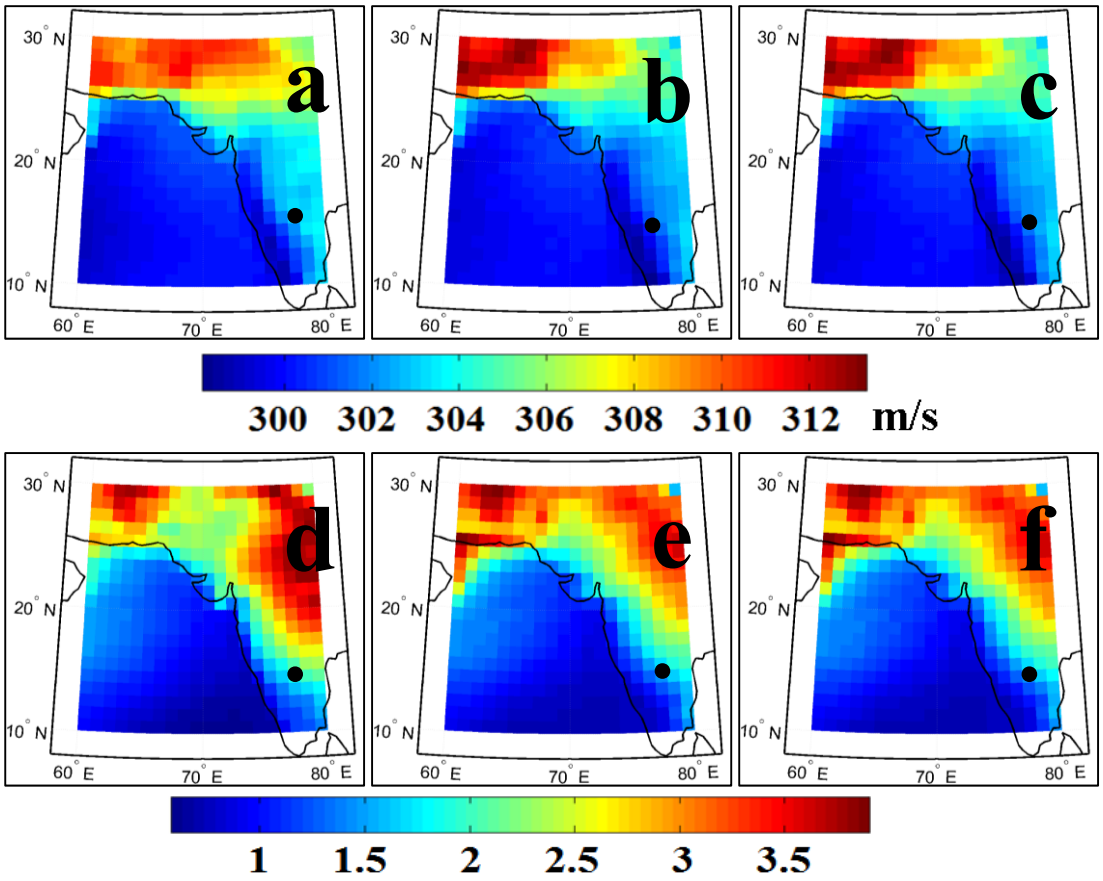
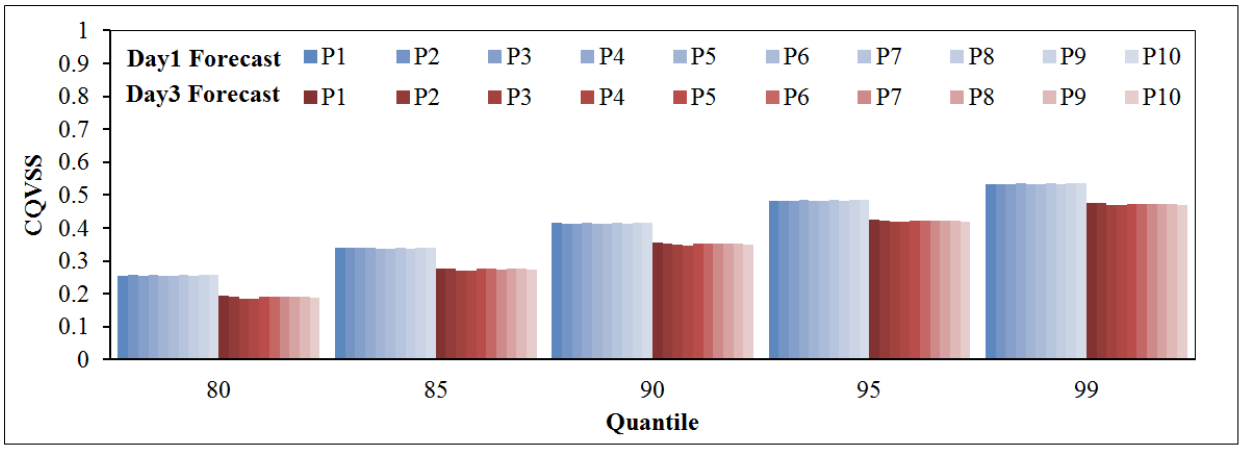


Figure 4.3 Bias correction methodology (standardization) applied to GEFS-simulated predictors.

Application of the method is demonstrated with u-wind. Uncorrected GEFS simulations 1000h Pa u-wind for 1985–2007, monsoon mean(a) and standard deviation(d) fail to capture the spatial distribution as observed with mean(b) and standard deviation(e). However, the bias

corrected simulations capture spatial variation both in terms of mean(c) and standard deviation (f).



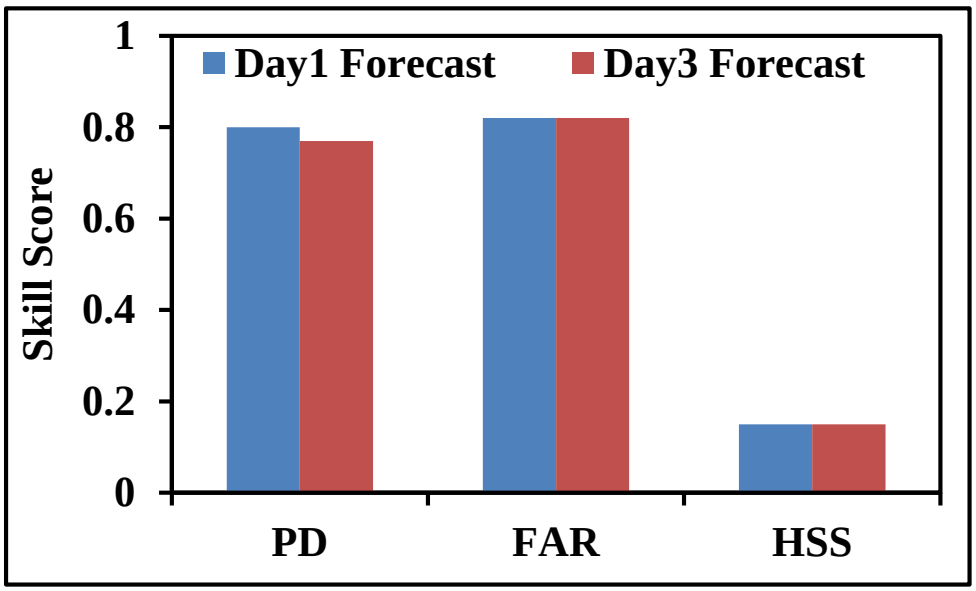


Figure 4.4 Skill score for GEFS forecast at lead time 1 and 3 days

Censored Quantile Verification Skill Score (CQVSS) against the quantiles (0.8 ,0.85, 0.9, 0.95, 0.99) (a), The shading indicates GEFS perturbations. Probability of Detection(PD), False Alarms(FA) Ratio and Heidke Skill Score(HSS), of extreme rainfall (at 95th percentile) corresponding to day 1 and day 3 forecast. The forecast skill reduces with lead time.

4.5.3 The three step censored quantile regression model

The conditional quantile model is trained on selected EOFs of the ERA Interim reanalysis data, and it is then applied to NCEP high-resolution GEFS 12-h forecasts. We use an estimate of conditional quantile coefficients using the subsample of data. Standard Quantile Regression module provided by R. Gilchrist (2001, p.1) describes a quantile as 'the value that corresponds to a specified proportion of an ordered sample of a population'. A clear definition of Quantile may be given as: the points taken at regular intervals in the cumulative distribution function of

a random variable. Introduced by Koenkar and Basset in 1978, quantile regression is an extension of the linear regression model. It is used to estimate quantiles of a response variable distribution conditional on a given set of predictors using a linear model. It provides a more complete statistical view than the classical expectation value regression. With quantile regression models the relationship between predictor variables and specific percentiles or quantiles of the response variable is established (Koenker 2005). We adopt the three step procedure proposed by Chernozhukov and Hong (2002) for estimation of quantile values. Friederichs and Hense (2007) used the same approach for statistical downscaling for extremes. In the first step conditional probability of the occurrence of precipitation π = prob(Y>0/X) is estimated using a generalized linear model (GLM) with a logit function (Fahrmeir and Tutz 1994). The estimated probability of precipitation denoted as $\hat{\pi}$ indicates probability of not censoring. Based on this estimate, a subsample J_0 is chosen with $J_0 = \{i:\hat{\pi}_i = prob(Y> 0/xi) >1-\tau\}$. Using the subsample J_0 an initial estimate of quantile coefficients β_τ is calculated in second step.. The quantile regression is performed using the R open source statistical computer program (R Development Core Team 2003) and the R quantreg-package of Roger Koenker. An updated subsample with $J_1 = \{i: \hat{\beta}_\tau^T x_i >0\}$ is selected. Finally with third step an estimate of the three-step estimator $\hat{\beta}_\tau$ of the censored QR is calculated based on the updated subsample J_1 again using standard QR. An optional step four repeats the update of the subsample based on the three-step estimator $\hat{\beta}_\tau$ with $J_2 = \{i: \hat{\beta}_\tau^T x_i >0\}$ and updates the estimate $\hat{\beta}_\tau$.

4.5.4 Quantile Forecast verification

The goal of verification is to access the performance of quantile regression model along with defining adequate summary measures or scoring rules for valuation of a resultant compared with a reference. Gneiting and Raftery (2005) proposed a proper scoring rule for valuation of a result compared with a reference by using the (censored) LAD function that directly applies to estimate conditional quantiles of a censored resultant variable. In order to assess the relative gain in performance with respect to a reference, a censored quantile verification skill score (CQVSS) is derived for a specific τ quantile. CQVSS is analogously to the Brier skill score derived for one category as:

$$CQVSS(\tau) = 1-CQV(\tau)/CQVref(\tau)$$

The CQVS is a positive definite function that can takes values on an interval $(-\infty,1)$. Its expected minimum is obtained if the resultant corresponds to conditional τ-quantile and the expectation is zero if regression resultant is perfect therefore if underlying process

deterministic indicating no gain with respect to reference forecast, while value of one indicates a perfect and deterministic forecast. Murphy (1973) states that the CQVSS as well as Brier skill score are only asymptotically proper for very large samples.

Along with the CQVSS, we also adopt the quality measures relating to the prediction the amount of precipitation based on the contingency table of forecast against observations of precipitation. We use the widely accepted Relative Operating Characteristic (ROC) tool for probabilistic weather forecast verification. The ROC has been used for decades in engineering, biomedical, and psychological applications. The ROC measures the hit rate of a forecast against its false-alarm rate as the decision threshold (quantile of a probabilistic forecast) is varied. For ensembles, the ROC is a curve that indicates the relationship between hit rate and false alarm rate as different sorted ensemble members are used as decision thresholds. As per the conventional method of calculation the 2×2 contingency tables are generated, with separate contingency tables tallied for each sorted ensemble member.

We deduce the Probability of Detection (PD), and the False Alarm (FA) corresponding the extreme events exceeding the 95th percentile over the testing time period. PD is calculated as the number of true forecast divided by the total number of events. Similarly, FA is calculated as the number of false positives divided by the number of non-events. Range of PD is zero to one, with a perfect score 1 and no skill at zero. PD is sensitive to hits but does not take into account of false alarms. While maximizing the number of hits and minimizing the number of false alarms is desirable, it is required that PD be examined together with FA. FA has a negative orientation, range of FA is also one to zero, with a perfect score zero. We also compute the Heidke Skill Schore (HSS) for the model forecast. The HSS is one of the most commonly used skill score that standardizes the number of correct hits and rejections to eliminate forecasts which would be correct due to random chance (Heidke 1926). Range of HSS is minus infinity to one, a perfect score 1, no skill forecast 0. In this context we explore the potential of using a downscaling model to produce short lead (day 1 and day 3) forecasts of extreme rainfall for the urban region of Mumbai.

4.6 Results and discussion

We first examine potential forecast skills of the raw GEFS forecasted rainfall over the grid including the city of Mumbai. Figure 4.5 presents the rainfall climatology, mean standard deviation and extremes as observed for the station Santacruz (Mumbai) and the same from the

GEFS ensemble precipitation forecast. A comparison clearly reveals that the GEFS ensemble precipitation forecast considerable underestimates the observed precipitation in terms of all the three selected parameters. These facts elaborate the need of precipitation downscaling to provide a correct forecast for the city of Mumbai.

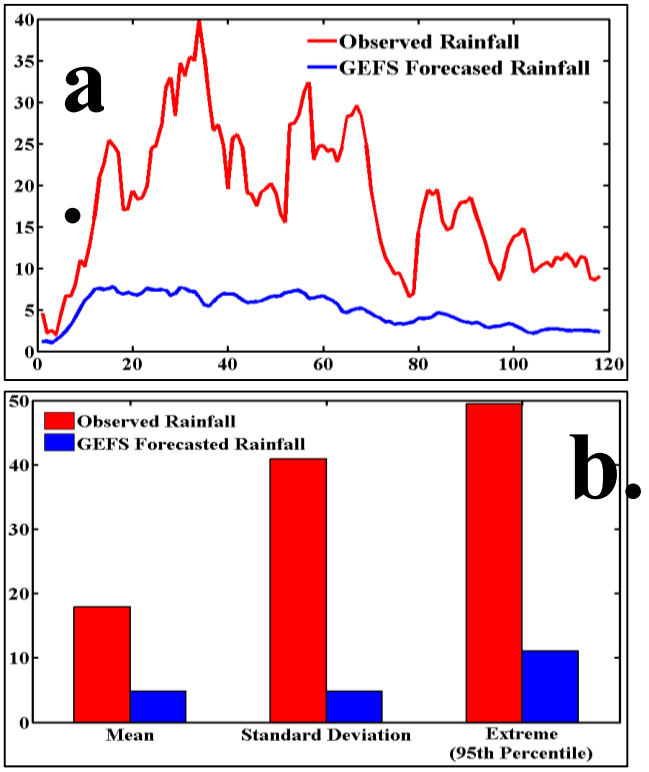


Figure 4.5 Association of the Observed and GEFS forecasted Rainfall.

a. Climatology (5days moving average) of the observed and GEFS forecasted rainfall

b. Mean, Standard deviation and Extremes(at 95^{th} percentile) of the observed and GEFS forecasted rainfall

The optimal number of EOF entering the QR model is defined at the first step of the downscaling process. We examine the potential of using a combined covariate vector of the selected variables. As described in the data section, the multivariate covariates consists of 21 × 21 grid point time series over station Santacruz (Mumbai) for 10 meteorological variables from the ERA-Interim reanalysis dataset. As reduction of spatial degrees of freedom of the covariate is necessary prior to model fitting the multivariate covariate is projected onto its leading empirical orthogonal functions (EOFs) using the PCA. The skill scores are calculated from the independent cross-validated forecasts, whereby a subsample of 15 years was used for model estimation and the remainder 14 years used for validation.

Figure 4.6 (a) shows the CQVSS for varying numbers of EOFs ranging from 200 to 7. The results are shown for lead time of 24 hours. The CQVSS is largest at 0.99, 0.95 quantile, the least skill is obtained for the 0.8 quantile. Table 4.1 shows the proportion of total variance explained by the predictors selected under each individual experiment.

Table 4-1 The EOF selected with different experiments.

Sr. No.	EXPERIMANT	No. of PC
1	**PCA 98**	200
2	**PCA 90**	48
3	**PCA 88**	39
4	**PCA 87**	35
5	**PCA 86**	32
6	**PCA85**	31
7	**PCA80**	21
8	**PCA75**	15
9	**PCA70**	11
10	**PCA65**	9
11	**PCA60**	7
12	**LASSO**	87
13	**PCA 98+LASSO**	59

The total variance is mainly captured by small number of EOFs. The CQVSS increases with increasing numbers of EOFs up to about 14 EOFs, afterward the skill score remains constant. The highest CQVSS is reached with about 12–14 EOFs. Key attributes of ensemble forecast quality include the reliability of the forecast probabilities and the ability of the forecasts to discriminate between different observed conditions (Jolliffe and Stephenson 2003). Forecast reliability is evaluated with the PD, HSS and FA following Hsu and Murphy 1986. The PD and HSS for varying the number of EOFs are presented with figure 4.6 b. As obtained with CQVSS the estimation of PD and HSS also reaches maximum at about 30-39 EOFs. Based on this analysis we select 35 EOFs that is convenient to represent the multivariate covariate. The selected first 14 EOFs explain about 87% of the total variance of the combined ERA-Interim field. Though ,with a high PD the HSS skill score remains low mainly because of the high False Alarm Ratio (FA). However the investigations reveal that the events of FA are mostly confined to the prior or after 1-2 days the occurrence of the extreme event as revealed with figure 5.6©. The best fit model using 14 EOFs is applied to ensemble forecasts of the bias corrected predictor variables from the NCEP-GEFS for lead times of 24 and 72 hours(lead time 1–3 days). Independent validation results, including reliability and discrimination are presented for ensemble forecast quality. The model performance with the GEFS predictor dataset is

estimated using the CQVSS score (figure 4.6(a)). The CQVSS reaches 50% for the higher quantiles.

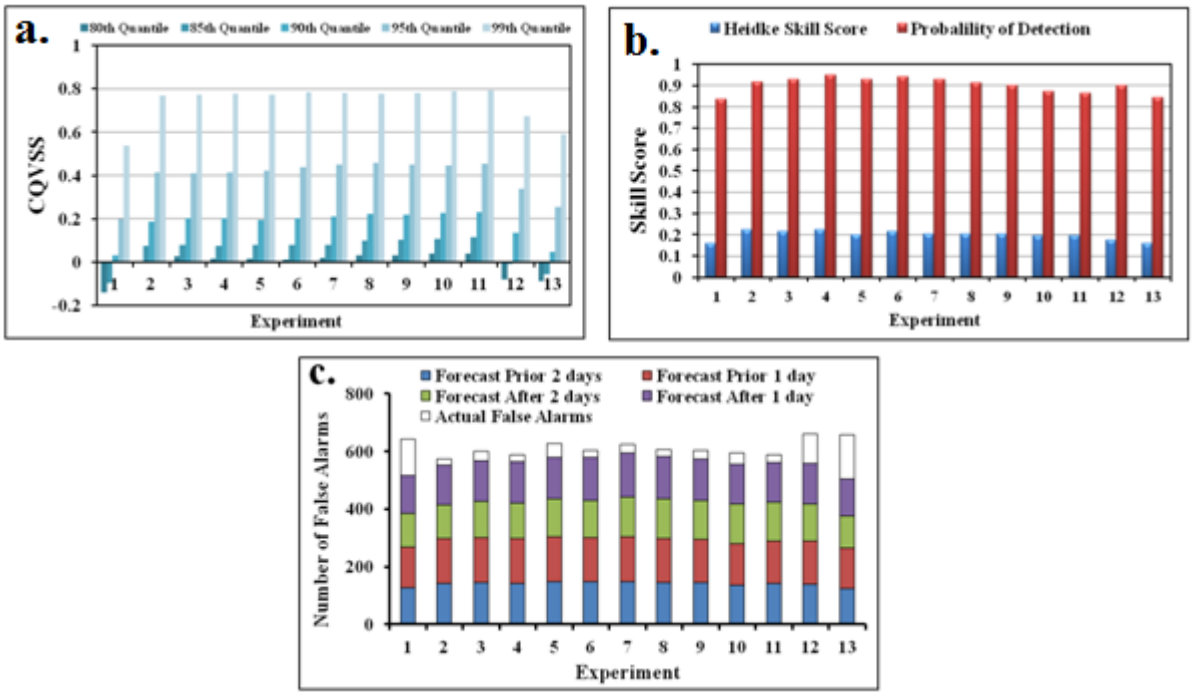


Figure 4.6 Skill score for varying numbers of EOFs.

Heidke Skill Score(HSS) and Probability of Detection(PD) of extreme rainfall (at 90th percentile) corresponding to different experiments conducted varying the number of EOFs of the combined covariate of the predictors(a). The reason of low HSS is the high number of False Alarms (FA). However the investigation shows that the FA are mostly confined around the occurrence of extreme events(b). The selection of EOFs with individual experiment is referred to table 4.1.

To further verify the quantile forecasts, we investigated whether the quantile forecasts are reliable, that is, the wellness of forecasting the individual extreme events. If a quantile forecast is reliable, then it is expected that the observed precipitation values lies below the forecast at conditional τ^{th} quantiles. The performance of the quantile model is verified for forecasts of the extremes occurring over the last two years (2006-2007). The verification samples comprised paired forecasts and observations at lead time of 24 hours. The observed extreme rainfall is compared with the model forecast results and bias-corrected GEFS forecasts results (figure 4.7). In figure 4.7 observed precipitation (the red solid line) and the forecast obtained with bias corrected GEFS ensemble forecast (green solid line) as well as the ensemble spread (green patch) are plotted with the ensemble mean forecast obtained from the model output (blue) and its ensemble spread (cyan). For all 29 extreme events over the verification time period, the

results reveal a higher uncertainty band associated with the bias corrected GEFS precipitation forecast (green band with each subplot). The results at the same time evidences reduction of uncertainty spread associated with in the GEFS forecasts with the proposed downscaling methodology over all the extreme precipitation events.

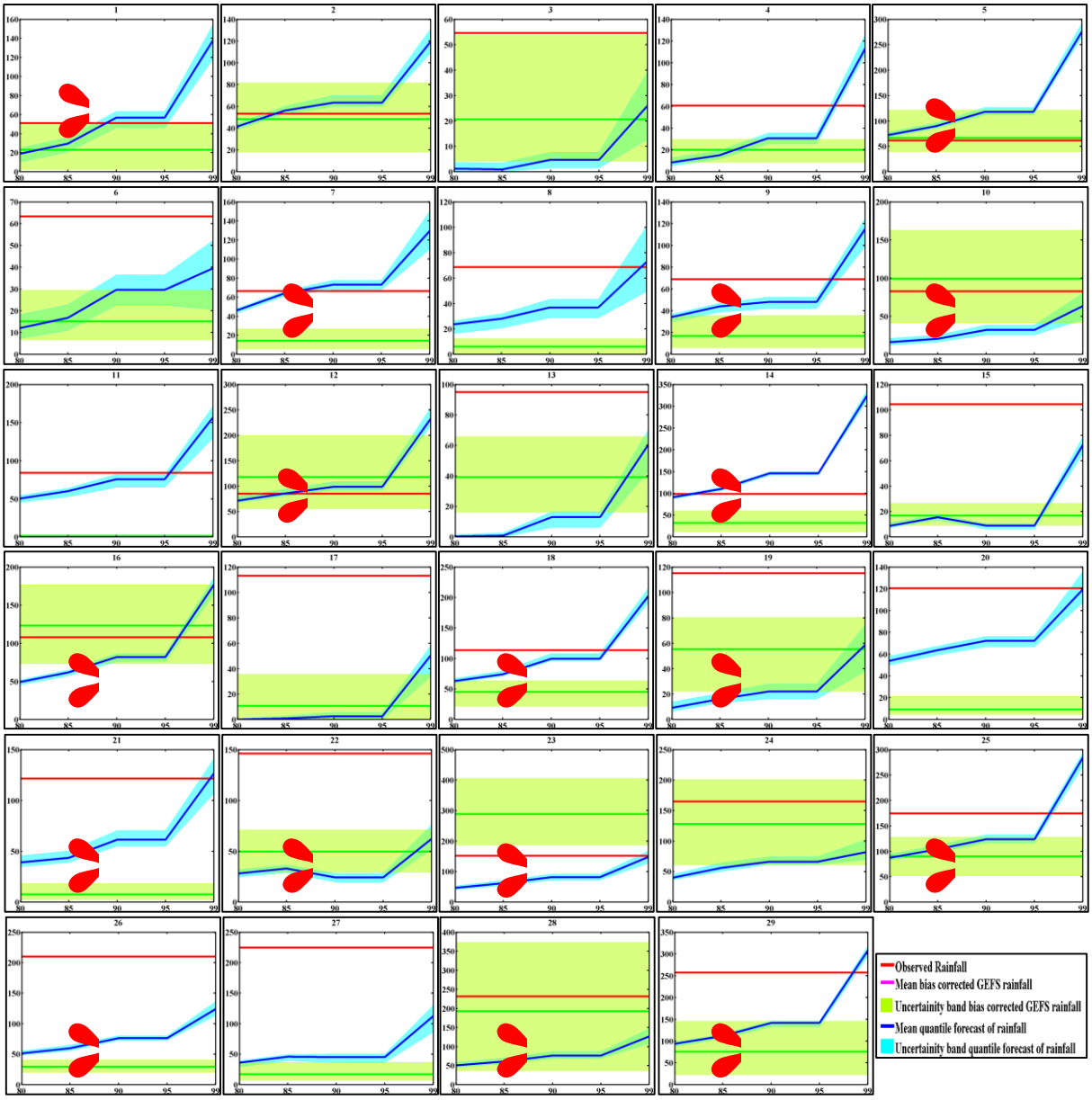


Figure 4.7 Performance of quantile forecast of rainfall and GEFS bias corrected rainfall.

Model performance is checked for forecasting of extreme events (at 90th percentile) over the years 2006-07. 29 extreme events occuring over the evaluation time period is shown here. The bias corrected GEFS rainfall shows a higher uncertainty (green band) and fails to capture the

extreme events with mean (green). However, the downscaled rainfall shoes a lower uncertainty (cyan band) as well as attempts to captures the events with mean value(blue) for 17 events (marked with red asterisk).

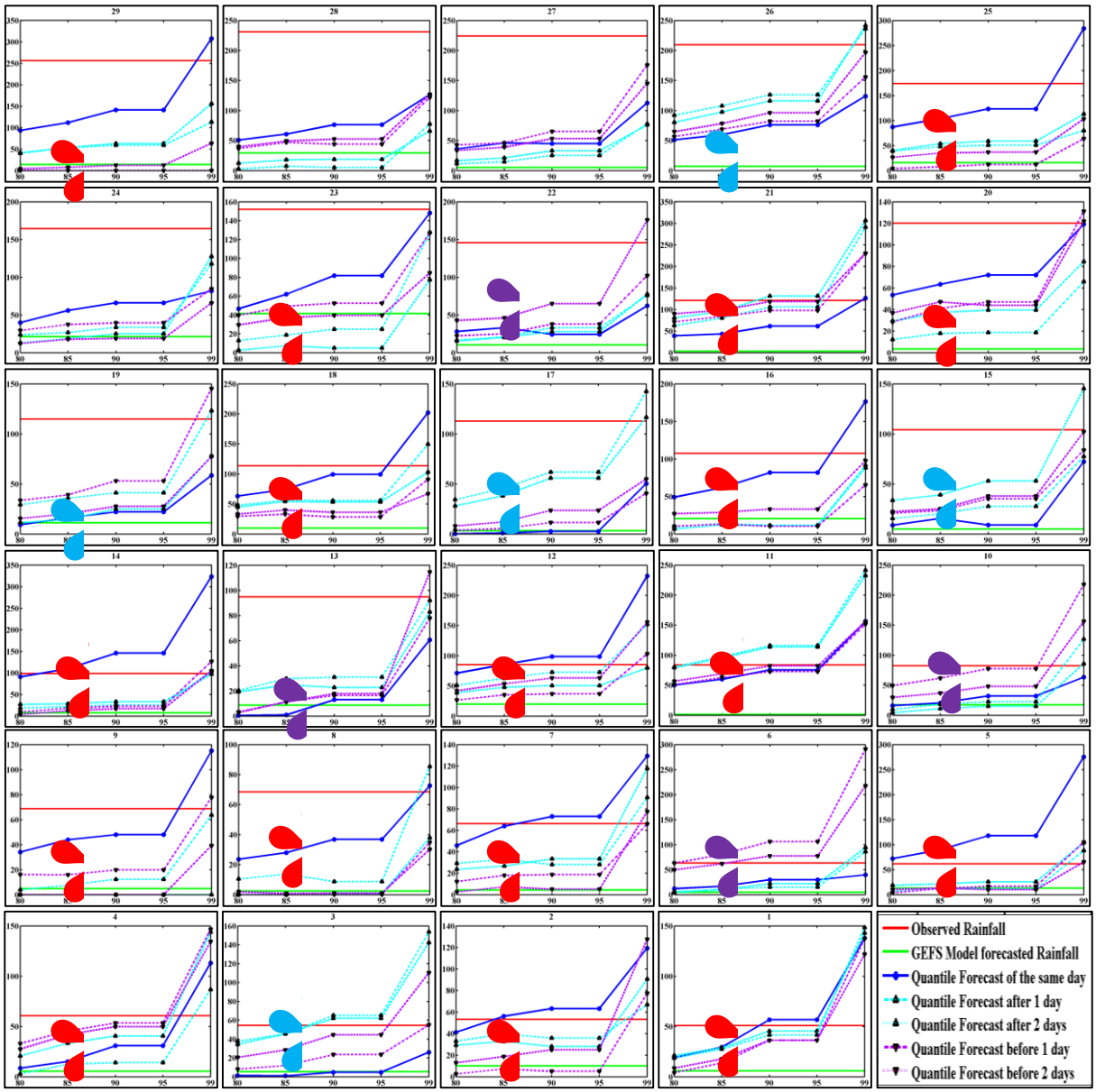


Figure 4.8 Cross validation of the quantile rainfall forecast.

The reduction in the ensemble spread provides a mechanism for better interpretation of the forecast at the present level of forecast errors. It is interesting to point out that the proposed model substantially improves the accuracy of precipitation forecast as well. 17 extreme events occurring over the verification time period are successfully captured with the model forecast

on the same day. Out of the remaining 12 extreme events, the further examination reveals that 9 events are forecasted on 1-2 days prior-after the occurrence of the extreme event (figure 4.8). Model performance in terms of forecasting of extreme events (at 90th percentile) observed in the years 2006-07. The GEFS model forecast(green) fails to capture all the extreme events, however the downscaled rainfall at different quantiles provides a considerable improvement over the GEFS precipitation. Out of the 29 events observed over the evaluation time period 17 are forecasted correctly with the model(marked with red asterisk). However the model captures 5 events 1-2 days prior to its occurrence (marked to with cyan asterisk) and 4 events after the occurrence (marked with purple asterisk).

4.7 Summary

In this paper we demonstrate that a reliable extreme precipitation forecasts in terms of quantiles can be made by means of quantile regression. The approach requires no strong assumptions. The inclusion of information from NWP models is also very flexible. The censored quantile regression based model is trained with the well simulated climate variables from the reanalysis data to constrain parameter estimates of extreme precipitation event. Ensemble forecasts of NCEP-GEFS is observed to contain a typical underestimation in the mean, spread, and higher moments of precipitation and remains non useful. A sufficiently reliable and unbiased forecast is obtained with the help of proposed downscaling model. The regime of extreme rainfall for the station Santacruz, Mumbai is correctly identified on a daily time scales with an estimate of the associated uncertainty. The downscaling lead to significant improvement in the ensemble distribution and ensemble based forecast with decreased spread and reduced systematic errors in the ensemble mean. For instance, the raw GEFS forecasts did not capture any of their verifying observation whereas, 90% of the observations were captured following the proposed downscaling methodology. The three-step censored QR following Chernozhukov and Hong (2002) is easy to apply using a standard QR procedure. However, a careful selection of predictors is crucial to obtain good forecasts. The model skill can be improved by an appropriate selection of forecast variables. A clear and precise measure of goodness is very important at the same time. The CQVSS provides a critical check on the performance of the model. The forecast skill reduces with the forecast lead time as well, due to the reduced indicator cross correlations at longer lead times. The improvement from downscaling also declines systematically with lower quantiles.

The quantile regression for the τ^{th} quantile assumes the linear model (Koenker 2005), the assumption of linear dependency seems reasonable for this case as the problem of crossing of the conditional quantiles constitutes is not encountered. However, there might be cases where the linear assumption is inconvenient or crossing quantiles indeed constitute a problem. The crossing of the conditional quantiles also occurs when the training period is short (Friederichs and Hense, 2006). Here, it is also important to state that a considerably long training period is very important to obtain good forecast skill for very high precipitation quantiles such as the 0.99 quantile.

The present study observes an improvement in objective weather forecasting through the combination of numerical and statistical models. Since the computational time needed to run this model is very short, the model can be easily adopted with input data from number of urban stations. The model can be run both automatically and on inputs provided by the forecaster to provide extreme rainfall forecast for the different urban centres of the country. It is also straight forward to apply the approach to other variables like temperature and wind speed. The study highlights importance of development of new and modification of old numerical and statistical models as useful tools for short range weather forecasting. The results obtained for the city of Mumbai emphasize the fact that the combination of dynamic and statistical methods are much valuable to progress in the field of objective weather forecasting. An increased effort in this direction can prove to be much beneficial for the fast developing country like India having huge number of urban centres without investment in the computational infrastructure.

Chapter 5 : PROBABILISTIC RAINFALL FORECASTS ON EXTREME DAYS IN MUMBAI

5.1 Introduction

The science of climatic extremes is important and critical in terms of modeling, socioeconomic impacts, damages, and adaptation. Occurrences of rainfall extremes are expected to increase in changing climate (Goswami et al. 2006; IPCC 2012), and hence, proper scientific understanding of extremes is crucial. Though there are significant research advancements in the last two decades in the science of extremes (Cavazos et al. 2008; IPCC 2012; Wheater 2002; Young 2002) to minimize the impacts, hazards, and losses, there are still a significant number of extreme events resulting in huge human and economic losses. One such example is the disastrous flood of Mumbai, India in 2005, which caused 409 deaths and an unprecedented loss of Rs. 5,000 Crore. More than one million people were rendered homeless because of the flood (Rajendra et al. 2006). Early prediction-based alert broadcasting may help in operating the existing flood control systems with maximum efficiency, which minimizes losses, such as evacuation, flow diversion, alerting the population, preparedness of the disaster mitigation team, etc. Such a warning system may alert the population for making their own arrangements to safeguard their lives and their properties. People under attack will be relocated well before the advancement of the flood. In addition, early warning may enable the preparation of international assistance actions (Pappenberger et al. 2008).

Extreme rainfalls are expected to be associated with exceptional atmospheric conditions. The possible reason of heavy rainfall occurrence over Indian regions (western coast) is the high-speed wind coming from the Arabian Sea with excess moisture. Heavy rains over different parts of India are usually caused by (Rakhecha and Pisharoty 1996):

1. Formation and subsequent movement of cyclonic, low-pressure regions with anticlockwise wind circulation disturbances across the country. These cyclonic disturbances originate from the Bay of Bengal and Arabian Sea.

2. Breaks in monsoons, when the rainfalls are confined to the Himalayas and the Indian regions. During a break, the monsoon trough gets displaced northwards from the normal position and cause heavy rain near the area on the Himalayan side.

Extreme rains of short duration are the consequence of convective instabilities in moist air

in small spatial location (Goswami et al. 2006). Although the fraction of extreme rain events is caused by synoptic disturbances (Francis and Gadgil 2006), a large number of extremes are caused by processes like thunderstorms and are more uniformly distributed with space and time (Goswami et al. 2006). Thunderstorms result from the accelerated upward movement of moist and warm air. Two environmental conditions, which encourage the development of thunderstorms, are (1) very warm and moist air and (2) an environmental lapse rate in which temperature decreases more rapidly with altitude than it does for either the dry or wet adiabatic lapse rates. Western Indian heavy rain events are associated with a low-pressure region over the northwest Bay of Bengal, development of high-magnitude vortices, and increase in vertically integrated moisture in the region (Joseph 2006). Although majority of intense rainfall events over the west coast of India are associated with large-scale systems such as the Tropical Convergence Zone and organized convection over a large-scale oriented in the east–west direction, some of them are also associated with offshore convective systems and/or mid-tropospheric cyclones.

The intensity–duration–frequency (IDF) approach for de-fining design precipitation for water resources systems is used to develop a relation between intensity, duration, and frequency. Analyzing IDFs show that extreme rainfall has low frequency. Since extreme events are very rare, it is difficult to acquire more information about them and, hence, difficult to forecast them. The India Meteorological Department (IMD) uses both linear and nonlinear regressions for rainfall forecasting. At present, large numbers of models are involved in finding out possible combinations of predictors for long-range forecasting; only few models with best skill are selected. The National Center for Medium Range Weather Forecasting (NCMRWF), an Indian government agency, provides daily weather and rain forecasts based on physics-based models, but the models are not capable of predicting heavy rainfall accurately (Khaladkar et al. 2007). Here, we present a brief over-view of statistical and dynamical approaches used for extreme rainfall prediction.

Numerical weather prediction (NWP) models which are based on dynamical weather equations are used to provide short-range forecasts based on the present weather conditions. These models use systems of differential equations based on the laws of physics, fluid motion, and atmospheric chemistry. The model run takes place with initialization which includes feeding the present weather conditions to the model. The partial differential equations for the dynamics of weather are solved: some models use finite-difference methods for all three spatial

dimensions, while other global models and a few regional models use spectral methods for the horizontal dimensions and finite-difference methods in the vertical dimensions (Strikwerda 2004). Commonly used NWPs are T80, T170, fifth-generation mesoscale model (MM5), and Eta model. These models are primarily run to predict mesoscale (20–200 km) precipitation, along with other meteorological variables, using synoptic scale (200– 2,000 km) weather conditions. In India, the MM5 model is run on triple-nested domains at 90-, 30-, and 10-km grid resolutions using initial conditions from the T80 global model of NCMRWF (Gupta et al. 2004). However, these models are poor in simulating extreme rainfall in India (Khaladkar et al. 2007). One possible reason may be be-cause Indian regions are nonhomogeneous with respect to the land–ocean interaction, terrain distribution, and prevailing weather systems. This aspect of non-homogeneity in Indian regions makes it important to take into account the mesoscale conditions of the atmosphere, along with synoptic scale circulations, for the prediction of regional or mesoscale heavy rainfall events. Dodla and Ratna (2010) analyzed extreme precipitation event which occurred over the west coast region of India on 26 July 2005. The results showed that mesoscale atmospheric modeling can be very effective in predicting extreme heavy rainfall events. Mesoscale NWPs are run on high spatial resolution of 0.5–1 km for local and regional weather forecasts, which demands huge computational effort; however, such a high-resolution modeling is necessary for urban regions to incorporate urban land surface processes. Urban climate modeling is complex because of the variations of solar radiation, temperature, and wind conditions with spatially changing topography and local surroundings. The model should also be able to capture urban heat island (UHI) effect, momentum, moisture, and heat transfer mechanism caused by urban land use pattern. UHI are areas like urbanized cities, which are warmer than the surrounding areas, thereby making the high temperature persist for a time longer than normal. Because the area is warmer, the capacity of air to retain more moisture increases. The UHI effect plays an important role in perturbing thermal and dynamic processes (Lin et al. 2001). The use of inappropriate initial conditions in urban areas can severely affect the model output. Although these models are multiply nested and use highly meticulous methods of parameterization techniques and higher order numerical mathematical equations to represent the dynamics of the atmosphere, output from the models are not perfect; hence, there is a need for corrections of flaws and discrepancies (Nott et al. 2001). NWP models are found to be efficient in predicting weather; however, these models are not efficient enough to predict heavy rainfall events well in advance (Březková et al. 2010; Hong and Lee 2009; Khaladkar et al. 2007; Selvam 1988). Extreme rainfall threats are expressed through the deterministic quantitative precipitation forecast (QPF)- a spatial and

temporal forecast of the potential amount of precipitation over a region, probabilistic excessive rainfall outlook- categorical probability of rainfall exceeding given thresholds, and experimental probabilistic quantitative precipitation product suite-continuous probabilities of rain-falls exceeding thresholds. Novak et al. (2011) observed that, on average, the deterministic QPF improves over the numerical model; however, the deterministic QPF is still often inadequate for the depiction of extreme rainfall events.

Among the statistical and probabilistic models used for the prediction of heavy rainfall events, analog methods are most popular. These methods are based on the similarity of atmospheric conditions on extreme days. Two atmospheric states are said to be analogous if there is certain resemblance. Ideally, two states should be considered similar only if the three dimensional distributions of wind, pressure, temperature, water vapor, and other parameters such as sea surface temperature, snow cover, etc. are similar (Lorenz 1969). Different criteria may be applied along with different variables to select the analog days. It is assumed that the resulting rain from both the states will be similar in quantity. Criteria for analogous weather pattern for heavy rainfall, as used by Altava et al. (2006), are criterion of proximity in n dimensions (n is the number of grid points) and the criterion of correlation between variables that characterize the atmospheric state. Only the states that have parameters higher than a prefixed threshold are selected to be analogous. The results of the analog method are compared with the deterministic model MM5 mesoscale output, and the results show that, while MM5 overestimates the rainfall, the analog method underestimates it. Daoud et al. (2011) used analog methods to predict precipitation over Saone River basin, France. Past meteorological conditions are used as predictors from which analogs will be sorted, and past rainfall amount are used as predictands. However, the performance is not satisfactory, suggesting the need for significant improvements in the algorithms.

Abundance of literature is available for predicting rainfall over Indian catchments (Bhowmik and Durai 2010; Mitra et al. 2011; Rajeevan 2001; Sahai et al. 2000); however, there is scarcity of literature exclusively for predicting heavy rainfall from weather patterns. Researchers have attempted to link the extreme rains with eccentric atmospheric behavior (Hart and Grumm 2001; Panziera and Germann 2010; Tymvios et al. 2010); it has been observed that extreme rainfall events have a great dependence on anomalous weathers. There are barely any studies which have analyzed atmospheric conditions during historic extreme rainfall events in order to improve the predictability of these extreme events over Indian regions. However, globally, few

recent studies have related extreme rainfall with mesoscale and synoptic scale weather conditions.

The fingerprinting technique, developed by Root et al. (2007), is a recently developed statistical method that uses clustering technique to detect the atmospheric variables and areas undergoing significant changes during extreme events. These variables and areas become the fingerprint of the extreme events. For the prediction of extreme events, a pattern recognition technique is used to match the fingerprint and the weather of a given day to be classified as extreme or non-extreme day. Root et al. (2007) used the fingerprinting technique to predict extreme weather phenomena like severe snow fall, flash flood, severe thunderstorm winds, etc.; the method has been found to be able to distinguish between types of events, for example, whether it is a snow event or a severe thunderstorm wind event. Here, we apply the same for predicting the extreme rainfall event in Mumbai; however, such a method is observed to result in a huge number of false alarms. This motivates us to use a machine learning technique: support vector machine (SVM)-based algorithms involving the anomalous weather pattern to predict extreme weather events for alert broadcasting. Here, we first describe the data used for this study followed by the application of the finger-printing approach for the Mumbai case study with it limitations. With this background, we present in detail the proposed algorithms with the improvements achieved over the results of the fingerprinting approach.

5.2 Data

The major data, which are required to be collected for extreme weather prediction, are the rainfall data and historical weather data. The rainfall data of Mumbai is collected from the rain gauge station at Colaba. The Colaba rain gauge station in Mumbai is located towards the south of Mumbai, the commercial capital of the country, located over Western Ghats of western India. Mumbai is a highly urbanized city and is usually threatened by short-term intense rains during south-west monsoon. Colaba station is one of the observatories of IMD in Mumbai; it is located at 18.93° N and 72.85° E. The Indian summer monsoon (June, July, August, and September) hourly rainfall data is collected for the period 1969–2008 from IMD. Except for 3 months (June 1991, September 1993, and September 2006), data are available without any break. The prediction is performed in this work at 6 to 48 h lead time. The weather data used for this analysis is the six-hourly National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis data. The atmospheric variables

used in the study are at surface level and at the 850-, 600-, and 400-hPa levels. Surface level variables include air temperature, mean sea level pressure (MSLP), precipitable water (PW), relative humidity (Rhum), U-wind, and V-wind. The 850-, 600-, and 400-hPa level variables include air temperature, vertical wind velocity (omega), relative humidity, U-wind, and V-wind. The atmospheric variables for the area with latitude from 5° to 40° N and longitude from 65° to 100° E are obtained from the NCEP/NCAR global reanalysis data set (Kalnay et al. 1996). The reanalysis data set is available for 17 pressure levels and has a 2.5×2.5° spatial grid resolution. The data set is available from 1948 to date. Six-hourly atmospheric variables from 1 January 1969 to 31 December 2008 are downloaded from the NCEP/NCAR download page (http://www.esrl.noaa.gov/psd/data/gridded/ data.ncep.reanalysis.html#temp) in NetCDF format.

The data from 1969 to 2008 are divided into a training set and a validation set. The training set consists of extreme rainfall events from 1969 to 1999 and the validation period is from 2000 to 2008. As per the IMD definition, heavy precipitation event (HPE) and extreme HPE are defined as 24-h precipitation events, when rainfall exceeds 120 and 200 mm, respectively. Here, we consider "6-h" events and define extreme rainfall, when the amount exceeds 75 mm (≥75 mm). This is slightly on the higher side in terms of magnitude, as compared to the IMD definition. This is intentionally set, considering the short duration (which has high intensity com-pared longer duration) and high monsoon rainfall in Mumbai, compared to other regions in India. A total of 66 such events are identified over Mumbai from 1969 to 2008. Out of all, two events are having rainfall more than 200 mm; the highest being 338.00 mm which occurred on 25 June 1985 and the other one with rainfall amount of 234.8 mm occurring on 5 July 1974. Based on the above definition, the event database is generated which is presented in Table 5.1. A total of 66 extreme events are identified; 50 events are used for training and 16 events are used for validation/testing.

Table 5-1 Extreme rainfall events database

S. no.	Year	Month	Day	Time	Amount (mm)
1	1970	6	16	1	87.2
2	1970	6	17	1	141.4
3	1970	8	18	3	102.2
4	1970	9	22	1	108.6
5	1971	6	23	1	138.2
6	1971	6	23	3	118.1
7	1971	8	26	1	85.4
8	1972	6	28	1	120.2
9	1974	7	4	1	143.7
10	1974	7	5	1	234.8
11	1974	8	4	1	169.7
12	1975	7	30	3	96.5
13	1975	9	2	2	115
14	1977	6	17	2	84.7
15	1977	7	21	4	91.6
16	1977	9	2	2	122
17	1978	6	15	1	89.6
18	1979	7	31	2	95.8
19	1981	9	22	4	96.2
20	1982	6	22	1	118.5
21	1983	7	12	1	90
22	1983	7	17	2	82.8
23	1983	8	15	1	105.2
24	1983	9	24	4	93.7
25	1984	6	13	2	98.8
26	1984	6	30	1	110.5
27	1984	7	1	4	153.5
28	1984	7	3	4	131.3
29	1984	9	12	2	90.5
30	1985	6	16	4	142
31	1985	6	25	2	338
32	1986	8	8	1	90.5
33	1987	7	1	2	113.5
34	1989	6	16	4	85.5
35	1990	6	15	4	138.8
36	1990	6	16	3	128.8
37	1990	6	26	1	102.3
38	1990	8	15	1	153.9
39	1992	7	16	4	89
40	1994	7	12	2	96.6
41	1995	9	1	1	139.9
45	1997	8	22	4	86.4
46	1997	9	26	1	98
47	1998	6	27	3	128.9
48	1998	8	9	4	185.7
49	1998	8	25	4	88.1

50	1999	6	23	1	100.6
51	2000	7	3	4	99.5
52	2000	7	8	1	105.5
53	2000	7	12	2	159
54	2001	7	8	1	106.6
55	2003	6	19	1	89.2
56	2004	7	29	3	97.2
57	2004	8	2	2	90.6
58	2005	6	23	1	124.1
59	2005	7	26	2	NAN
60	2005	9	10	1	164.8
61	2006	7	4	3	107.6
62	2006	8	6	2	89.5
63	2007	6	23	3	110.5
64	2007	8	3	2	111.8
65	2008	7	1	1	83.6
66	2008	7	27	3	92

5.3 Fingerprinting extreme rainfall events

The fingerprinting technique employed for predicting extreme rainfall events has been developed by Root et al. (2007). The technique is developed based on the hypothesis that extreme event days have a specific weather pattern (fingerprint) which is different from the normal day's weather pattern. As for example, the vertical velocity or the Lagrangian rate of change in pressure, commonly known as omega in meteorology, pattern over India, before 6 h of heavy rainfall event occurred in Mumbai on late night 21 July 1977, is presented in Fig. 1b, which is different from that of the normal condition (Fig. 5.1(a)). The underlying hypothesis of the fingerprinting approach is that such an omega pattern is only expected before the occurrence of extreme event and not on a normal weather day.

It is observed from Fig. 5.1 that the omega pattern near the Tibetan Plateau, in Central India, and in the Arabian Sea for extreme events is different from those of the normal condition. The omega values in these regions deviate from the mean condition before extreme events and such patterns usually form the fingerprint for an extreme event. Such retrieved patterns are used for predicting extreme events, well in advance.

The fingerprinting technique considers maximum anomalies in the atmospheric variables before extreme events to determine the regions that are consistently having highly anomalous values during extreme events. Standardized anomalies, i.e., deviations from the climatological mean, are calculated all over the domain considered for analysis: the maximum (peak) and

minimum (valley) anomalies for all extreme events are located; their values and locations are obtained. The clustering technique is applied to figure out the areas which show highly anomalous weather behavior during extreme rainfall. The present study uses strong point analysis, as used by Root et al. (2007), to cluster the obtained peaks and valleys. Depending upon the resemblance of cluster with weather pattern of extreme rainfall events, significance of cluster is determined. Clusters with high significance form the fingerprint of an extreme event. For assigning a day as extreme or non-extreme, the day's score is calculated which gives an estimate of how close the weather pattern of the day is to the fingerprint (pattern) of the extreme event.

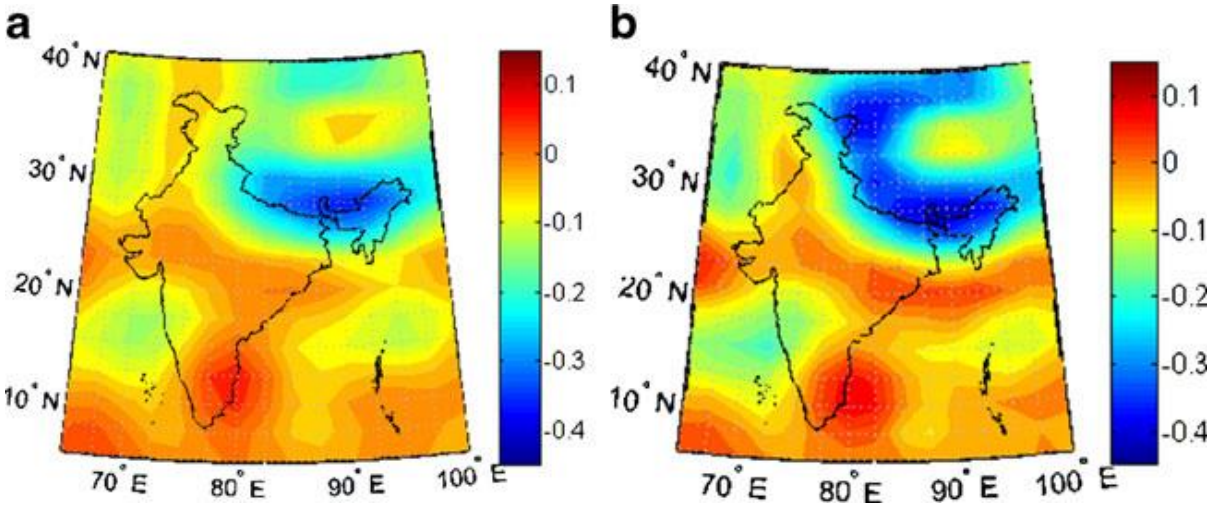


Figure 5.1 Fingerprinting of extreme events
Condition of omega at 850 hPa pressure level for (a) normal condition and (b) extreme condition. The patterns near the Tibetan Plateau, in Central India, and in the Arabian Sea form the fingerprint

The methodology as followed by Root et al. (2007) begins with developing an event database which is presented in Table 5.1. The next step is to create the climatology of weather variables (here, we use the NCEP/NCAR reanalysis data as proxy to be observed), i.e., to find the temporal mean and standard deviation of all the atmospheric variables over specific grid points. The specific grid points include a region larger than the study area. The region considered in the present study extends from a latitude of 5° to 40° N and a longitude of 65° to 100° E at 2.5° grid resolutions. The primary reason behind this selection is to capture the mesoscale (>5 km) and synoptic scale (>1,000 km) weather patterns before heavy rainfall. The climatology of atmospheric variables is created in order to determine the atmospheric variable's anomalies during events. The 40-year (1969–2008) climatology is prepared for each of the variable on each day of the monsoon (June to September). For example, the first July climatology of temperature is defined by its mean and standard deviation over 40 years. The climatology

process uses a 21-day centered mean technique on the means and standard deviations computed over 40 years, i.e., taking the 21-day mean of the means. This procedure is as per Root et al. (2007).

To start with, anomaly values of weather variables before 48 to 6 h of each event are calculated at each grid point. For each meteorological variable, the climatological mean, X, is subtracted from the NCEP/NCAR variable values (x) and divided by the standard deviation (σ) to arrive at a normalized climatological anomaly. The formula for calculating is given in Eq. 1:

$$N = \frac{x - X}{\sigma} \qquad 1$$

where N is the normalized climatological anomaly value at a particular grid point on a particular day. Generating and normalizing the anomalies converts a pseudo-normal distribution into a standard normal distribution (Hart and Grumm 2001).

The primary peak (PP; maximum normalized anomaly value in the domain), secondary peak (SP; second maximum normalized anomaly value in the domain), primary valley (PV; minimum normalized anomaly value in the domain), and secondary valley (SV; second minimum normalized anomaly value in the domain) with their normalized anomaly values (or simply anomalies) are identified in the whole do-main for each extreme event in the training. It should be noted that the positions of peaks and valleys will change with the change of variables. The positions and values of the PP, SP, PV, and SV are obtained for each variable for each extreme event. In order to determine the regions which are consistently different in terms of weather pattern during extreme rainfall events, clusters of positions of critical points (PP, SP, PV, and SV) are obtained. A cluster for a particular atmospheric variable is defined as the number of grid points which have high anomalous value of the variable during most of the extreme events; each cluster forms a part of an extreme event's finger-print. Corresponding to each of the atmospheric variables, there are four clusters resulting from PP, PV, SP, and SV. Conventional clustering techniques can be used for clustering, but in order to make clusters less compact and have variegated shapes, strong point analysis, developed by Root et al. (2007), is used which includes members in a cluster that have contiguous locations with a high frequency of peak or valley occurrences. Strong point analysis involves two steps for clustering: defining strong points and constructing clusters. In defining strong points, a term called grid point density is used. Grid point density of PP/PV/SP/SV at a grid point is defined as the number of occurrences of PP/PV/SP/SV on that grid point before extreme events. It identifies the grid points where the weather variables are constantly responsible for an extreme

event. A gridded map of the region considered for the study is used for strong point analysis and is shown in Fig. 5.2. A detailed description of constructing clusters using strong point analysis can be found in Root et al. (2007).

After constructing the cluster for each peak type for each variable, it is required to determine the significance (or importance) of each cluster to extreme rainfall events. The significance of the cluster is estimated in terms of the spatial location of the cluster and in terms of the anomaly values which the cluster grid points have. The significance of the cluster location is the estimate of how important the spatial location of the cluster is for extreme rainfall event occurrence. Alpha (α) measures the spatial significance of a cluster and is calculated by Eq. 2 as suggested by Root et al. (2007):

$$\boldsymbol{\alpha_i = \frac{N_{clust,i}}{N_{event}} \left\{ max \left\{ \frac{(\rho_{clust,i} - \bar{\rho})}{(\rho_{clust,i} + \bar{\rho})} \right\}, 0 \right\}} \qquad 2$$

Where $N_{clust,i}$ is the number of members in the cluster (cluster of the i^{th} atmospheric variable) and N_{event} is the number of events in the database; ρ is the event density over the cluster's grid points and is defined as the average number of PP/PV/SP/SV of a variable on each grid point inside the cluster and $\bar{\rho}$ is the average number of PP/PV/SP/SV on each grid point all over the domain. $\frac{N_{clust,i}}{N_{event}}$ is called the consistency coefficient. The consistency coefficient will be higher if there are more numbers of members inside the cluster, which would in turn mean that more extreme events have their PP/PV/SP/SV of a variable inside the cluster. The higher the consistency coefficient, the more will be the impact of the variable's cluster on the fingerprint of extreme events. In the above equation, $\bar{\rho}$ is assumed to be the density of a completely random distribution of PP/PV/SP/SV. This equation provides the comparison between the number of PP/PV/SP/SV of a variable inside the cluster and the number of PP/PV/SP/SV of the variable all over the domain when there is random distribution of peaks or valleys. A denser cluster will have a larger α value, indicating that the peak and valley locations play a consistent role in the event's fingerprint. Negative values are not considered because they would mean that the cluster represented a local minimum in density rather than a local maximum. The values of α can range from 0 (totally random) to 1 (a perfect cluster) (Root et al. 2007). To measure the significance of anomaly values of a cluster, a measure phi (φ) is used which is defined as in Eq. 3 by Root et al. (2007); it measures the anomaly value significance of a cluster:

$$\boldsymbol{\varphi_i = \frac{N_{clust,i}}{N_{event}} \left\{ max \left\{ (1 - \sigma_{clust,i}) \left(\left| 1.5 - \left| \mu_{clust,i} \right| \right| \right) \right\}, 0 \right\}} \qquad 3$$

where $\sigma_{clust,i}$ and $\mu_{clust,i}$ are the standard deviation and mean of the cluster member's

PP/PV/SP/SV values of the variable i, respectively. The φ measures how important the values of PP/PV/SP/SV inside the cluster are to the event fingerprint.

This equation can be interpreted as the comparison between the standard deviation and mean of the cluster member's values and the standard deviation and mean of a completely random distribution of PP/PV/SP/SV values.

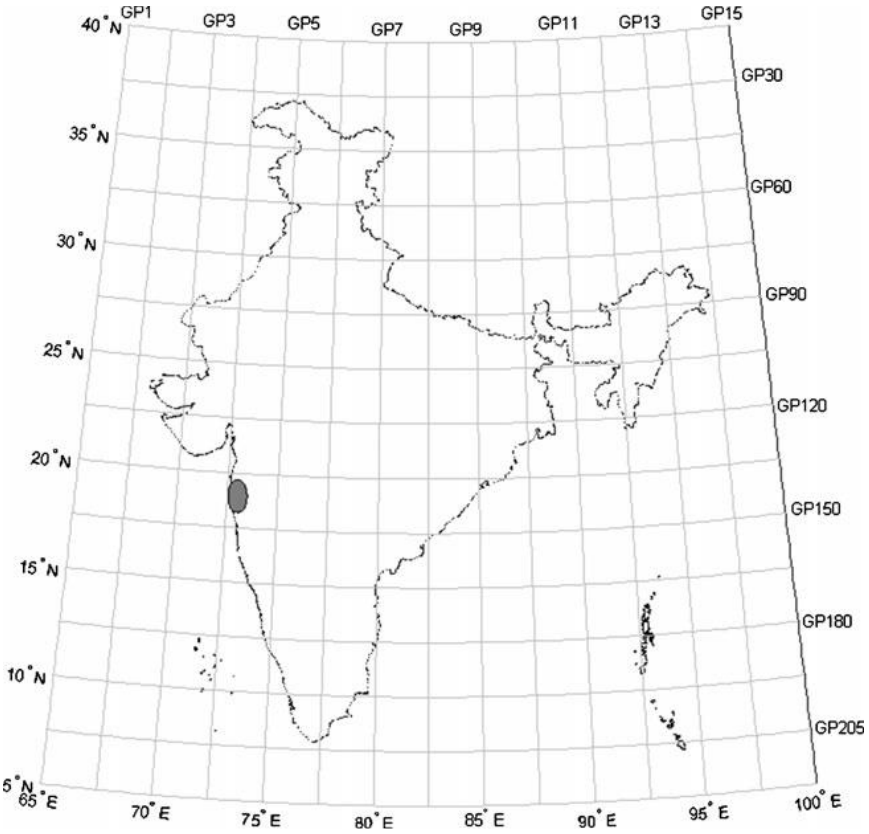


Figure 5.2 Location of Mumbai, India and location of grid points at which predictors are considered for extreme event predictions.

The grid points are numbered for better explanation

Larger values of φ indicate that the peak's and valley's standardized anomaly values provide a large part of the event fingerprint. The range of values for φ is from 0 (random) to infinity (impossible case) (Root et al. 2007). All the clusters which are significant in terms of both their spatial location and the anomaly values of their members are considered to be important for the prediction of extreme rainfall events. Prediction involves matching the location and value of future day atmospheric variable with the respective important cluster's location and values of members. Cluster matching is used to estimate the match between the value and location of a PP/PV/SP/SV of a variable on a particular day with the corresponding clusters PP/PV/SP/SV. This matching provides a partial assessment whether the day is an extreme or a non-extreme. The location of PP/PV/SP/SV of a variable is compared with the corresponding cluster's

member's PP/PV/SP/SV location, and spatial matching is calculated using a measure called gamma (γ). The γ is calculated using Eq. 4 (Root et al. 2007):

$$\boldsymbol{\gamma_i = \sum_j exp\left(\frac{-d_m^2}{c^2}\right)} \quad 4$$

Equation 4 represents a weighing factor used in Barnes' interpolation scheme (Barnes 1964; Sinha et al. 2006). Here, d_m denotes the distance between the observation point (j) and the grid point where the peak or valley has occurred; c controls the rate of fall of weighing function with distance and is calculated as given by Koch et al. (1983):

$$\boldsymbol{c = (5.052)^{\frac{1}{2}}\left\{\frac{2\Delta n}{\pi}\right\}} \quad 5$$

Where Δn = (area/number of stations). Weights for the observations farther than a distance of 5c (also called radius of influence) from a particular grid point are set to zero as suggested by Narkhedkar et al. (2008).

For matching the anomaly value of a potential future event, the value of PP/PV/SP/SV of all the variables is compared with the corresponding cluster's member's PP/PV/SP/SV values. The anomaly value matching is calculated using a measure called chi (χ). The χ measures how well the anomaly value of a peak or valley matches the set of anomaly values of the corresponding cluster members. As suggested by Root et al. (2007), for calculating χ, a histogram of the cluster member's anomaly values is created. The χ is calculated by ascertaining in which bin the forecast PP/PV/SP/SV anomaly value belongs. Then, a weighted average of the adjacent histogram bins yields χ as in Eq. 6:

$$\boldsymbol{\chi = 0.25\omega_{k-1} + 0.5\omega_k + 0.25\omega_{k+1}} \quad 6$$

where ω_k is the cluster's histogram frequency for bin number k. Larger values of χ indicate that the anomaly value of the forecast peak or valley fits better with the fingerprint. Here, α and φ are called cluster importance metrics and γ and χ are called pattern matching metrics. The pattern matching metrics, γ and χ, combined with the cluster importance metrics, α and φ, are summed to obtain the weighted matching value for each cluster, i, represented as vi given in Eq. 7:

$$\boldsymbol{v_i = \left\{\left(\frac{\gamma_i}{\gamma_{max,i}}\right)(\alpha_i)\right\} + \left\{\left(\frac{\chi_i}{\chi_{max,i}}\right)(\varphi_i)\right\}} \quad 7$$

This combined measure yields maximum values for a forecast peak or valley when it closely matches the position and anomaly values of a dense cluster. The inclusion of α_i and φ_i assigns weights based upon how important that field is to that fingerprint of extreme events. The sum of v values corresponding to all the atmospheric variables is the event score for that event as given by Eq. 8. An event score is computed for an event based on the expected fields (PP/PV/SP/SV) for a given forecast. The larger the event score, the better the expected peaks or valleys match that fingerprint:

$$\boldsymbol{Event\ Score} = \sum \boldsymbol{v_i} \qquad 8$$

For predicting extreme rainfall events, a threshold is developed based on the event score which is at the first quartile of event scores determined by hind casting. This approach does not test the hypothesis that the fingerprints are absent on non-extreme days, and hence, it is associated with false alarms, i.e., predicting non-extreme days as extreme days.

5.4 Two-phase SVM for extreme events prediction

The limitations of the fingerprinting approach are as follows:

1. The fingerprints identified by the approach may also be present on a non-extreme day, which may result in false alarms.

2. There may be multiple numbers of weather patterns, which may result in extreme events; however, the finger printing approach considers only one fingerprint. As for example, we observe that, for Mumbai, there exist two weather patterns, which result in extreme rainfall.

Hence, we propose a support vector classifier for classifying a specific day's weather pattern to extreme or non- extreme day, and it considers both weather patterns of extreme or non-extreme days. Multiple (two for Mumbai) weather patterns are taken into account in multiple (two) phase support vector (SV) classifiers. For example, let us assume that there are two weather patterns, wp1 and wp2, resulting in two types of extreme rainfall events, e1 and e2. The first phase SV classifier first classifies a day into either of the groups: (1) e1 or (2) e2 or non-extreme. If the day is classified as group 2, the second phase SV classifier is applied to classify to e2 or non-extreme. Hence, the pro-posed methodology takes care of both limitations of the fingerprinting approach. First, we discuss here a brief over-view of SVM as a prerequisite

and then discuss in detail the proposed methodology.

5.4.1 Overview of the SVM

Although SVMs were introduced in 1992 by Boser et al. (1992) in the Conference on Computational Learning Theory (COLT)—1992, there formal development took place in 1995, when Cortes and Vapnik (1995) developed the SVM classifier for binary classification. SVM is a ma-chine learning tool used as supervised statistical learning algorithm for classification and regression. The SVM classifier is widely used in bioinformatics (and other disciplines) due to its high accuracy, ability to deal with high-dimensional data such as gene expression, and ability in modeling diverse sources of data (Schölkopf and Smola 2002). SVMs belong to the general category of kernel methods (Cristainini and Shawe-Taylor 2000). The essence of SVM lies in four main concepts (Noble 2006): the separating hyper plane, the maximum margin hyper plane, the soft margin, and the kernel function. The separating hyper-plane separates the classes using the training instances, the maximum margin hyper plane is the optimal hyper plane which is located at a maximum separation from both classes, the soft margin allows some erroneous instances to be misclassified so that overall result is not affected, and the kernel function allows the mapping of input space to feature space with least computational efforts.

Using the feature of two SVM class constructs in a plane called separating hyperplane, an n−1 dimensional plane corresponding to an n dimension space separates the classes apart. The hyperplane is so selected that both classes are separated by a maximum distance from the plane; this plane is called the maximum margin hyper plane. The equation of the separating hyper plane is given in Eq. 9:

$$\boldsymbol{w} * \boldsymbol{X} + \boldsymbol{b} = \mathbf{0} \qquad 9$$

where X is the d-dimensional feature matrix consisting of features of classes to be separated, b is the bias, w is normal to the hyper plane, $|b|/||w||$ is the perpendicular distance from the hyper plane to the origin, and $||w||^2$ is the Euclidean norm of w.

In real-world problems, the data is not easily classified because of errors in some features of instances. SVM deals with the errors in data by letting few instances be misclassified by introducing a margin called soft margin; it allows outliers to be misclassified without affecting the final result. However, it is necessary that the soft margin should not allow too many instances to be misclassified; for this, a control parameter is provided to limit the number of instances to cross the hyper-plane and enter into the opposite class. In real-world problems, the

data is not separable by a linear hyper plane; for example, a two-feature data may not be separated by a linear hyper plane, but it can be separated by a nonlinear hyper plane. Constructing the optimal curve to fit the data is a difficult task; instead, the data is transformed from the input dimension space to some higher dimension space called the feature space where the data becomes linearly separable. Kernel functions do the important task of transforming (mapping) input space into feature space. If the transformed data and the separating plane are brought back to input space, a highly complex separating curve can be observed. Once the optimal hyper-plane, which maximizes the separating distance between the two classes, is constructed, either in input space or in feature space, classification is easy; points are classified depending on the position of their residence with respect to the separating hyper plane given in Eq. 10:

$$\boldsymbol{f(X) = sgn(\varphi(X)w + b)} \quad 10$$

Here, ϕ is the function which transforms data from the input space, R with d features, to the higher dimension feature space, H:

$$\boldsymbol{\varphi: R^d \rightarrow H} \quad 11$$

The performance of any data-driven prediction model depends on the correct selection of predictors. Here, in the present study, we propose the anomaly frequency method (AFM), where based on the frequency of high anomaly weather variable values, before extreme events, the predictors are selected. The following subsection presents the details of the AFM.

5.4.2 Anomaly frequency method

The AFM is an efficient technique in extracting the features which discriminate extreme events and non-extreme events. The name "Anomaly Frequency Method" comes from the frequency of anomalies, which is used to extract features of extreme events. Atmospheric variables at all the four pressure levels, at all the grid points over the domain considered, and at different time steps before the occurrence of the event are used to extract the features of the extreme. An anomaly is defined, based on experience from the fingerprinting technique, as the deviation of more than 1.25 times climatological standard deviation from the climatological mean. The anomaly thresh-old is calculated as in Eq. 12:

$$\delta^+ = \bar{X} + 1.25\overline{SD}$$
$$\boldsymbol{\delta^- = \bar{X} - 1.25\overline{SD}} \quad 12$$

where δ^+ and δ^- is the positive anomaly threshold and negative anomaly threshold respectively,

for some variables, say V-wind, having a climatological mean $\bar{X}$ and climatological standard deviation SD at a particular time instant at a grid point. If the variable value at a time instant on a grid point is more than the positive anomaly threshold, then the value is named as positive extreme anomaly, and if it is lesser than the negative anomaly threshold, it is named as the negative extreme anomaly. All the grid points are analyzed for finding the anomalies preceding extreme events. Before extreme events, variable value is compared with the respective anomaly threshold to find the extreme anomalies, if any, on each grid point; they can be positive or negative. The total number of extreme anomalies on each grid point is calculated: this gives the anomaly frequency on each grid point. The methodology helps to identify the variables and the grid points which are showing extreme anomalous behavior consistently during extreme events.

For a variable, those grid points are selected as feature grid points which have a very high frequency of extreme anomalies. Consider the region shown in Figure 5.2, all the grid points are analyzed, say, for 50 extreme events.

Anomaly frequency at each grid point for surface level V-wind before 6 h of event occurrence is determined. Figure 5.3 shows the frequency corresponding to each grid point in the region.

From figure 5.3, it is observed that grid point number 26 has the highest frequency of 26, which means that, out of 50 extreme events considered, 26 events have highly anomalous V-winds at grid point number 26.

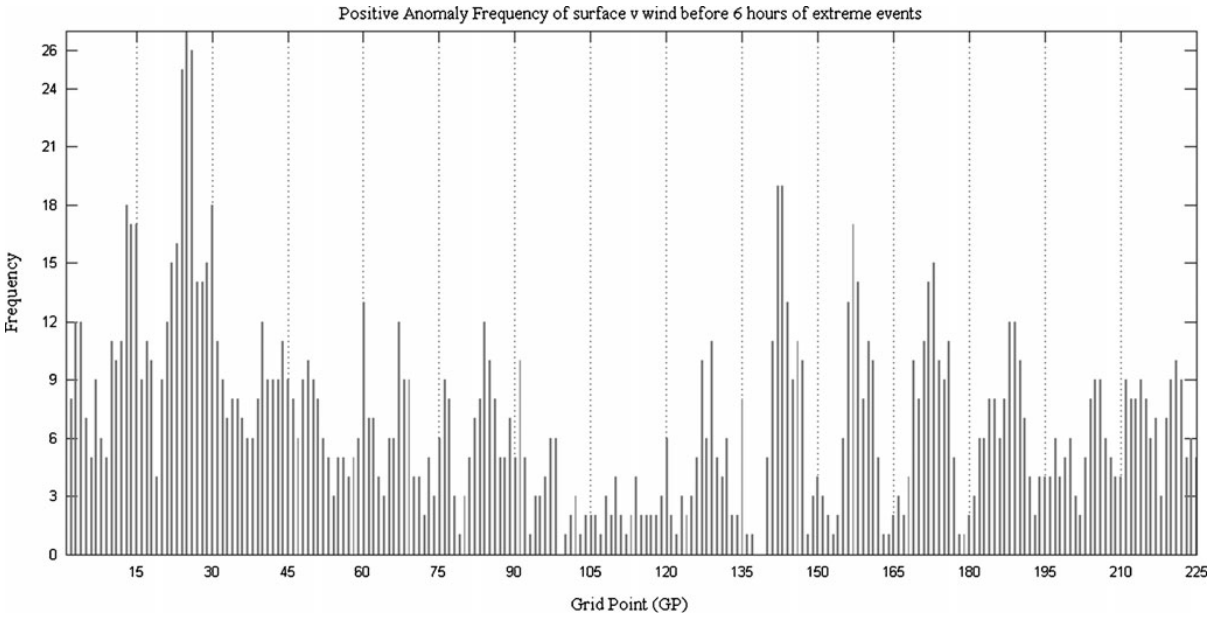


Figure 5.3 Frequency of high positive anomaly of V-wind velocity at the surface level
At different grid points, 6 h before the extreme events. Fifty extreme events are considered for this

AFM identifies two weather patterns corresponding to extreme rainfall over Colaba, Mumbai. One weather pattern corresponds to daytime (1800–0600Z) extreme events and the second corresponds to nighttime (0600–1800Z) extreme events. As the extreme rainfall events over Mumbai are associated with two weather patterns, it is essential that a model is developed which can simulate both weather pat-terns separately: a two-phase SVM is developed. The two-phase SVM uses two trained SVM models to classify a time period based on the day's weather pattern as an extreme or a non-extreme event.

5.4.3 Two-phase SVM

This section describes SVM model building for classifying extreme and non-extreme rainfall events based on the weather features extracted using the anomaly feature tech-nique. For the training of SVM, features of both extreme and non-extreme classes are used. It is found that there are two weather patterns associated with Mumbai rainfall. Conventional classification model based on single SVM may not work well because it considers a single weather pattern for the classification. Figure 4 presents the two-phase model.

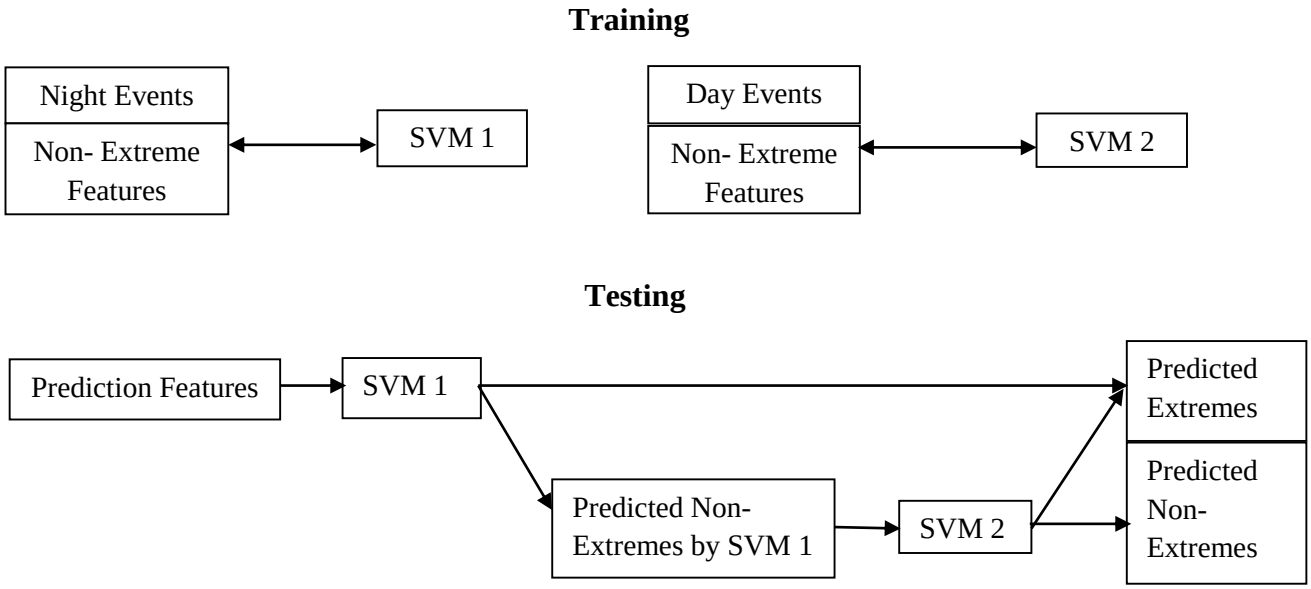


Figure 5.4 Flowchart of the two-phase SVM model

SVM model building is mostly a hit and trial process; in its general form, it includes finding the best kernel and its parameters, weights, and bias. The classification performance may also depend on the method used to find the hyper plane, i.e., to estimate the weights and bias. In the study, quadratic programming, sequential minimal optimization, and least squares method are used. These methods estimate the parameters in order to maximize the distance between the two classes. The best possible combination of kernel and kernel parameter need to be searched by hit and trial procedure. SVM determines the optimal separating hyper plane based on the support vectors selected from the data used for training. A large number of support vectors may not form a good classifier as it has an adverse effect on generalization (Quang-Anh et al. 2003); also, in contrast, a smaller number of support vectors increases the computational cost (Zhan and Shen 2005). In the present study, the extreme occurrences are rare compared to non-extremes. Hence, a large number of support vectors may have resulted from non-extreme events and to avoid that, here, we limit the number of non-extreme instances for training SVM. In the present study, it is observed that selection of the number of instances, especially the non-extremes, have a significant impact on classification. An extensive hit and trial method is employed to use the optimal number of non-extremes for training. It is found that using 50 and 30 non-extreme instances for training two models of SVM corresponding to night events and day events, respectively, yielded the best performance for testing. As shown in figure 5.4, two SVM models are trained: SVM1 is trained with 32 night extreme instances and 50 non-extreme instances; SVM2 is trained with 18 day extreme instances and 30 non-extreme in-stances.

Figure 5.4 also shows validation/prediction which is carried out in two steps: first, the instances to be predicted are fed in SVM1 for classification; if SVM1 predicts any instance as extreme, then the instance is extreme and is expected to have a weather pattern similar to night events weather pattern. If SVM1 predicts the instance as non-extreme, it is fed to SVM2 for classification. Depending on SVM2 prediction, the instance will be classified as extreme or non-extreme. An important note here is that SVM can easily help in discerning the class of instance by just looking at the sign of the hyper plane function, but the interpretation of parameters is difficult as the feature have been transformed to a higher dimension feature space.

5.5 Results and discussion

Extreme rainfall prediction of Mumbai, India is first carried out using the fingerprinting technique. As will be discussed in the next subsections, the fingerprinting technique performs poorly in predicting extreme rainfall events over Mumbai. A critical inspection is carried out to find the real cause for the poor performance of the fingerprinting technique and the reasons of extreme rainfall events over Mumbai. An attempt is made to model the features of extreme and non-extreme rainfall events. The following subsections provide with important results obtained from the study.

5.5.1 Fingerprinting results

The analysis yields information about the atmospheric variables which are the most important predictors of heavy rainfall events. These variables form the fingerprint of heavy rainfall events over Mumbai. By comparing the fingerprint with any future weather pattern, the possibility of heavy rainfall over Mumbai can be predicted before 48 to 6 h.

Analysis of atmospheric variables 48 h before extreme events show that the SV of negative anomalous relative humidity, having an α value of 0.7055, can be considered as the most important predictor. Figure 5.5(a) shows the 850-hPa relative humidity pattern over the domain before 48 h of extreme event which occurred on 22 June 1982. Negative relative humidity prevails over the eastern part between latitudes 22.5° N to above 30° N near 65° E which is common to other extreme events.

The second most important predictor is the location of the SP of PW having an α value equal

to 0.6650. Figure 5.5(b) shows the surface level PW pattern over the domain, 48 h before the extreme event of 22 June 1982. It can be observed that there is high PW content over a region close to Mumbai. The most important predictor by the anomaly value is the surface level V-wind with PP having a φ value of 0.4857, the next being the 850-hPa level V-wind SP and the 850-hPa level PP. Figure 5.5(c) shows the surface level V-wind pattern over the domain, 48 h before the extreme event on 22 June 1982. It can be observed from figure 5.5(c) that high-velocity winds near the east coast are present, which is common to most of the extreme events; it is possible that V-wind in the region having similar velocity may be an important predictor of extreme events.

Table 5.2 shows the most important predictors of heavy rainfall over Mumbai, when all the given atmospheric levels are considered. The average individual score for the above variables from 50 training events is given in Table 5.3; the higher the individual score of the variable, the higher is its overall importance in predicting extreme rainfall event.

From Table 5.3, it can is observed that, among all variables, the negatively anomalous U-wind at surface level has a maximum score of 1.11, hence is the most important predictor of extreme events. U-wind is followed by negative anomalous V-wind at the 850-hPa level with a score of 1.04.

(a) Relative humidity at 850 hPa(%) **(b) Precipitable water at surface (kg/m^2)** **(c) V-wind velocity at surface (m/s)**

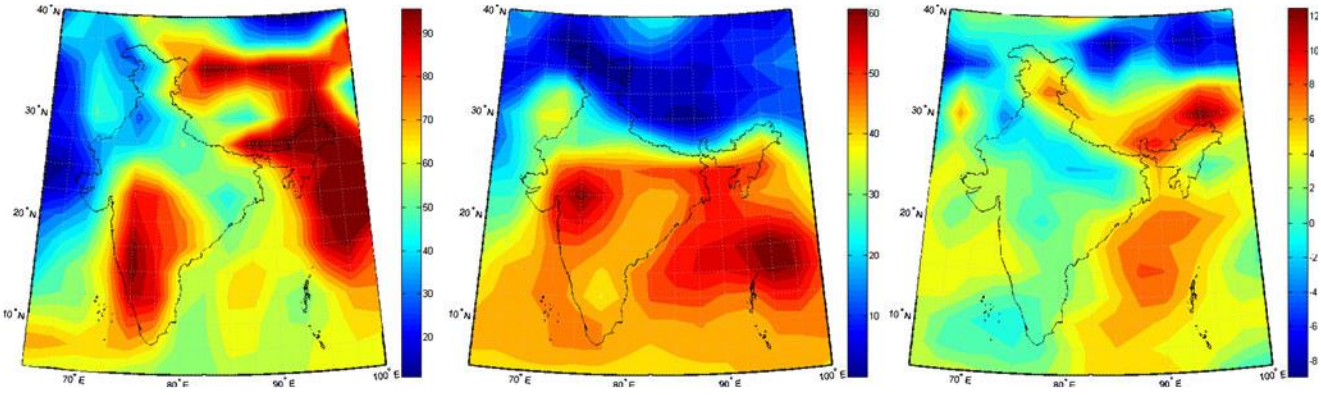


Figure 5.5 Fingerprints of key predictors before extreme events:
(a) relative humidity at 850 hPa (in percent), (b) PW at surface (in kilograms per square meter), and (c) V-wind velocity at surface (in meters per second). The SV of relative humidity, SP of PW, and PP of V-wind are the three most important predictors for fingerprinting analysis

Similarly, the most important atmospheric variables before different time intervals are determined: Table 5.4 shows the contribution from each important variable towards the prediction of extreme events before 36, 24, 12, and 6 h of the occurrence, respectively.

Table 5-2 Most important predictors for fingerprinting

Climate variable	Peak type	Alpha (α)	Phi (φ)
U-Wind_Surface	PP	0.54230	0.40379
V-Wind_850-hPa	SP	0.649364	0.46566
V-Wind_Surface	SP	0.648426	0.29718
V-Wind_850-hPa	PP	0.632212	0.485784
V-Wind_850-hPa	PV	0.633954	0.338598
Precipitable Water_Surafce	PV	0.61841	0.318942
Precipitable Water_Surface	SP	0.6650	0.047502
Rhum_850	PV	0.70551	0.024804

It can be observed that negative anomalous U-wind and V-wind consistently provide a good contribution to the fingerprint of the extreme event as their score values are comparatively high. Also, before 36 and 24 h of the occurrence of events, the positive anomalous relative humidity score increases at the 850-hPa level which signifies that there is an increase in relative humidity before 36 and 24 h of occurrence of the event. PW is also a significant indicator of extreme event; at few locations it decreases, while at other locations it increases before 48 h of event occurrence; however, it does not show a decrease before 6 h of event occurrence. Although there is a negative anomalous MSLP present at a certain location before 12 h of occurrence of the event, there is an increase in MSLP before 6 h of the occurrence of the event.

Table 5-3 Individual score of most important predictors before 48 h

Climate variable	Peak type	Score
V-Wind_850	PP	0.7920565
V-Wind_Surface	SP	0.7106992
V-Wind_850-hPa	PV	1.0446386
V-Wind_850-hPa	SP	0.9798098
U-Wind_Surface	PV	1.1117549
PPWTR_Surface	PV	0.4463701
PPWTR_Surface	SP	0.5812755
Rhum_850	PV	0.227172

The fingerprinting technique eventually determines the event score of any rainfall event based on the match of the weather pattern with the fingerprint of extreme rain-fall event. Depending on the event score, the weather pattern will correspond to extreme rainfall event or non-extreme rainfall event. For the purpose of classification, past extreme events are used to train the model (hind casting) and their scores are calculated. All the 50 training events are used to train the model and thresholds are obtained. The threshold is determined as the 1st quartile or 25th percentile of all the 50 scores. The thresholds for 48, 36, 24, 12, and 6 h before event occurrence are set as 5.26, 6.85, 4.83, 4.85, and 5.38, respectively. The fingerprinting technique is validated using data from 2000 to 2008. A total of 4,392 (9 years×122 days per year×4 parts of a day) instances are present for model validation. The model predicted more than 900 instances as extremes; however, there are only 16 extreme events in the validation data set. Overall, the fingerprinting technique generates a large number of false alarms for Mumbai and, hence, is not implementable. There are few discrepancies in the method-ology which may impede in the exact prediction of extreme events. Fingerprinting calculates α and φ metrics on the basis of extreme departure from the normal; only peak and valley anomalies are considered. These peaks and valleys may occur on any of the non-extreme days which are not verified. The model is calibrated only with the extreme events, and it may be possible that the PP/SP/PV/SVs may also correspond to the non-extreme days. As the model is not calibrated with the non-extreme events, it fails to predict correctly non-extreme days. Hence, there is a need to use the appropriate model for the prediction of extreme events using appropriate predictors, which may be identified not by the positions of PP or PV, but by careful investigations of weather patterns before not only extreme events, but also before non- extreme events.

Table 5-4 Individual score of most important predictors

Climate variable	Peak type	Score
Before 36 h		
AirTemprature_Surface	PV	0.368948781
PPWTR_Surface	PV	0.553225719
PPWTR_850-hPa	SV	0.418404224
MSLP_Surface	PP	0.437826703
MSLP_850-hPa	SV	0.372663564
U-Wind_850-hPa	PV	1.049401092
U-Wind_Surface	SV	1.38020148
V-Wind_Surface	PV	0.930699221
Table 5-4 continued		
V-Wind_Surface	SV	0.804723383

Rhum_850	PP	0.970292614
Rhum_850	SV	0.559064671
Before 24 h		
V-Wind_850-hPa	PP	0.8959578
V-Wind_Surface	SP	0.7590087
V-Wind_Surface	SV	0.6059619
U-Wind_850-hPa	SP	0.8815831
Rhum_600-hPa	SP	0.666301
PPWTR_850-hPa	PV	0.5365885
PPWTR_Surface	SP	0.7146289
PPWTR_850-hPa	SV	0.5558639
Before 12 h		
PPWTR_Surface	PV	0.5655226
MSLP_Surface	PV	0.4558538
MSLP_Surface	SV	0.5158948
U-Wind_850-hPa	PV	1.4429308
U-Wind_850	SV	1.093575
V-Wind_Surface	PV	0.7093148
V-Wind_Surface	SV	0.8882968
Before 6 h		
V-Wind_Surface	PP	0.67595
V-Wind_Surface	PV	0.8117482
V-Wind_Surface	SP	0.8504578
V-Wind_850	SV	1.1930083
U-Wind_850	SP	0.7988597
MSLP_Surface	PP	0.4927509
MSLP_Surface	SP	0.5385422
PPWTR_850	SP	0.6103919

5.5.2 Feature extraction through AFM

Anomaly frequency of all the variables at all the grid points for time instances of 84 to 6 h at 6-h interval before the extreme event occurrence is carried out to determine the variables and grid points which can be the potential weather features that cause extreme events. Using 50 extreme events for calculating anomaly frequency at each grid point for all variables, it is found that some variables had a maximum frequency of around 32 consistently. Having a frequency of 32 for a variable on a grid point at some time instant implies that, out of 50 extreme events, anomalous behavior of the variable is related to only 32 extreme events; the variable does not show any anomalous behavior before the remaining 18 extreme events. Since this happens with

many variables, an investigation is carried out to determine which anomalies are related to which (32 out of 50) events. Interestingly, it is observed that the 32 events, which are related with positive or negative anomaly of different variables on a grid point at some time instance, are the same. It is also observed that the rest of the events are related with anomalies of the same variables but at some other time instances and the anomalies types may be the same or different. Also, it is observed that the first 32 events occurred in the nighttime (0600–1800Z), whereas the last 18 events occurred in the daytime (1800–0600Z). Now onwards, the 32 and 18 events are named as nighttime and daytime events, respectively. Table 5.5 shows the variables which have different anomaly types (positive and negative) during night events and during day events at different time instances on the same grid point.

Table 5-5 Variables having different anomaly types at different time instants before the event

S. no.	Variable	Level	Grid point (GP)	Hours before event (night events)	Anomaly type (night events)	Hours before event (day events)	Anomaly type (day events)
1	Air temperature	Surface level	90	72	Negative	72	Positive
2	Air temperature	850-hPa	70	12	Positive	48	Positive
3	Air temperature	850-hPa	90	24	Negative	24	Positive
4	Air temperature	850-hPa	105	24	Negative	24	Positive
	Table 5	continued					
5	Rhum	Surface level	54	24	Positive	6	Positive
6	Rhum	850-hPa	54	48	Positive	6	Positive
7	U-wind	Surface level	50	24	Negative	78	Negative
8	U-wind	Surface level	3	72	Negative	78	Negative
9	V-wind	Surface level	24	84	Negative	84	Positive

10	V-wind	850-hPa	29	24	Positive	24	Negative	

From Table 5.5, it can be observed that air temperature at the surface level has a negative anomalous value at 72 h before the event at GP90 during night events, and it has a positive anomalous value during day events. Similarly, air temperature at the 850-hPa level on GP90 and GP105, V-wind at the 850-hPa level at GP24, and V-wind at the 850-hPa level at GP29 have contrasting anomalies during night events and day events. Analyzing the table provides very important information which the fingerprinting method could not capture: there are two weather patterns which cause extreme rains over Mumbai- one that is associated with night events and the second that is associated with day events.

All the atmospheric variables are analyzed using the anomaly frequency technique, and the variables with high frequency (more than 23 out of 32 and more than 12 out of 18) are selected as features of extreme events; for example, air temperature at the surface level on GP90 before 72 h of the extreme event having a frequency of more than 23 is a feature of night extreme events. These features are identified at all the four atmospheric levels considered in the study. The temporal variation of features having a frequency of more than 30 (out of 32) and 16 (out of 18) is analyzed. Figure 5.6(a) shows the median temporal variation of surface level air temperature on GP92 during night events.

Frequency of positive anomalies of air temperature at GP92 before 12 and 36 h is 32, i.e., all the night extreme events have positive anomalous temperature at GP92 be-fore 12 and 36 h of the event. Figure 6b shows the shaded contours of air temperature before 12 h of the extreme event which occurred late night 21 July 1977. In figure 5.6(b), the arrow mark shows the location of GP92 for which the temporal variation is shown in figure 5.6(a). GP92 and surrounding regions have a high temperature before 12 h of the event. Omega- vertical wind velocity has high frequency at many grid points, hence provides features of extreme events. Figure 6c shows the temporal variation of the 850-hPa level omega on GP25 12 h before the night extreme events. The frequency of negative anomaly of the 850-hPa level omega on GP25 before 60 h of extreme event is 31. Conditions are similar at GP39. Figure 6d shows the shaded contour map of omega before 12 h of the previously mentioned extreme event. The arrow mark shows GP25 for which the temporal variation is shown in Fig. 6c. The frequency of negative anomaly of the 850-hPa V-wind on GP25 before 60 and 84 h of the events is 31 and 30,

respectively. Similar conditions are observed at GP24 and GP26. Figure 6f shows the shaded contour map of V-wind before 12 h of the previously mentioned extreme event. The arrow mark shows GP25 for which the temporal variation is shown in Fig. 6e. It is interesting to note that the climate variables at GP24 and GP26 are also among the features of extreme events. GP25 lies in the Tibetan range, which has a considerable impact on southwest monsoons (Flohn 1968). U-wind speed, which is more than 10 m/s at these locations, may provide some link between heavy rains over Mumbai and Tibetan mountains.

As mentioned earlier, the weather patterns responsible for the extreme events during the night is quite different from that during the day. This is evident from figure 5.7. One reason may be diurnal temperature variation which affects the process. Hence, a similar analysis with AFM is performed for the extreme events during the day for predictor identification. The frequency of positive anomaly of surface level air temperature on GP89 before 72 and 48 h is 16 out of 18 day events. The frequency of negative anomaly of surface level air temperature on GP9 before 72 and 48 h is 16.

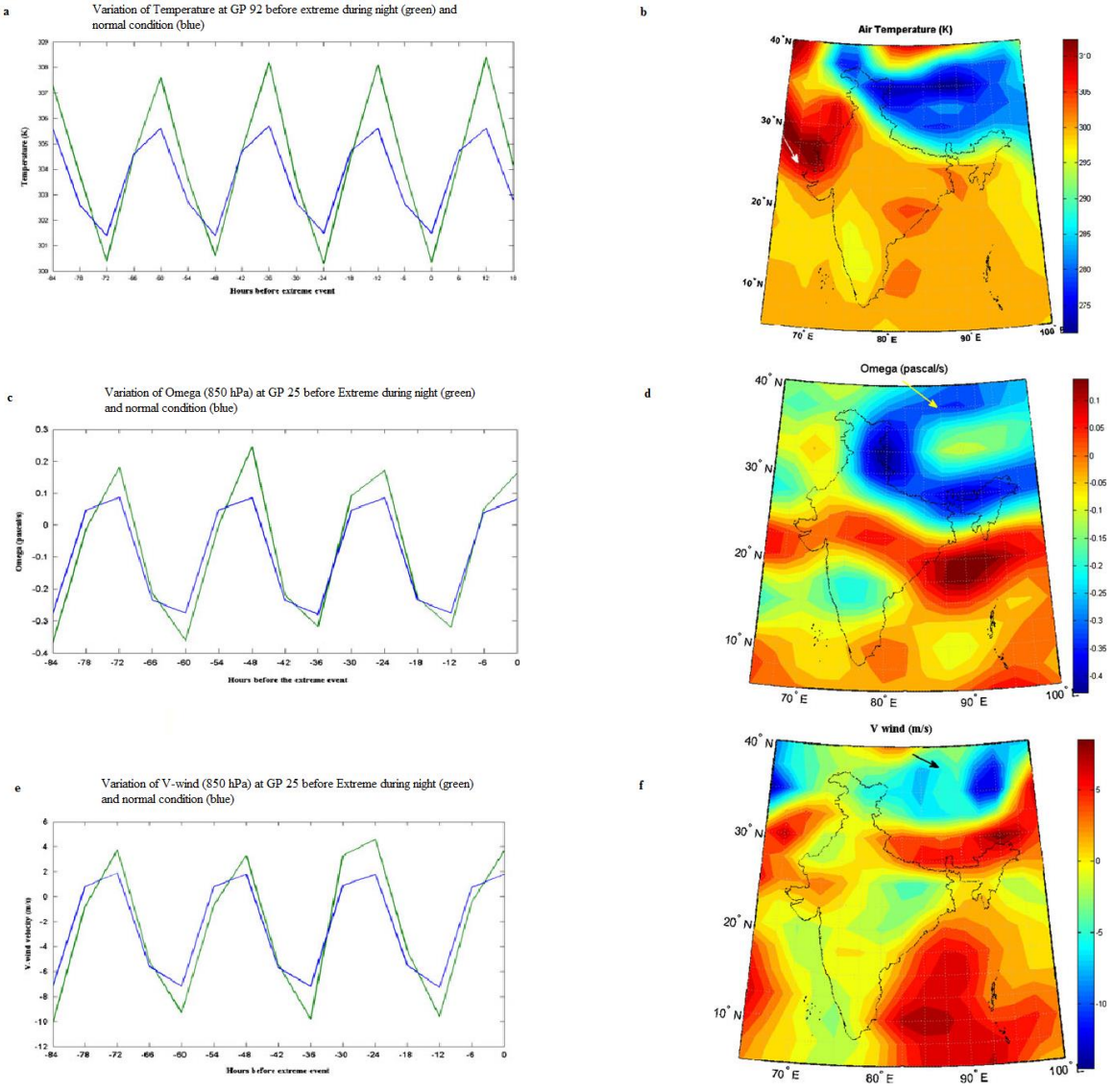
AFM is applied to other derived atmospheric variables which have a significant contribution in extreme event occurrence (Joseph 2006). It is interesting to note that there are certain variables which have significantly high anomaly values at the neighboring locations of Mumbai before extreme events, irrespective of their timing (day or night). These variables include zonal wind shear and vorticity. Figure 5.8 (a–c) shows that vorticity patterns around Mumbai before the extreme event during both the daytime and night-time are similar, but quite significantly different from the normal condition. Figure 5.8(d) shows the frequency of positive anomalies of vorticity at the 850-hPa level over all 225 grid points.

From figure 5.8(d), it is observed that GP141 has a highest frequency of 41 which implies that 41 events out of a total of 50 show a positive anomalous behavior of the 850-hPa level vorticity on GP141 at 6 h before the event. It can also be observed that grid points GP123 to GP126 and GP138 to GP141 have a frequency near 35 at 6 h before the event; these grid points are around Mumbai, which suggests that, during extreme events, the regions around Mumbai are having very high vorticity before 6 h. Similarly, AFM is applied to analyze zonal wind shear all over 225 grid points; it is observed that zonal wind shear has a high frequency of positive anomalies at GP126 and GP141 before 6 h of the extreme event. Figure 8e shows the temporal variation of vorticity around Mumbai (GP123 to GP126 and GP138 to GP141). It is

evident that the increase of the 850-hPa vorticity is mild till 18 h before the event, but after that it becomes steep. Vorticity is always much higher than the climatological mean (blue line). Similar results are also obtained for zonal wind shear at 850 hPa.

Both vorticity and zonal wind shear at the 850-hPa level have high anomalous values during almost all the 50 events before 6 h of the event. These also form important features of extreme rainfall events. After identifying the features of extreme events, the SVM classifier is used for training.

Figure 5.6 Key predictors used in SVM for predicting extreme events during night
(a) air temperature variation at grid point 92 (GP92) before the extreme event (green) and during normal condition (blue); (b) arrow shows the location of GP92 with the spatial variation of air tempera-ture, (c) omega (at 850 hPa) variation at GP25 with an arrow showing its location in (d); (e) 850 hPa V-wind variation at GP25, with an arrow showing its location in (f)



To start the AFM, all the available variables from the NCEP/NCAR reanalysis project are considered. It is observed from the results that few of the atmospheric variables contribute to the extreme weather patterns, and the most critical variables are presented in Table 5.5. They are mostly air temperature, humidity, U-wind, and V-wind. A possible physical explanation could be that temperature differences (high anomaly temperature in a region) may result in the formation of low-pressure zone which causes high wind velocity with moisture (commonly known as vertically integrated moisture content [VIMT]). It is observed in the literature (Fasullo and Webster 2003; Konwar et al. 2012) that VIMT is highly associated with Indian rainfall and the variables responsible for high VIMT are actually identified by the AFM.

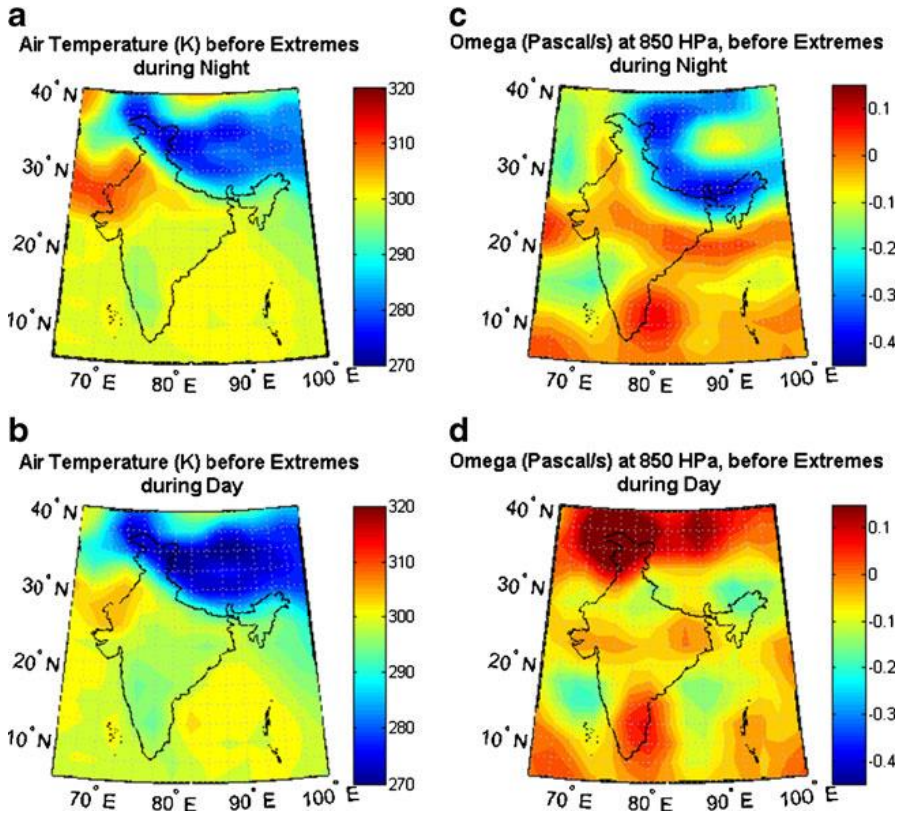


Figure 5.7 Different weather patterns before extreme events during the day and night (a, b) air temperature, (c, d) omega at 850 hPa

5.5.3 Two-phase SVM results

Trial and error is performed to select the best SVM parameter, viz., kernel functions, the sigma value (similar to band width) of the radial basis kernel function used, etc. Initially, training of the SVM classification models is performed using the values of all the features (having a frequency more than 23 for SVM1 and having a frequency more than 12 for SVM2). Validation of the models is carried out for all the instances in the validation period 2000–2008. The validation results show poor performance for the best model obtained with trial and error: out of 4,392 instances, around 1,000 are predicted as extremes where in actual only 16 instances are extremes; this prediction is not acceptable. Also, there are very few extremes which are predicted with lag of 6 h or more.

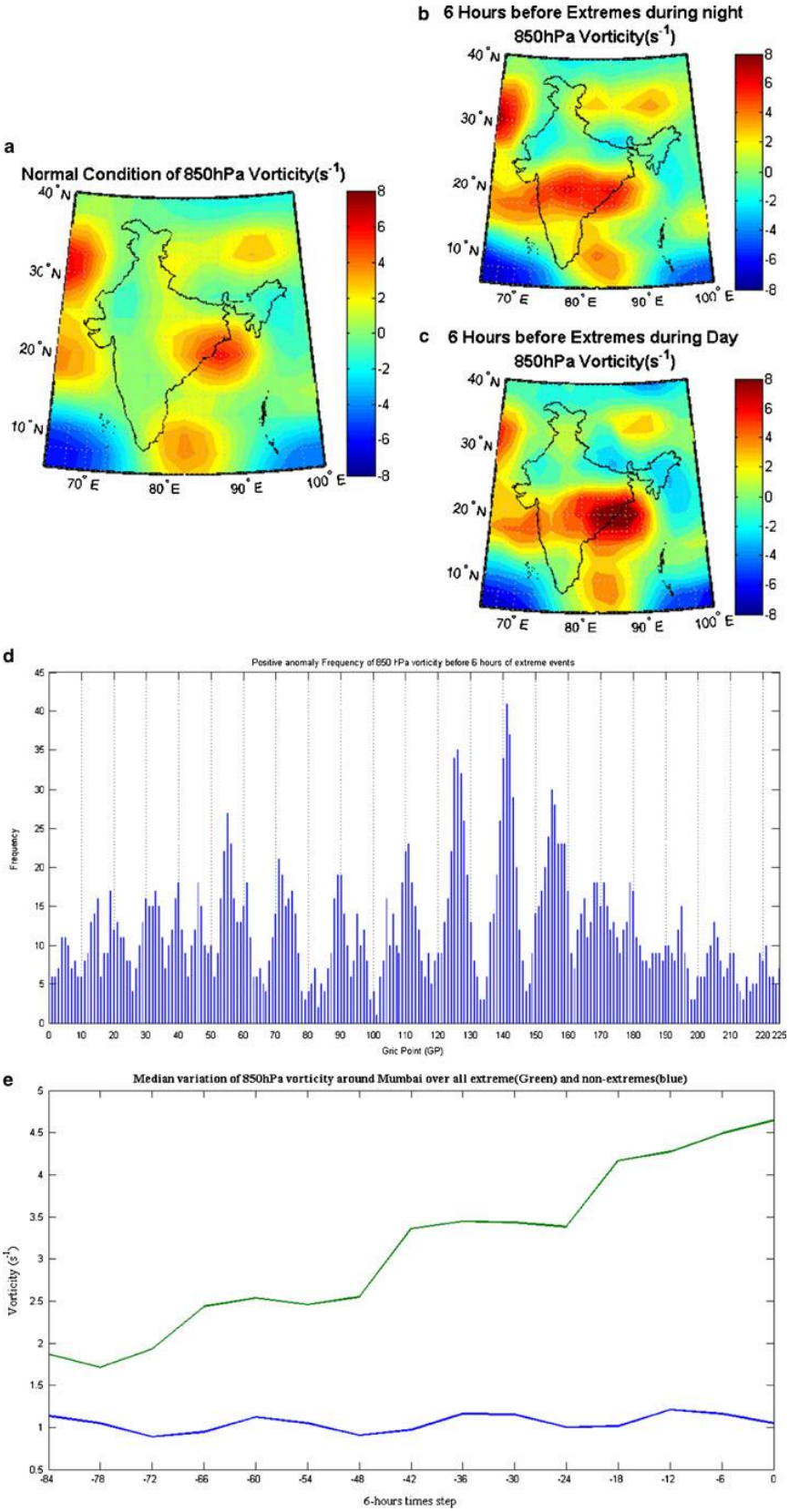


Figure 5.8 Vorticity around Mumbai with similar spatial pattern before extreme events

During both night (b) and day (c), which are different from normal condition (a). (e) Frequency of positive anomaly of 850 hPa vorticity, at different grid points, 6 h before the extreme events, 50 extreme

events are considered for this. (f) Median temporal variation of 850 hPa vorticity at neighboring regions of Mumbai during all (day + night) extremes (green) and non-extreme (blue)

The reasons behind such poor performance may be: out of all features, few have a high potential triggering extreme event and the rests have a lesser potential for extremes. Furthermore, some features might not be too different from non-extremes to be differentiated by SVM. To cope up with the first caveat, we assume that features with high frequency are better predictors than features with low frequency, and hence, we introduce a weighting factor to the feature matrix for prediction depending on the frequency. The weights to the features are assigned between 0 and 1 depending on the frequency. The higher the frequency of the feature, the higher is its weight. The summation of weights of all features is equal to 1. Using weighed features, the hit and trial method is applied to arrive at the best possible model. Classification is improved considerably using weighed features; out of 4,392, 424 events, as compared with the 1,000 in the first case, are classified as extremes. Also, only two events are predicted 12 h ahead; however, the classification performance is still not satisfactory. From the results obtained with the two experiments, it is assumed that feature values of extremes and feature values of non-extremes may not have enough difference that SVM models can discern. This is also observed in figures comparing extremes and normal weather. The differences are small for some predictor variables, such as 850 hPa omega at GP25, although it is statistically significant at the 1 % level. This difference, even though statistically significant, may not be very well captured by the SVM models. One possible reason may be that these values may be similar to the outliers for non-extreme cases and the soft margin fails to classify them. Since the intuition of SVM for finding maximal hyper plane (training) is geometric (Boswell 2002), it may be possible that scaling the feature values of extremes and non-extremes may provide better classification. The difference between the variable values and anomaly threshold are taken to constitute a new feature vector instead of raw features. Equation 13 calculates the difference between threshold and raw feature value:

$$\boldsymbol{f} = \boldsymbol{x} - \boldsymbol{\delta}^{\pm} \qquad 13$$

where f is the new feature value, x is the raw feature values, i.e., the value of atmospheric variable corresponding to feature, and δ± is the positive or negative anomaly threshold.

Table 5-6 SVM1 (night events training) training with different frequencies

Frequency	Total events predicted	Actual extremes predicted in testing period[a]	SVM1 kernel used
23	459	3	Linear
24	547	4	Linear
25	460	3	RBF
26	307	6	RBF
27	252	7	RBF
28	108	10	RBF
29	67	12	RBF
30	210	9	RBF
31	1,002	12	RBF
32	1,072	8	RBF

[a] Actual number of night extreme events in validation/testing period is 12

Table 5-7 SVM2 (day events training) training with different frequencies

Frequency	Total events predicted	Actual extremes predicted in testing period[a]	SVM2 kernel used
12	650	7	Linear
13	178	5	Linear
14	82	4	Quad
15	122	6	RBF
16	245	9	RBF

[a] Actual number of day extreme events in validation/testing period is 4

Using the feature obtained from Eq. 13, experiments are performed to determine which feature will provide the best results. All the features (having a frequency of 23 to 32 for SVM1 and 12 to 18 for SVM2) are used to compute the optimum frequency for best prediction performance. Table 5.6 shows the classification performance of best SVM1 models using features of different frequencies. Table 5.7 shows the classification performance of the best SVM2 models using features of different frequencies. It is observed from Table 5.6 that the SVM1 model, trained with the feature values corresponding to frequencies of more than 29, provides the best result—67 predictions in total, 12 being the actual extremes. Similarly, from Table 5.7, it is observed that the SVM2 model, trained with feature values corresponding to frequencies of more than 14, provided the best result—82 predictions in total, 4 being the actual

extremes. The details of best (final) SVM models (as obtained with trial and error) are presented in Table 5.8. It is observed that, when the support vectors are further analyzed, for SVM1, all extreme features are support vectors which give an indication that further improvement is required in feature extraction for better classification of rare events using SVMs.

A total of 133 non-extremes are predicted as extremes. Prediction of 149 (133 non-extremes+16 extremes) events as extreme events is quite satisfactory as compared with fingerprinting technique prediction which predicted 908 events as extremes. Large numbers of consecutive instances (which are non-extremes) are predicted as extremes in fingerprinting techniques, whereas such limitation is not observed in the present method for most of the cases. The average rainfall during the extremes predicted using the fingerprinting technique is 4.50 mm per instance (6 h), whereas average rainfall during extremes predicted through the two-phase SVM is 20 mm per instance (6 h). These results suggest that the present two-phase SV classifier is modeling extremes much better than the fingerprinting technique. These results are encouraging towards predicting extreme rainfall through weather pattern recognition. The results may be further improved with the use of high-resolution weather data (predictors) as well as Doppler radar data. Development of proper data assimilation techniques with fine-resolution data from multiple sources and using them in fine-resolution urban extreme weather prediction is a potential research area and may be considered as the future scope of the present work.

It should be noted that the region for predictors are selected arbitrarily. The region selected is significantly large with a primary assumption that the AFM will select the critical grids. We also perform another experiment, where we consider a larger region, not only considering just the subcontinent but also the entire Arabian Sea (−2.5° to 45° N and 55° to 102.5° E) and redo the same analysis. Almost similar set of variables are identified as key predictors, with slightly different critical locations, and it reduces the number of false alarms; however, it fails to generate the alarm for an extreme event. This is probably because the weather patterns responsible for this specific event get lesser weight for considering higher area and, hence, are not participating in SVM. A comparison of results for both regions is presented in Table 9.

The other limitation of the model results from the avail-ability of the data. The length of the data used for the present analysis is 1979–2008, 40 years data. As the occurrences of extremes are rare, it limits the use of a significantly long data set for validation, as for the training and development of a model, a good number of extremes is required. Due to the same reason, only 16 extremes could be used for validation, out of which 4 belongs to daytime events derived with SVM-2. Validation of the model with four extremes may not provide a clear picture on

model performance; however, if we increase the validation period, it will result to a training data set with <10 number of daytime extreme events, which may result in an inaccurate model. Also, it is a common practice in hydroclimatology to use a minimum of 30 years as baseline (Wilby et al. 2004; Dibike and Coulibaly 2005) or training data set for a model and the same is used here. To check the model performance with shorter length training data set, the entire data set is divided into two equal halves, 50 % for training and 50 % for validation. For the new set, it is observed that, out of 33 extreme events in validation, 26 are predicted, and this may be because not all types of extremes in training (with less training data) were considered.

Table 5-8 Best SVM architecture

SVM1		SVM2	
Kernel function	RBF	Kernel function	Quadratic
Kernel function argument (sigma)	0.8900	Bias	0.9489
Bias	0.3999	Support vectors	45×4
Support vectors	48×32	Optimization method	SMO

Table 5-9 Sensitivity of selection of regions for predictors

Region for predictors	Total no. of instances in validation period (2000–2008)	No. of actual extreme in validation period	Total number of extremes predicted	No. of actual extremes predicted as extremes	False alarms	Actual extremes predicted as non-extremes
Indian subcontinent	4,392	16	149	16	133	0
Arabian Sea + Indian subcontinent	4,392	16	62	15	47	1

5.6 Summary and concluding remarks

The hypothesis for the prediction of extreme rainfall events is that the weather pattern(s) causing extreme rainfall are entirely different from the normal weather days. Weather patterns associated with past extreme events are identified using pattern recognition techniques. Based on the resemblance of a future day weather pattern with that of extreme events, the future day

weather is predicted as extreme or non-extreme. The fingerprinting technique is used to predict six-hourly extreme rainfall events over Colaba station, Mumbai. Highly anomalous U-wind and V-wind around Mumbai and low-pressure regions near the northwest Bay of Bengal are identified to be the most important variables in producing extreme rains over Mumbai. Although the finger-printing method does well to predict all 16 extreme events in the validation period (4,392 instances), more than 900 non-extreme instances are predicted as extreme events. Hence, the use of the fingerprinting technique to predict extreme rainfall events over Mumbai in the future may be misleading. There may be many reasons for the poor performance of the fingerprinting technique in predicting rainfall events over Mumbai, the most appropriate are as follows: (1) the fingerprinting method is calibrated for extreme days only and (2) fingerprinting identifies only one weather pattern corresponding to extreme events and Mumbai has two weather patterns responsible for causing extreme rains. Identifying two weather patterns is possible using respective events as training instances for fingerprinting, but there is a disadvantage that the training data sets will be small for both of the weather patterns, hence results will be highly uncertain given that the method employs only extremes for calibration. AFM is developed in the present study to determine the variables and the grid points which are consistently being affected during extreme rainfalls over Mumbai. AFM is applied on past 50 extreme rainfall events over Mumbai to extract the features of extreme rainfall events over Mumbai. It is observed that, during night events, temperature is higher than normal before 12 h of extreme event in region near 25° N, 67.5° E. Negatively anomalous 850-hPa omega before 60 h and positive anomalous surface level U-wind and V-wind before 12 h of extreme events around the Tibetan plateau are also identified to be critical in triggering extreme rainfall events over Mumbai. During almost all the extreme events, it is observed that vorticity and zonal wind shear around Mumbai start increasing from 84 h before the event and reach their peaks just before event occurrence; hence, it is asserted that steep increase in vorticity and zonal wind shear, especially at lower atmospheric levels, over Mumbai is a precursor of extreme event. It is concluded that Mumbai extremes are not caused by exceptional variation of a single atmospheric variable but different atmospheric variables get altered and interact in a complex manner to trigger short-term extreme rains over Mumbai. AFM yields an important conclusion that there are two weather patterns causing extreme rainfalls over Mumbai. In this context, a two-phase SVM is developed in this present study. SVM is not able to classify extremes and non-extreme by using mere variable values as features matrix for training, even though the features of extremes and non-extremes are different with a high level of statistical significance, hence weighing and scaling the features is sought to provide better classification. Based on the

hit and trial method of training applied to weighed and scaled features, a best two-phase SVM is formulated, which predicts a total of 149 extreme events in the validation period out of which 16 are actually extreme. The two-phase SVM model shows a significant improvement over the finger-printing technique in predicting extreme events over Mumbai; however, there is still a good number of false alarms. This may be improved with the use of high-resolution weather pattern or with the use of Doppler radar data. Statistical development of a rare event classifier may also improve the performance. However, the use of multiple data may need proper data assimilation techniques, which should be further coupled with the event classifier. The quality of reanalysis data for extreme events is also another major factor, which affects the performance.

It should be noted that fine-resolution atmospheric models may provide better estimates with less number of false alarms. However, due to the unavailability of fine-resolution observed/reanalysis data, the coarse-resolution data is used. A possible alternative option may be the use of fine-resolution Climate Forecast System Reanalysis da-ta set (http://rda.ucar.edu/pub/cfsr.html), but the data is not available for the entire 40 years. Further reduction of data length will seriously affect the model performance and hence is not practiced for the present analysis.

Chapter 6 : COUPLED IMPACTS OF CLIMATE CHANGE AND URBANIZATION ON EXTREME RAINFALL IN MUMBAI

6.1 .Introduction

The intensity and frequency of extreme rainfall events are reported to increase over the most parts of the world (Alexender et al 2006). The probability of extreme precipitation events is increased by a factor of about 2 by the end of the 21st century (Kharin and Zweirs (2004; 2007). These changes largely associated with changes in location over most of the globe Kharin et al (2013). The increase in annual maximum daily rainfall intensity was further estimated to occur at a higher rate in tropics and high latitudes of Northern hemisphere (Groisman et al 2005). The longer time scale output from General Circulation Models (GCMs) also suggests the intensification of extreme rainfall events with global warming (Meehl et al 2007). At the same time the very large intermodal disagreement in the tropics suggests that dome physical processes associated with extreme precipitation are not well represented in the models (Kharin and Zweirs, 2007). Overall the sensitivity of precipitation extremes to a warming climate remains uncertain with important regional variations (O'Gorman 2015). This highlights the need of observations and simulations of the physical factors that govern the precipitation extremes.

Though with these realizations of the importance to understand the occurrences of extreme rainfall events from both scientific and impact cantered perspective, studies resolving the physical processes that cause sub daily extremes are scarce (Wakazuki et al 2008; Kendon et al 2010). A complete depiction of how sub daily extreme rainfall might change in the future remains unresolved (Boucher et al 2013; Collins et al 2013). The theoretical basis of extreme rainfall increases with that of air temperature can be explained using Clausis and Clapeyron (CC) theory. As temperature increases the intensity of extreme rainfall increases as warmer air capable of holding more water and has the potential to provide more moisture to rainfall events (Trenborth et al 2003). Though the moisture availability does not increase endlessly with temperature and may limit the extreme rainfall intensification above a specific temperature threshold. On the global scale the atmosphere energy balance sets a limit to mean precipitation change by 1-3% per degree global average surface temperature (Westra et al 2014).

Climate modeling studies have difficulty in resolving the processes that cause short duration rainfall extremes while simultaneously modeling the large scale conditions in which local

extreme rainfall events are embedded (Westra et al 2014). Limited number of studies which indicate sub daily extremes to increase is due to increases in air temperature (Lenderik and Van Mejigard , 2008).

The first formal detection of human influence on observed daily extreme rainfall intensification was provided by Min et al (2011). In this study the human-induced increases in greenhouse gases have contributed to the observed intensification of heavy precipitation events over approximately two-thirds of data-covered parts of Northern Hemisphere land areas (Min et al 2011). Significant amount evidence exists that links specific extreme weather events to anthropogenic characteristics is continuing to build up (Coumou and Rahmstorf, 2012; Min et al 2011). The evidences are obtained for sub-daily extreme rainfall intensification due to anthropogenic activities and increases in extremes with rise in atmospheric temperature (Westra et al 2014).

Regional climate models (RCMs) provide higher-resolution climate projections that partially resolve finer scale variability associated with topography and land cover and hence are increasingly being used in studies aimed at helping society adapt to climate change (Salathé et al., 2010). The extreme rainfall events simulated by RCMs (Regional Climate Models) may be more reliable due to better representation of smaller-scale topographic features and physical processes (Gutowski et al., 2010; Leung et al., 2004). However, these studies remain scarce underlying the huge computational requirements (Wakazuki et al 2008; Kendon et al 2010). Various local, regional and continental level studies are undertaken the projected changes in occurrences of extreme rainfall events. The selected return levels were obtained over the Canada (Mladjic et al 2001); the impact of future climate change on ISMR (Rajendran and Kitoh 2008), North American Regional Climate Change Assessment Program (NARCCAP) (Gutowski et al 2010, Mishra et al 2012) over Europe (Deque et al 2007) , UK (Fowler and Wilby 2010) are to name a few.

Several studies investigate the occurrence extreme events with a statistical modeling approach. Washington Luiz Félix Correia Filho et al (2014) diagnosed the extremes over Northeast Brazil (NEB) with application of statistical model and observed the maximum temperature, zonal wind component, evaporation, specific humidity and RH having profound influence on precipitation extremes. More importantly the undergoing the higher inter-annual variability and the factors confounding the same, detection of anthropogenic climate change at regional scale is difficult. The changes and uncertainties in extreme precipitation using three statistical models (Sunyer et al 2015).

Several studies pertaining to Indian summer monsoon rainfall extremes reported in the literature. For instance Goswami et al 2006 reported significant rising trends in the frequency and the 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. Rajeevan et al (2008) supported the hypothesis that the increasing trend of extreme rainfall events in the last five decades could be associated with the increasing trend of sea surface temperatures and surface latent heat flux over the tropical Indian Ocean.

A significantly increasing trend in the frequency of heavy rainfall climatology over urban regions of India during the monsoon season, adding that urban regions experience less occurrences of light rainfall and significantly higher occurrences of intense precipitation compared to non-urban regions (Kishtawal et al. 2010). The trends are further observed to lack of uniformity in occurrences but increasing spatial variability over rainfall extremes (Ghosh et al 2011).

The changes in extreme rainfall characteristics estimated in terms of intensity, duration and frequency of extreme rainfall is reported to exhibit a non-stationarity due to different drivers and spatially non-uniform pattern is observed in the changes over India (Mondal and Mujumdar 2015). Projections of precipitation extremes over India, obtained with a state-of-art regional climate modeling system, PRECIS (Providing REgional Climates for Impacts Studies) reveals marked increase in precipitation towards the end of the 21st century under scenarios of increasing greenhouse gas concentration and sulphate aerosols. The experimentations over PRECIS also observed the changes in terms of a shift in extreme precipitation and dry spells with a substantial increase in the spread having an increased probability of occurrence of both floods and drought events (Rao et al 2013).

6.2 Motivation and Objectives

Different studies highlighted the need for systematic examination of global versus regional drivers of trend in Indian rainfall extremes for flood hazard preparedness and water resource management. The understanding of influence of urban signature on extreme rainfall on two locations Mumbai and Alibag reveal the sensitivity of extreme rainfall events to the increased urbanization (Shastri et al 2014). This study is undertaken as to observe the changes in projection of rainfall extremes corresponding to the development of urbanization. However, these observational studies remain computationally expensive to apply for a longer time period.

We propose an integration of statistical and dynamical downscaling methodology to include the feedback from urban expansions on the occurrence of extreme rainfall.

6.3 Statistical Downscaling

The statistical models are based on statistical relationships between large scale climate features and hydrologic variables. It is assumed that the statistical relation developed during historical period and the same relation is also valid for the future.

As mentioned, the main shortcomings of the dynamic downscaling are that RCMs still require considerable computing resources and are as expensive to run as global GCMs themselves; these models still cannot meet the needs of spatially explicit models of ecosystems or hydrological systems and that there will be the need to downscale the results from such dynamic downscaling models to individual sites or localities for impact studies (Wilby and Wigley, 1997). Moreover, dynamic downscaling is inflexible in the sense that expanding the region or moving to a slightly different region requires redoing the entire experiment (Crane and Hewitson, 1998). However, the 'Statistical Downscaling' techniques are computationally easy and the efficiency of results of statistical downscaling are comparable with 'Dynamic Downscaling' techniques (Schoof, 2013).

6.3.1 Data Used

The statistical downscaling procedure utilizes three preliminary datasets namely: the Observed Rainfall data, Climate reanalysis data, GCM data.

The observed rainfall data at a daily resolution for the station Santacruz Mumbai from 1979-2007 is obtained from India Meterological Department (IMD).

The selection of predictor data set plays important role in the performance of the statistical downscaling model. It is broadly assumed that the predictors directly affect rainfall process. Predictors should be so chosen such that in the climate change context the predictors essentially capture effect of global warming. Humidity plays an important role in capturing changes in water holding capacity of atmosphere under global warming (Wilby and Wigley 1997). Temperature , U wind , V wind, Mean sea level pressure (MSLP) add considerable power to predict short and long term changes in precipitation. The predictors selected for this analysis are air temperature, wind velocities (U and V wind), mean sea level pressure, specific humidity

at surface level. In addition to this the 500 & 850 hPa level U and V wind velocities are included as predictors.

The selected predictor data is obtained from NCEP-NCAR reanalysis and GCM data sets.

6.3.1.1 NCEP-NCAR Reanalysis data

Reanalysis data is surrogate for observed data for any predictor variable. The NCEP-NCAR Reanalysis data set is a continually updating gridded data set representing the state of the Earth's atmosphere, incorporating observations and numerical weather prediction (NWP) model output dating back to 1948. It is a joint product from the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR), NOAA. For the current projections, the reanalysis data was downloaded. The resolution is 2.5º lat × 2.5º long. The base line period considered for the present study is from 1961-2000 which is of sufficient duration to establish a reliable climatology (Ghosh and Mujumdar 2008). The NCEP/NCAR reanalysis-I data (Kalnay et al., 1996) provide global atmospheric data which is a mixture of physical observations and model forecasts. Kalnay et al., 1996 have used different data assimilated systems such as global raw aircraft data, satellite data, and surface land synoptic data, advanced microwave surface wind speed data etc. to with a T62 resolution and 28 vertical sigma levels to calculate the reanalysis data products for various climate variables. The data is available form http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html.

6.3.1.2 GCM data

The original GCM (General Circulation Models) outputs of precipitation as well as the simulations for predictors are obtained from Program for Climate Model Diagnosis and Intercomparison (PCMDI). The present analysis projection is performed with the 5 selected GCMs are as listed with Table 6.1.

The future scenarios considered for this analysis are (Representative concentration pathways) RCP 4.5 and RCP 8.5. Both the scenarios are considered as worst case scenarios for the respective generations of models. The future period considered for this analysis is 2030-2050 and the projected changes are computed with respect to the base line period 1969-2005.

Table 6-1 GCMS used in the present study

Sr.No	Short name	PROJECT
1	GFDL	GFDL –ESM2 Geophysical Fluid Dynamics Laboratory
2	CNRM	CNRM-CM5 Centre National de Recherches Meteorologiques
3	MRI	MRI-CGCM3 Meteorological Research Institute
4	CCCma	CanESM2 Canadian Centre for Climate Modelling and Analysis

6.3.2 Methodology

With the statistical downscaling methodology the coarse-resolution predictors are linked to the fine-resolution predictand using a statistical relationship. The flow diagram for stepwise mathematical operations performed on the data for the rainfall projections are given with figure 6.1.

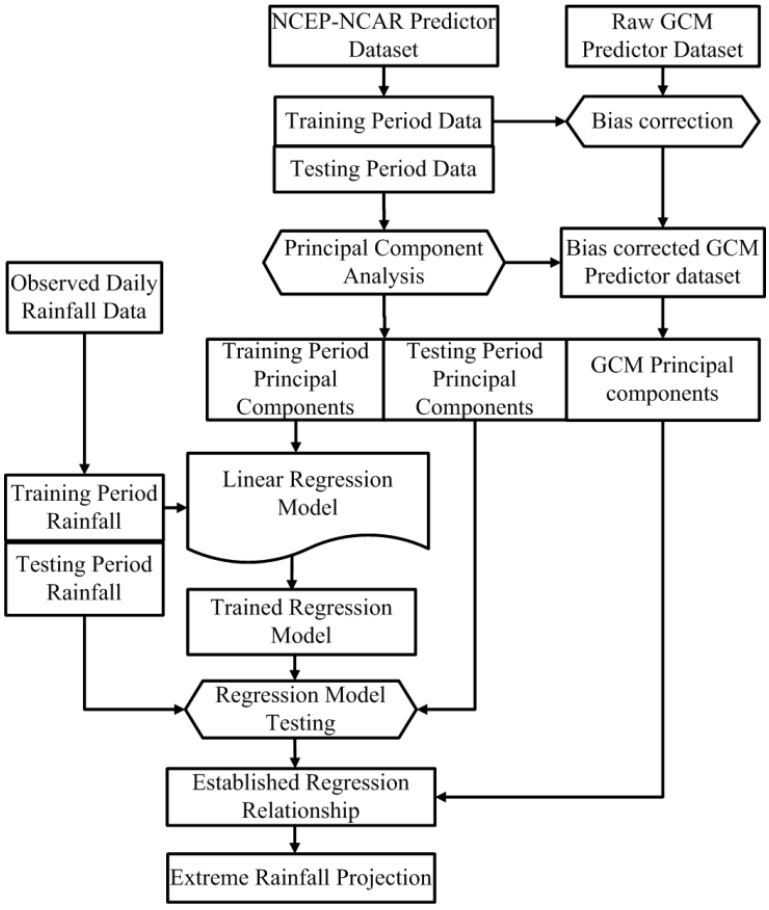


Figure 6.1 Flow diagram describing Statistical Downscaling Methodology

The GCM-simulated predictors and the observed rainfall, as a predictand, undergo different mathematical operations before actually becoming statistically linked. The predictors undergo a bias correction operation where the systematic error is removed using a quantile-based remapping technique (Li et al., 2010). The bias-corrected predictors go through PCA that involves the application of orthogonal transformation on a set of correlated predictor variables, producing principal components. The resulting principal components are dimensionally reduced and uncorrelated to one another. Principal components carry almost the same variability as that of the original data. Hence, the PCA helps to reduce both dimensionality and multicollinearity. A reduction in the dimensions also results in a reduction in the computational effort.

The daily observed rainfall data and the bias-corrected predictors, which all undergo principal component analysis, are key inputs to the regression model for establishing the statistical relationship for the training period. Assuming that the relationship holds for the future, projection are generated; with the help of pre-established relationships and predictors for the future period. By applying liner regression, rainfall is projected for the city of Mumbai. For the model validations, the projections are obtained using NCEP/NCAR reanalysis climate variables as predictors. The period 1969–1984 is the time slice considered as the training period for which the statistical relationship is established. The period from 1985–2007 is the time slice considered as the validation period for which projections are obtained and compared with the observed rainfall data for the same time period. The 20th century projections are obtained from the historic GCM data for the time period slice 1969–2005. Future projections are obtained from the same GCM for the time period from 2030–2050. The future projection for two different scenarios namely RCP 4.5 and RCP 8.5 are calculated.

6.4 Dynamic Downscaling

Dynamic downscaling leads to the development of finer-scale physics-based models known as regional climate models (RCMs) that take input from GCMs' simulations as initial and boundary conditions, incorporate the subgrid features, and produce very high resolution results. Using various RCMs, dynamic downscaling has been attempted successfully for rainfall projections. The methodology has a major limitation of being highly computationally expensive. Here we perform a limited WRF-LSM runs over the city of Mumbai.

6.4.1 Data Used

The dynamic downscaling methodology utilizes two main datasets namely climate data and urbanization development data.

6.4.1.1 Climate data

Initial conditions are provided by European Centre for Medium-range Weather Forecasts (ECMWF) reanalysis or ERA-I data (Dee et al., 2011). The six-hourly outputs (00, 06, 12, 18 UTC) at 0.75°×0.75° horizontal resolution over 30 pressure levels are used for this experiment. Characterization of land surface, which acts as a bottom boundary condition for the coupled model, is an important step towards simulating the Urban Boundary Layer (UBL). We use MODIS land use data set for this study. The coarse domain consists of 74x93 spatial matrix, whereas the domain of interest has a 232x304 mesh covering the Mumbai metropolitan and Alibaug with 4 km horizontal grid spacing.

6.4.1.2 Land use land cover data

The urban agglomeration of Mumbai is one of the largest and fastest-growing urban regions in the world, and this growth has unprecedented effects on urban sprawl and population dynamics (Chakrabarti, 2001; United Nations, 2012). Here, we explicitly addressed the simulation of future urban growth patterns of Mumbai with the dynamic downscaling model. Here we adopt the maps generated by Hossain and Marco; 2013. The prevailing dynamism, of spatiotemporal variation with the landuse mapping conducted over the years 1973 and 2001. The classified urban maps show clear urban expansion and demonstrate that urban growth dynamics are strongly linked to population dynamics. The urban growth modeling is applied to develop an understanding of booming and vital spatial developments in Mumbai. The temporal mapping of built-up areas and the simulations for 2050 indicate that the projected urban expansion occurring between built-up areas and open land and croplands. Here we consider that the regulatory protection takes into account environmental considerations for reserved areas around the Sanjay Gandhi National Park as well as the wetlands around the city. This analysis supports predictions by Taubenböck et al. (2012), who anticipated the emergence of satellite towns. The urban maps obtained for the year 1973, 2001 and 2050 are shown in the figure 6.2.

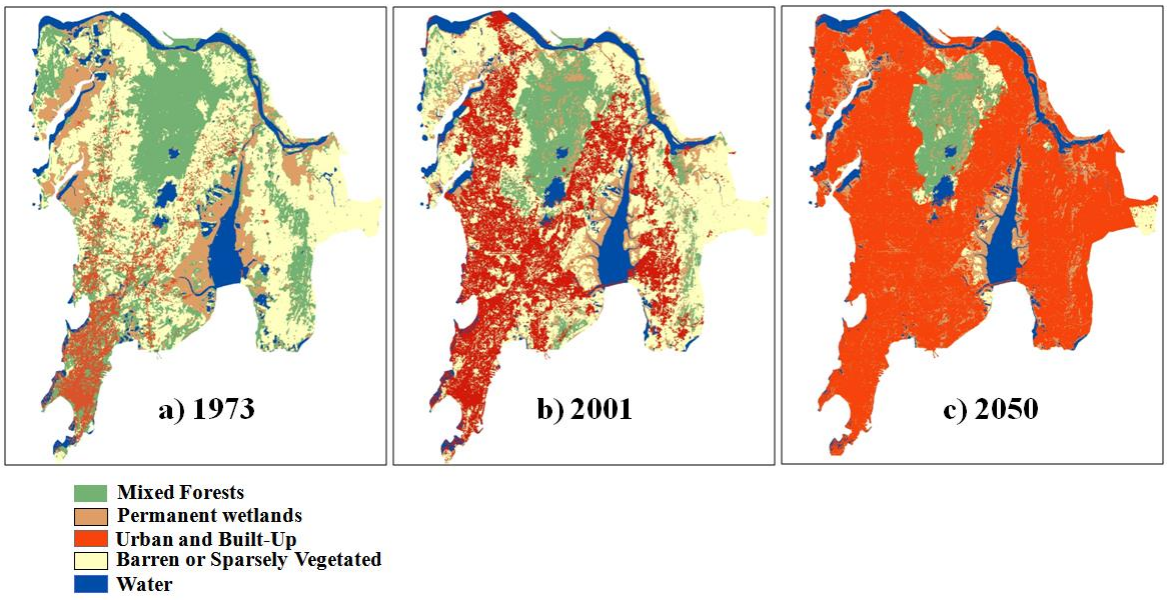


Figure 6.2 Land use map of the city of Mumbai.

6.4.2 Methodology

WRF version 3.6 is used for this study. WRF is a non-hydrostatic terrain mesoscale model (Skamarock and Klemp 2008), used widely for both operational forecast and climate research. The simulations are performed over Mumbai metropolitan area using three nested domains with horizontal grid resolutions of 36 km (d01), 12 km (d02) and 4 km (d03) respectively (figure 6.3).

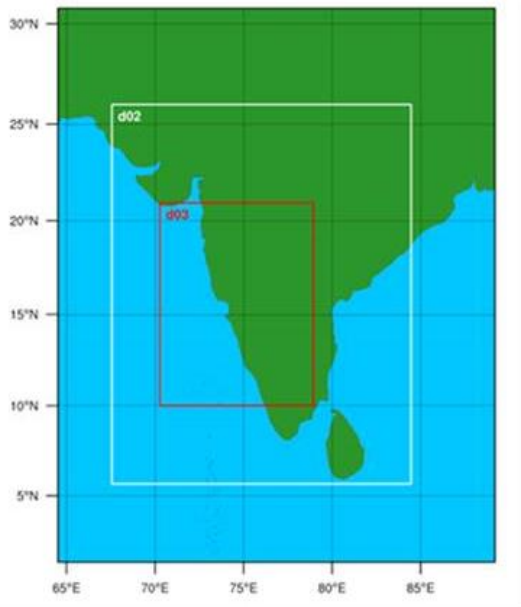


Figure 6.3 Configuration of the WRF model domains

To simulate the urbanization impacts on rainfall, we couple the mesoscale model WRF with single layer urban canopy model (UCM) (Kusaka et. al., 2001). The model is initialized at 00Z 4th July 2006 UTC allowing for 24 hr spin up with a horizontal resolution 4 km at the Brief descriptions of physical parameterizations used for simulation is presented in Table 6.2.

Table 6-2 Model configuration and setup

Physics	**Setup**
Cloud Micro physics	Thompson scheme (Thompson et. al., 2008)
Sub grid scale cloud	Kain-Fritsch (new Eta) scheme (Kain, John S., 2004)
Boundary-layer	Mellor-Yamada-Janjic TKE scheme (Janjic and Zavisa I., 1994,Janjic 1994, 2002)
Long wave radiation	RRTM scheme (Mlawer et al.,1997)
Short wave radiation	Dudhia scheme(Dudhia, J., 1989)
Surface-layer	Monin-Obukhov -Janjic scheme (Monin A. S., and A. M. Obukhov, 1954, Janjic, Z. I., 1994)
Land-surface	Unified Noah land-surface model(Tewari et al. 2004)
Urban Parameterization	Single-layer, UCM (Kusaka et al. 2001)

6.4.2.1 Selection of simulation time period

We perform the dynamic experiment with WRF for 5 years of monsoon season to understand the effect of urban expansions over the extreme precipitation process. The selection of simulation time period is carried out to obtain a representative condition of interannual monsoonal variations. The 5 years are selected on the following basis.

Table 6-3 Selection basis for the rainfall years for model run

Sr. No.	Criteria	Year
1	The year with occurrence of highest no of extremes	1983
2	The year with occurrence of lowest no of ext	2002
3	The year with mean rainfall closest to time series mean	2000
4	The year with lowest mean rainfall	1986
5	The year with highest mean rainfall	1990

We perform the WRF simulation for three different cases, pre urban (with land use 1973), current urban (with land use 2001), projected urban (with projected land use of 2050). The differences in results between them quantify the impacts of urbanization development.

6.5 Integration of Statistical and Dynamic Downscaling

We evaluate the quantile values of the downscaled projections with both statistically and dynamical methods as well as the observed rainfall. As the dynamic downscaling considers the physical process corresponding to occurrence of rainfall it is expected to have simulated values closer to the observed values, especially at the higher quantiles. We perform the quantile transformation of statistically downscaled projections on the basis of dynamically downscaled projections obtained different quantiles. Firstly a base transform is performed with the dynamic downscaling values obtained over WRF-LSM run considering the 1973 land use pattern over the city of Mumbai. This base transform is essentially required as the statistically downscaled projection highly underestimates the extreme events. The subsequent integrated projections for historical and future time period is obtained considering WRF-LSM run for 2001 and 2050 land use pattern respectively. This methodology is similar to the bias removal with quantile value mapping proposed by Li at. al. 2010.

6.6 Results

The urban rainfall projection for historical as well as future scenarios is achieved through the statistical downscaling technique. The multi model average is calculated to remove the intra model uncertainty in the rainfall projections (Salvi et al. 2013). The 5 years rainfall projections are obtained with dynamic downscaling considering pre, current and future urbanization over the city of Mumbai. The comparison of the mean and standard deviation and extreme for observed data and projected data for historical time period (1969-2005) is made to ascertain the performance of downscaling procedures. The Figure 6.4 indicates the performance of multi model averages of statistical and dynamic downscaling and model in reference with that of observed rainfall for the period 1979 -2005.

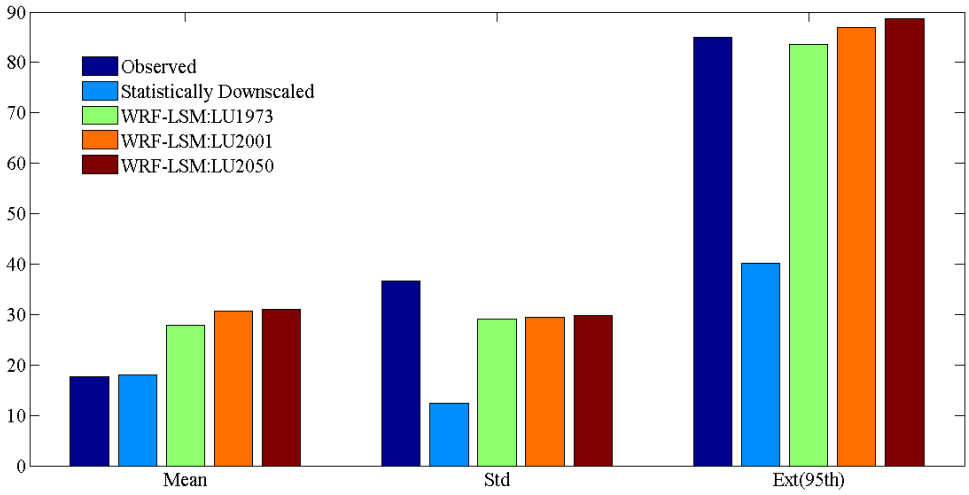


Figure 6.4 Performance of Statistical and Dynamic Downscaling methodology

The mean rainfall projections from GCM after downscaling show results equivalence with observed rainfall. This indicates that the rainfall projected from selected predictors is very well simulated with the statistical downscaling model under its mean conditions. However the standard deviation (Std) and extreme rainfall projections (95^{th} quantile) are not simulated at the same degree of accuracy. The dynamic downscaling at the same time over estimates the mean condition but at the same time simulates the standard deviation and extremes with a greater accuracy. We again observe am increase in the extreme precipitation amount with the amplified urbanization over the city of Mumbai provided within the model simulations as land use land cover maps.

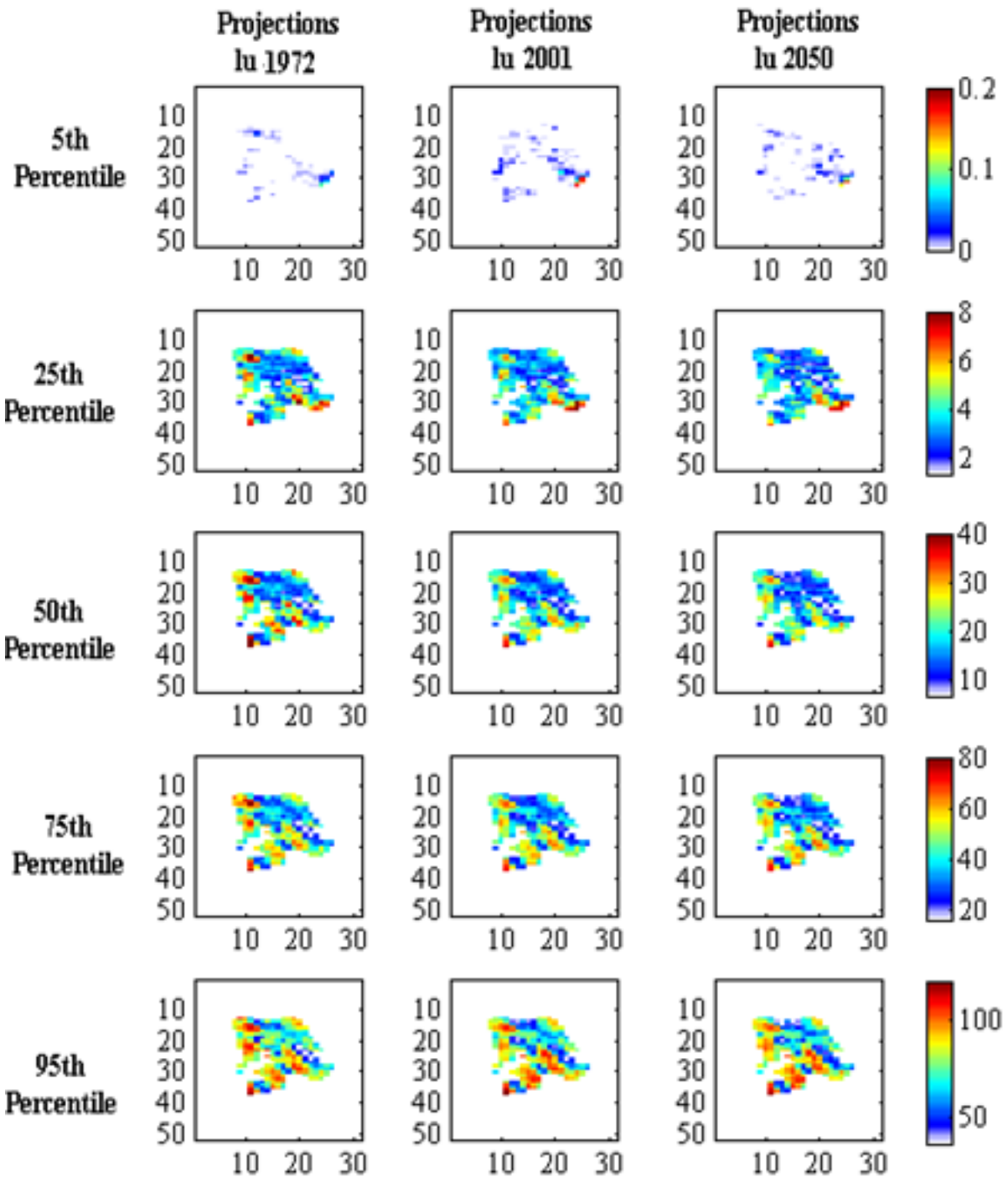


Figure 6.5 Spatial variability comparison between WRF-LSM

The outputs at different quantile levels considering the three stages of urban development of the city of Mumbai

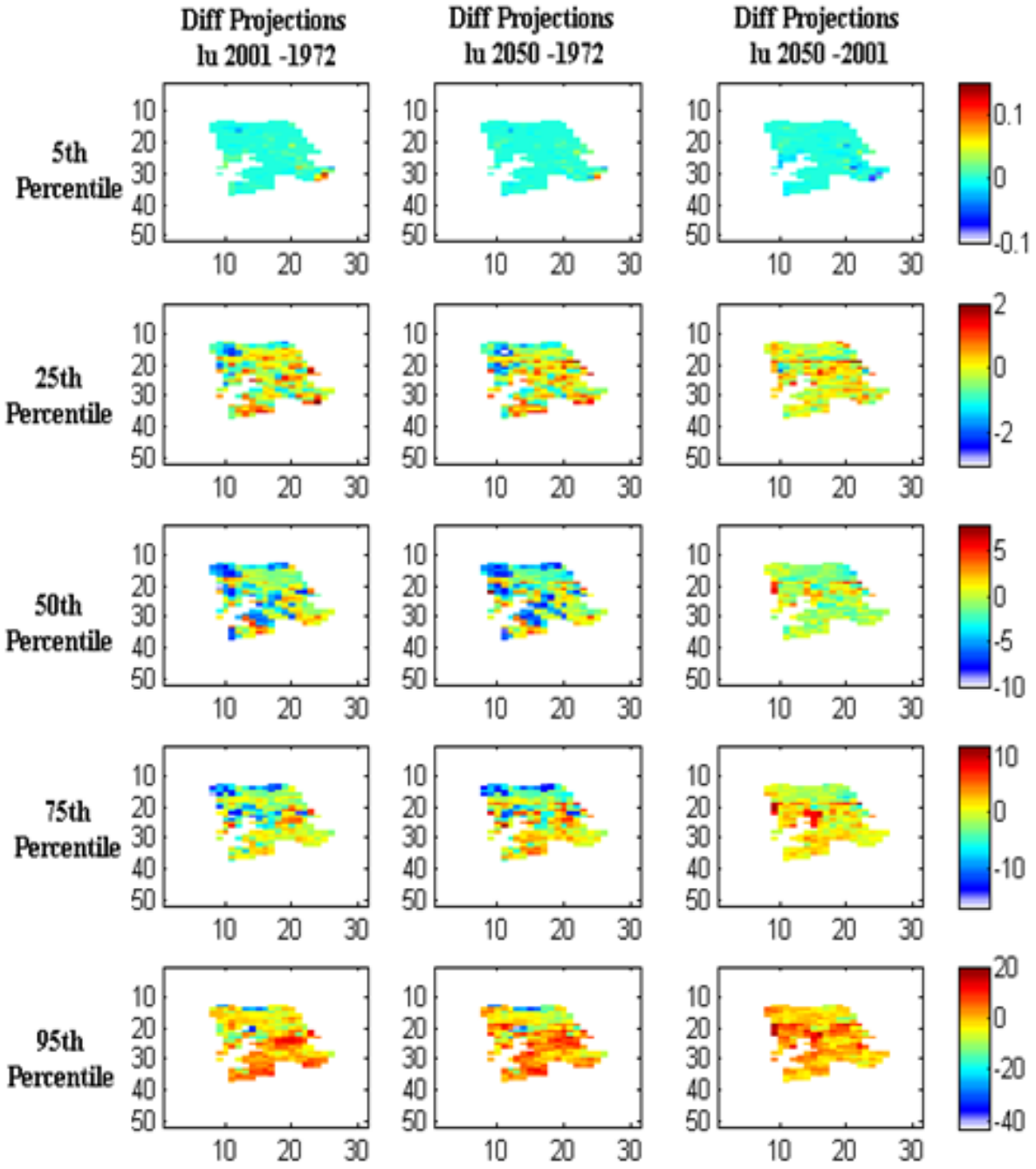


Figure 6.6 Difference of spatial variability comparison between WRF-LSM outputs

The difference between model simulated precipitation at different quantile levels considering the three stages of urban development of the city of Mumbai

The Figure 6.5 shows the spatial pattern of simulated rainfall at 4 km resolution over the city of Mumbai. The difference with increased urbanization is observed at higher quantiles as indicated with figure 6.6. Considering the fact that the changes are observed mainly at the

higher quantiles the quantile integration are applied at the higher quantiles. The performance of integration methodology is as shown with figure 6.7.

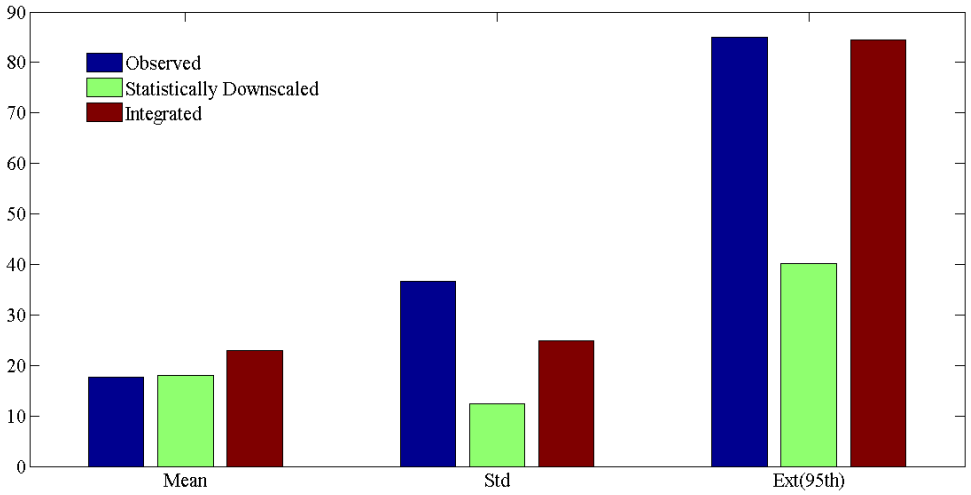


Figure 6.7 Performance validations of integrated model projections

The mean standard deviation and extremes of statistical downscaled and integrated projections are compared with the observed rainfall over the same time period

The projections of future rainfall for the period 2030 -2050 are carried out for both RCP 4.5 and RCP 8.5 scenario. The obtained statistical and integrated projections are as shown with figure 6.8.

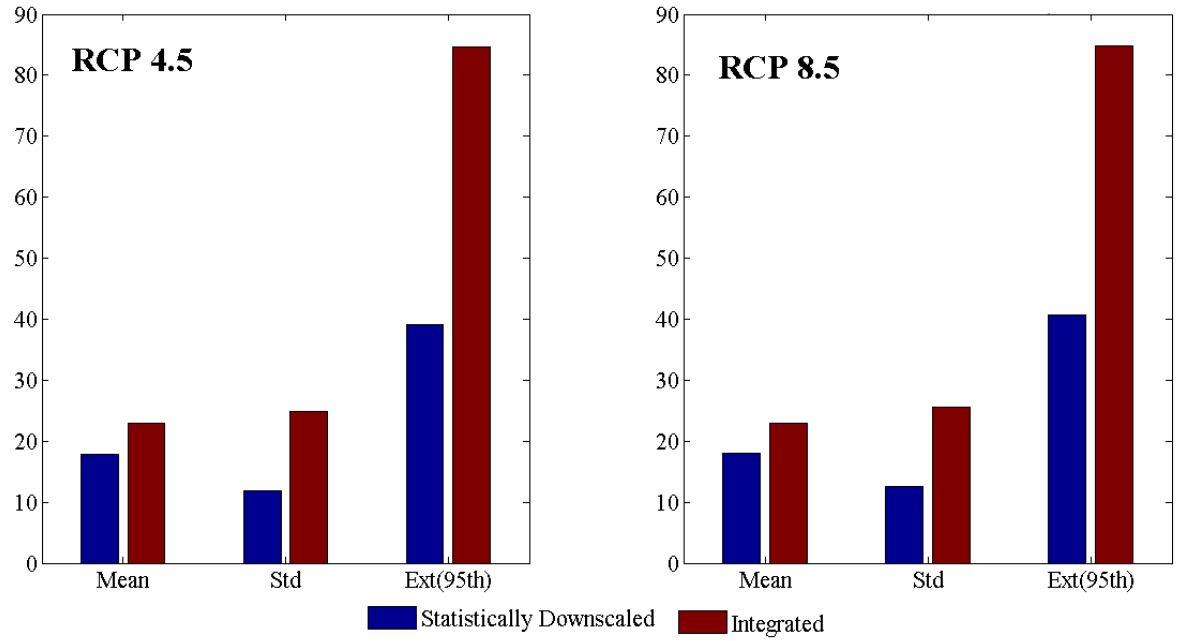


Figure 6.8 Integrated model projections for the future time period

6.7 Summary

This analysis is undertaken to estimate the effect of urban expansion of the city of Mumbai over last 30 years and further up to 2050 over the rainfall extremes. The urban expansion of the city of Mumbai over last 30 years and further up to 2050 is provided to the WRF-LSM. The projections obtained with limited runs of dynamic downscaled output are integrated with the longer time period statistically downscaling outputs. The assessment of proposed methodology over the historic time period reveals an improvement of the projected rainfall in terms of standard deviation and extremes. We obtain the integrated projections for the future time period considering two different emission scenarios. Here it is important to note that the projected rainfall increases with increased urbanization.

Climate projections are very important to understand the behaviour of different atmospheric processed under the changing global climate. Precipitation downscaling improves the coarse resolution and poor representation of precipitation in global climate models, and helps end users to assess the likely hydrological impacts of climate change. The statistical downscaling remains relatively non expensive but does not take care of the physical processes. The major limitation of dynamic downscaling is the requirement of very high computational efficiency. The proposed novel approach integrating the outputs from the statistical and dynamic downscaling methodology is useful for large number of urban regions of the country especially of extreme precipitation projection.

Chapter 7 : SUMMARY AND FUTURE SCOPE OF STUDY

7.1 Research contribution from the present study

Overall, this study highlights the important role of land use /land cover and urbanization for understanding the rainfall changes as part of regional climate change. Although only a small percentage of land covers urban areas significantly alter the climate, biogeochemistry, and hydrology at local and regional scales.

The work presented here firstly includes literature review on the impact of urbanization on rainfall and identifies a research gap in terms of a theoretical and numerical study to highlight role that the UHI plays in precipitation enhancement for Indian summer monsoon rainfall. The application of remote sensing data and is proved to be an objective and effective method for understanding spatial characteristics and dynamic changes of urban thermal landscape and providing decision-making reference for urban planning and management. We observe the daytime SUHI characteristics mainly governed by the natural factors whereas the nighttime SUHI is driven by the anthropogenic factors. The research highlights the higher occurrences of extreme temperature in nearby non-urban regions than over the urban region as opposite to the reported effect of UHI in different parts of the world. The analysis of gridded precipitation data is undertaken to demonstrate an overall picture of the impacts of urbanization on Indian summer monsoon rainfall extremes. The key finding lies in understanding the non–uniformity of these impacts on extreme rainfall across the country. The presented analysis results of extremes with station data of Mumbai and Alibaug indicate that the summer monthly rainfall amount shows an increasing trend under the influence of urbanization changes and the UHI effect. These results highlight the need of an effective flood forecasting system to safeguard the large population of the metropolitan Mumbai against the occurrences of extreme rainfall events. The statistical quantile forecast models relationship between the large scale circulation and rainfall extremes to provide a short range (1-3days) extreme rainfall forecast. The proposed forecasting model provides a remarkable improvement over the coarser scale NWF output to provide an efficient extreme rainfall forecast over the city of Mumbai. However the false detection of extreme around 3-5 days of its actual occurrence remains a challenge. The extreme rainfall projections obtained with dynamic downscaling shows an increase with the expansion of urban landuse over the city of Mumbai. However, this methodology is computationally expensive hence difficult to apply for the longer time period and several locations. The

integration of dynamic downscaling outputs with the projections obtained with statistical downscaling provides a remarkable improvement to the extreme rainfall projections. Further this methodology is computationally inexpensive and hence can be easily applied for a longer time periods and many different urban locations.

In the same context more case studies are needed for the Indian subcontinent to verify the effects of cities on climate and particularly the extreme rainfall events. The work has implications for policy makers, urban planners, water resource managers, and agriculture professionals who may use an understanding of urban rainfall climate in the design of better drainage systems, planning of land use, or identification of optimal areas for agricultural activity.

In the above mentioned context, the analysis contributes significantly to the literature by having demonstrated the effect of urban growth over the country to the development of UHI and its effect over the heavy rainfall climatology covering the short range forecast as well as long range projection.

7.2 Future scope of work

Following research activities are planned as next phase the of current work;

- A complete understanding of the SUHII behaviour under the changing emission scenario may be correctly understood utilizing the emissions in chemical transport models. For example to calculate columnar concentrations of BC, and subsequent radiative effects, which could influence the surface radiation balance and SUHII therefore.
- The proposed quantile regression model revealing the urbanization effects on rainfall extremes based on the relationship between the large scale circulation and rainfall extremes. The calibration of rainfall model with past data may not guarantee good simulations for future considering the urban developments. Further to this, the quantile regression is a black box model and hence it cannot identify the physical mechanisms, which are affected by urbanization and are responsible for changing the patterns of rainfall extremes. This needs to carry out a follow on research activities in understanding the fingerprints of extremes (Nayak and Ghosh, 2012) and then identifying the changes in the fingerprints at local scale during extremes for pre and post urbanization periods.

- The satellite based cloud detection is an important source of information for short range rainfall forecasting. The role of integrating the satellite cloud data along with climate data in a forecasting model may be checked to further improve the skill of proposed forecasting model in order to reduce the uncertainty associated with it and obtain a more realistic forecast. Development of computationally inexpensive forecasting models based on statistical methodology is of a high importance for the large number of fast developing urban centers of the country.
- A realistic rainfall projection is valuable information for urban water resources development and management. Considering the spatial non uniformity in trends of different rainfall parameters with different climate zones and urban centers of the country. A specialized effort is required to obtain a correct urban rainfall projection for different urban centers of the country. This may include considering the dynamical aspect of land use change with a coupled statistical and dynamical modeling as demonstrated here.

REFERENCES

Allan, R. P., & Soden, B. J. (2008). Atmospheric warming and the amplification of precipitation extremes. *Science*, *321*(5895), 1481-1484.

Altava-Ortiz, V., Barrera, A., Llasat, M., Prat, O. P., Gibergans-Báguena, J., & Barnolas, M. (2006). Application of the MM5 and the analogous method to heavy rainfall event, the case of 16? 18 October 2003 in Catalonia (NE Spain). *Advances in Geosciences*, *7*, 313-319.

Anas, A., Arnott, R., & Small, K. A. (1998). Urban spatial structure. *Journal of economic literature*, *36*(3), 1426-1464.

Arnfield, A. J. (2003). Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *International journal of climatology*, *23*(1), 1-26.

Baik, J. J., Park, R. S., Chun, H. Y., & Kim, J. J. (2000). A laboratory model of urban street-canyon flows. *Journal of applied meteorology*, *39*(9), 1592-1600.

Barnes, S. L. (1964). A technique for maximizing details in numerical weather map analysis. *Journal of Applied Meteorology*, *3*(4), 396-409.

Barredo, J. I., Kasanko, M., McCormick, N., & Lavalle, C. (2003). Modelling dynamic spatial processes: simulation of urban future scenarios through cellular automata. *Landscape and urban planning*, *64*(3), 145-160.

Baur, D., Saisana, M., & Schulze, N. (2004). Modelling the effects of meteorological variables on ozone concentration—a quantile regression approach. *Atmospheric Environment*, *38*(28), 4689-4699.

Ben Daoud, A., Sauquet, E., Lang, M., Bontron, G., & Obled, C. (2011). Precipitation forecasting through an analog sorting technique: a comparative study. *Advances in Geosciences*, *29*(29), 103-107.

Bornstein, R., & Lin, Q. (2000). Urban heat islands and summertime convective thunderstorms in Atlanta: three case studies. *Atmospheric Environment*, *34*(3), 507-516.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152). ACM.

Boswell D (2002) Introduction to support vector machines. Available at http://www.work.caltech.edu/~boswell/IntroToSVM.pdf. 15 pp

Braham, R. R., Semonin, R. G., Auer, A. H., Changnon Jr, S. A., & Hales, J. M. (1981). Summary of urban effects on clouds and rain. *METROMEX: a review and summary, Meteor. Monogr*, (40), 141-152.

Bremnes, J. B. (2004a). Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. *Monthly Weather Review*, *132*(1), 338-347.

Bremnes, J. B. (2004b). Probabilistic wind power forecasts using local quantile regression. *Wind Energy*, *7*(1), 47-54.

Brezková, L., Šálek, M., Soukalová, E., & Starý, M. (2007). Predictability of flood events in view of current meteorology and hydrology in the conditions of the Czech Republic. *Soil and Water Research*, *2*(4), 156-168.

Cannon, A. J. (2011). Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Computers & geosciences*, *37*(9), 1277-1284.

Cavazos, T., Turrent, C., & Lettenmaier, D. P. (2008). Extreme precipitation trends associated with tropical cyclones in the core of the North American monsoon. *Geophysical Research Letters*, *35*(21).

Chakrabarti, P. D. (2001). Urban crisis in India: new initiatives for sustainable cities. *Development in practice*, *11*(2-3), 260-272.

Chander, G., & Markham, B. (2003). Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. *Geoscience and Remote Sensing, IEEE Transactions on*, *41*(11), 2674-2677.

Changnon, S. A. (1992). Inadvertent weather modification in urban areas: Lessons for global climate change. *Bulletin of the American Meteorological Society*, *73*(5), 619-627.

Changnon, S. A., Shealy, R. T., & Scott, R. W. (1991). Precipitation changes in fall, winter, and spring caused by St. Louis. *Journal of applied meteorology*, *30*(1), 126-134.

Changnon, S. A., Huff, F. A., Schickedanz, P. T., & Vogel, J. L. (1977). Summary of METROMEX, Volume 1: Weather anomalies and impacts.

Chernozhukov, V., & Hong, H. (2011). Three-step censored quantile regression and extramarital affairs. *Journal of the American Statistical Association*.

Climate Research Committee. (2005). *Radiative Forcing of Climate Change:: Expanding the Concept and Addressing Uncertainties*. National Academies Press.

Conway, D., Wilby, R. L., & Jones, P. D. (1996). Precipitation and air flow indices over the British Isle. *Climate Research*, *7*(2), 169-183.

Cortes, C., & Vapnik, V. (1995). Machine learning. *Support-vector networks*, *journal*, *20*, 273-297.

Crane, R. G., & Hewitson, B. C. (1998). Doubled CO2 precipitation changes for the Susquehanna basin: down-scaling from the GENESIS general circulation model. *International Journal of Climatology*, *18*(1), 65-76.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Dash, P., Göttsche, F. M., Olesen, F. S., & Fischer, H. (2002). Land surface temperature and emissivity estimation from passive sensor data: theory and practice-current trends. *International Journal of remote sensing*, *23*(13), 2563-2594.

Dash, S. K., Saraswat, V., Panda, S. K., & Sharma, N. (2013). A study of changes in rainfall and temperature patterns at four cities and corresponding meteorological subdivisions over coastal regions of India. *Global and Planetary Change*, *108*, 175-194.

Déqué, M., Rowell, D. P., Lüthi, D., Giorgi, F., Christensen, J. H., Rockel, B., ... & van den Hurk, B. J. J. M. (2007). An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections.*Climatic Change*, *81*(1), 53-70.

Dibike, Y. B., & Coulibaly, P. (2005). Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models. *Journal of hydrology*, *307*(1), 145-163.

Dirmeyer, P. A., Niyogi, D., de Noblet-Ducoudré, N., Dickinson, R. E., & Snyder, P. K. (2010). Impacts of land use change on climate. *International Journal of Climatology*, *30*(13), 1905-1907.

Dodla, V. B. R., & Ratna, S. B. (2010). Mesoscale characteristics and prediction of an unusual extreme heavy precipitation event over India using a high resolution mesoscale model. *Atmospheric Research*, *95*(2), 255-269.

Dudhia, J. (1989). Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, *46*(20), 3077-3107.

Ehlers, M., Jadkowski, M. A., Howard, R. R., & Brostuen, D. E. (1990). Application of SPOT data for regional growth analysis and local planning.*Photogrammetric Engineering and Remote Sensing*, *56*(2), 175-180.

Fall, S., Niyogi, D., Gluhovsky, A., Pielke, R. A., Kalnay, E., & Rochon, G. (2010). Impacts of land use land cover on temperature trends over the continental United States: assessment using the North American Regional Reanalysis. *International Journal of Climatology*, *30*(13), 1980-1993.

Fahrmeir, L., & Tutz, G. (2013). *Multivariate statistical modelling based on generalized linear models*. Springer Science & Business Media.

Fasullo, J., & Webster, P. J. (2003). A hydrological definition of Indian monsoon onset and withdrawal. *Journal of Climate*, *16*(19), 3200-3211.

Field, C. B. (Ed.). (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*. Cambridge University Press.

Flint, E. P., & Richards, J. F. (1991). Historical analysis of changes in land use and carbon stock of vegetation in south and southeast Asia. *Canadian Journal of Forest Research*, *21*(1), 91-110.

Flohn, H. (1968). Contributions to a meteorology of the Tibetan Highlands.*Atmospheric science paper; no. 130.*

Forster, B. C. (1985). An examination of some problems and solutions in monitoring urban areas from satellite platforms. *International Journal of Remote Sensing*, *6*(1), 139-151.

Francis, P. A., & Gadgil, S. (2006). Intense rainfall events over the west coast of India. *Meteorology and Atmospheric Physics*, *94*(1-4), 27-42.

Frey, C. M., Parlow, E., Vogt, R., Harhash, M., & Abdel Wahab, M. M. (2011). Flux measurements in Cairo. Part 1: in situ measurements and their applicability for comparison with satellite data. *International Journal of Climatology*, *31*(2), 218-231.

Friederichs, P., & Hense, A. (2007). Statistical downscaling of extreme precipitation events using censored quantile regression. *Monthly weather review*, *135*(6), 2365-2378.

Gallo, K. P., Tarpley, J. D., McNab, A. L., & Karl, T. R. (1995). Assessment of urban heat islands: a satellite perspective. *Atmospheric Research*, *37*(1), 37-43.

Gamba, P., & Herold, M. (Eds.). (2009). *Global mapping of human settlement: experiences, datasets, and prospects*. CRC Press.

Gero, A. F., & Pitman, A. J. (2006). The impact of land cover change on a simulated storm event in the Sydney basin. *Journal of Applied Meteorology and Climatology*, *45*(2), 283-300.

Ghosh, S., Das, D., Kao, S. C., & Ganguly, A. R. (2012). Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes.*Nature Climate Change*, *2*(2), 86-91.

Ghosh, S., & Misra, C. (2010). Assessing hydrological impacts of climate change: modeling techniques and challenges. *The Open Hydrology Journal*,*4*(1).

Ghosh S, Mujumdar PP. Statistical downscaling of GCM simulations to stream flow using relevance vector machine. Adv Water Res 2008; 31(1): pp 132-146.

Gilchrist, W. (2000). *Statistical modelling with quantile functions*. CRC Press.

Goswami, B. N., Venugopal, V., Sengupta, D., Madhusoodanan, M. S., & Xavier, P. K. (2006). Increasing trend of extreme rain events over India in a warming environment. *Science*, *314*(5804), 1442-1445.

Groisman, P. Y., Knight, R. W., Easterling, D. R., Karl, T. R., Hegerl, G. C., & Razuvaev, V. N. (2005). Trends in intense precipitation in the climate record. *Journal of climate*, *18*(9), 1326-1350.

Gupta, M. D., Das, S., & Ashrit, R. (2004). MM5 3D-Var data assimilation and forecast system over Indian Subcontinent—Results from recent experiments. In *5th WRF/14th MM5 Users' Workshop NCAR*.

Gutowski Jr, W. J., Arritt, R. W., Kawazoe, S., Flory, D. M., Takle, E. S., Biner, S., ... &

Mearns, L. O. (2010). Regional extreme monthly precipitation simulated by NARCCAP RCMs. *Journal of Hydrometeorology*, *11*(6), 1373-1379.

Hafner, J., & Kidder, S. Q. (1999). Urban heat island modeling in conjunction with satellite-derived surface/soil parameters. *Journal of applied meteorology*, *38*(4), 448-465.

Hamill, T. M., Whitaker, J. S., Fiorino, M., & Benjamin, S. G. (2011). Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Monthly Weather Review*, *139*(2), 668-688.

Han-qiu, X. U. (2005). A Study on Information Extraction of Water Body with the Modified Normalized Difference Water Index (MNDWI)[J]. *Journal of Remote Sensing*, *5*, 589-595.

Handl, A., (2000), Quantile,http://www.wiwi.uni-bielefeld.de/~frohn /Lehre /Datenanalyse /Skript/daquantile.pdf

Helsel, D. R., & Hirsch, R. M. (1992). Statistical Methods in Water ResourcesElsevier. *New York*.

Hjelmfelt, M. R. (1982). Numerical simulation of the effects of St. Louis on mesoscale boundary-layer airflow and vertical air motion: Simulations of urban vs non-urban effects. *Journal of Applied Meteorology*, *21*(9), 1239-1257.

Hong, S. Y., & Lee, J. W. (2009). Assessment of the WRF model in reproducing a flash-flood heavy rainfall event over Korea. *Atmospheric Research*, *93*(4), 818-831.

http://esa.un.org/unup.

http://www.actionbioscience.org

http://www.censusindia.gov.in/2011provresults/data_files/india/Final_PPT_2011_chapter 3. pdf.

http://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs

http://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-ensemble-forecast-system-gefs

Huff, F. A. (1986). Urban hydrological review. *Bulletin of the American Meteorological Society*, 67: pp 703–712.

Inamura, T., Izumi, T., & Matsuyama, H. (2011). Diagnostic study of the effects of a large city on heavy rainfall as revealed by an ensemble simulation: A case study of central Tokyo, Japan. *Journal of Applied Meteorology and Climatology*, *50*(3), 713-728.

Janjic, Z. I. (1994). The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly Weather Review*, *122*(5), 927-945.

Janjić, Z. I. (2002). Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso model. *NCEP office note*, *437*, 61.

Janjic, Z. I. (1994). The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly Weather Review*, *122*(5), 927-945.

Jauregui, E., & Romales, E. (1996). Urban effects on convective precipitation in Mexico City. *Atmospheric Environment*, *30*(20), 3383-3389.

Jensen, J. R., & Cowen, D. C. (1999). Remote sensing of urban/suburban infrastructure and socio-economic attributes. *Photogrammetric engineering and remote sensing*, *65*, 611-622.

Jones, R. G., Murphy, J. M., & Noguer, M. (1995). Simulation of climate change over europe using a nested regional-climate model. I: Assessment of control climate, including sensitivity to location of lateral boundaries.*Quarterly Journal of the Royal Meteorological Society*, *121*(526), 1413-1449.

Joseph PV (2006) Role of low level jetstream in intense monsoon rainfall episodes over the west coast of India. In: *National Workshop on Arabian Sea Monsoon Experiment* (ARMEX).

Kain, J. S. (2004). The Kain-Fritsch convective parameterization: an update.*Journal of Applied Meteorology*, *43*(1), 170-181.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., ... & Zhu, Y. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society*, *77*(3), 437-471.

Kalnay, E., & Cai, M. (2003). Impact of urbanization and land-use change on climate. *Nature*, *423*(6939), 528-531.

Karl, T. R., Diaz, H. F., & Kukla, G. (1988). Urbanization: Its detection and effect in the United States climate record. *Journal of Climate*, *1*(11), 1099-1123.

Katz, R. W., & Parlange, M. B. (1996). Mixtures of stochastic processes: application to statistical downscaling. *Climate Research*, *7*(EFLUM-ARTICLE-1996-001), 185-193.

Kendon, E. J., Jones, R. G., Kjellström, E., & Murphy, J. M. (2010). Using and designing GCM-RCM ensemble regional climate projections. *Journal of Climate*, *23*(24), 6485-6503.

Khaladkar, R. M., Narkhedkar, S. G., & Mahajan, P. N. (2007). *Performance of NCMRWF Models in Predicting High Rainfall Spells During SW Monsoon Season: A Study for Some Cases in July 2004*. Indian Institute of Tropical Meteorology.

Kharin, V. V., & Zwiers, F. W. (2005). Estimating extremes in transient climate change simulations. *Journal of Climate*, *18*(8), 1156-1173.

Kharin, V. V., Zwiers, F. W., Zhang, X., & Wehner, M. (2013). Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, *119*(2), 345-357.

Kishtawal, C. M., Niyogi, D., Tewari, M., Pielke, R. A., & Shepherd, J. M. (2010). Urbanization signature in the observed heavy rainfall climatology over India. *International Journal of Climatology*, *30*(13), 1908-1916.

Koch, S. E., Desjardins, M., & Kocin, P. J. (1983). An interactive Barnes objective map analysis scheme for use with satellite and conventional data. *Journal of Climate and Applied Meteorology*, *22*(9), 1487-1503.

Koenker, R. (2005). *Quantile regression* (No. 38). Cambridge university press.

Koenker, R., & Bassett, G. (1978). Regression quantiles. Econometrica 46 33–50. *Mathematical Reviews (MathSciNet): MR474644 Digital Object Identifier: doi*, *10*, 1913643.

Koenker, R., & Schorfheide, F. (1994). Quantile spline models for global temperature change. *Climatic Change*, *28*(4), 395-404.

Koenker, R., and Hallock, K. F. (2001).Quantile Regression, *Journal of Economic Perspectives*, 15(4), pp143–156.

Konwar, M., Parekh, A., & Goswami, B. N. (2012). Dynamics of east-west asymmetry of Indian summer monsoon rainfall trends in recent decades.*Geophysical Research Letters*, *39*(10).

Kusaka, H., & Kimura, F. (2004). Coupling a single-layer urban canopy model with a simple atmospheric model: Impact on urban heat island simulation for an idealized case. *気象集誌. 第 2 輯*, *82*(1), 67-80.

Lee, S. 2005: Lecture Notes for MECT1 Quantile Regression, http://www.homepages. ucl. ac.uk/~uctplso/Teaching/MECT/lecture8.pdf

Lei, M., Niyogi, D., Kishtawal, C., Pielke Sr, R. A., Beltrán-Przekurat, A., Nobis, T. E., & Vaidya, S. S. (2008). Effect of explicit urban land surface representation on the simulation

of the 26 July 2005 heavy rain event over Mumbai, India. *Atmospheric Chemistry and Physics*, *8*(20), 5975-5995.

Lettenmaier, D. (1995). Stochastic modeling of precipitation with applications to climate model downscaling. In *Analysis of Climate Variability* (pp. 197-212). Springer Berlin Heidelberg.

Lin, C. Y., Chen, W. C., Chang, P. L., & Sheng, Y. F. (2011). Impact of the urban heat island effect on precipitation over a complex geographic environment in northern Taiwan. *Journal of Applied Meteorology and Climatology*, *50*(2), 339-353.

Lorenz, E. N. (1969). Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric sciences*, *26*(4), 636-646.

Masek, J. G., Lindsay, F. E., & Goward, S. N. (2000). Dynamics of urban growth in the Washington DC metropolitan area, 1973-1996, from Landsat observations. *International Journal of Remote Sensing*, *21*(18), 3473-3486.

Meehl, G. A., Covey, C., Taylor, K. E., Delworth, T., Stouffer, R. J., Latif, M., ... & Mitchell, J. F. (2007). The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bulletin of the American Meteorological Society*,*88*(9), 1383-1394.

Memon, R. A., Leung, D. Y., & Liu, C. H. (2009). An investigation of urban heat island intensity (UHII) as an indicator of urban heating. *Atmospheric Research*, *94*(3), 491-500.

Meng, W., Yan, J., & Hu, H. (2007). Urban effects and summer thunderstorms in a tropical cyclone affected situation over Guangzhou city.*Science in China Series D: Earth Sciences*, *50*(12), 1867-1876.

Min, S. K., Zhang, X., Zwiers, F. W., & Hegerl, G. C. (2011). Human contribution to more-intense precipitation extremes. *Nature*, *470*(7334), 378-381.

Mishra, V., & Lettenmaier, D. P. (2011). Climatic trends in major US urban areas, 1950–2009. *Geophysical Research Letters*, *38*(16).

Mitra, A. K., Iyengar, G. R., Durai, V. R., Sanjay, J., Krishnamurti, T. N., Mishra, A., & Sikka, D. R. (2011). Experimental real-time multi-model ensemble (MME) prediction of rainfall during monsoon 2008: large-scale medium-range aspects. *Journal of earth system science*, *120*(1), 27-52.

Mladjic, B., Sushama, L., Khaliq, M. N., Laprise, R., Caya, D., & Roy, R. (2011). Canadian RCM projected changes to extreme precipitation characteristics over Canada. *Journal of Climate*, *24*(10), 2565-2584.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, *102*(D14), 16663-16682.

Mondal, A., & Mujumdar, P. P. (2015). Modeling non-stationarity in intensity, duration and frequency of extreme rainfall over India. *Journal of Hydrology*,*521*, 217-231.

Monin, A. S., & Obukhov, A. (1954). Basic laws of turbulent mixing in the surface layer of the atmosphere. *Contrib. Geophys. Inst. Acad. Sci. USSR*,*151*, 163-187.

Mukul Tewari, N. C. A. R. (2004, January). Implementation and verification of the unified Noah land surface model in the WRF model (Formerly Paper Number 17.5). In *20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction*.

Narkhedkar, S. G., Sinha, S. K., & Mitra, A. K. (2008). Mesoscale objective analysis of daily rainfall with satellite and conventional data over Indian summer monsoon region. *Geofizika*, *25*(2), 159-178.

Nichol, J. E. (1996). High-resolution surface temperature patterns related to urban morphology in a tropical city: a satellite-based study. *Journal of Applied Meteorology*, *35*(1), 135-146.

Noble, W. S. (2006). What is a support vector machine?. *Nature biotechnology*, *24*(12), 1565-1567.

Nott, D. J., Dunsmuir, W. T. M., Kohn, R., & Woodcock, F. (2001). Statistical correction of a deterministic numerical weather prediction model. *Journal of the American Statistical Association*, *96*(455), 794-804.

Novak, D. R., Brill, K., Eckert, M., Oravec, R., Sullivan, B., Bann, R., ... & Bodner, M. (2011). Quantifying extreme rainfall threats at the Hydrometeorological Prediction Center. In *91st American Meteorological Society Annual Meeting*.

O'Gorman, P. A. (2012). Sensitivity of tropical precipitation extremes to climate change. *Nature Geoscience*, *5*(10), 697-700.

Ochs III, H. T., & Semonin, R. G. (1979). Sensitivity of a cloud microphysical model to an urban environment. *Journal of Applied Meteorology*, *18*(9), 1118-1129.

Oke, T. R. (1988). The urban energy balance. *Progress in Physical geography*, *12*(4), 471-508.

Oke, T. R. (1995). The heat island of the urban boundary layer: characteristics, causes and effects. In *Wind climate in cities* (pp. 81-107). Springer Netherlands.

Olanrewaju, R. M. (2009). The Climate Effect of Urbanization in A City of Developing Country: The Case Study Of Ilorin, Kwara State, Nigeria. *Ethiopian Journal of Environmental Studies and Management*, *2*(2).

Panziera, L., & Germann, U. (2010). The relation between airflow and orographic precipitation on the southern side of the Alps as revealed by weather radar. *Quarterly Journal of the Royal Meteorological Society*,*136*(646), 222-238.

Pappenberger, F., Bartholmes, J., Thielen, J., Cloke, H. L., Buizza, R., & de Roo, A. (2008). New dimensions in early flood warning across the globe using grand-ensemble weather predictions. *Geophysical Research Letters*,*35*(10).

Parthasarathy, B., Rupa Kumar, K., & Munot, A. A. (1996). *Homogeneous regional summer monsoon rainfall over India: interannual variability and teleconnections*. Indian Institute of Tropical Meteorology.

Pielke, R. A., Marland, G., Betts, R. A., Chase, T. N., Eastman, J. L., Niles, J. O., & Running, S. W. (2002). The influence of land-use change and landscape dynamics on the climate system: relevance to climate-change policy beyond the radiative effect of greenhouse gases. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *360*(1797), 1705-1719.

Population reference bureau, transition in world's population, http://www.prb.org /publications/population bulletins/2004.

Potere, D., & Schneider, A. (2009). 13 Comparison of Global Urban Maps.*Global Mapping of Human Settlement: Experiences, Datasets, and Prospects*, 269.

Racsko, P., Szeidl, L., & Semenov, M. (1991). A serial approach to local stochastic weather models. *Ecological modelling*, *57*(1-2), 27-41.

Rajeevan, M., Bhate, J., Kale, J. D., & Lal, B. (2006). High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells. *Current Science*, *91*(3), 296-306.

Rakhecha, P. R., & Pisharoty, P. R. (1996). Heavy rainfall during monsoon season: Point and spatial distribution. *Current Science*, *71*(3), 179-186.

Rao, K. K., Patwardhan, S. K., Kulkarni, A., Kamala, K., Sabade, S. S., & Kumar, K. K. (2014). Projected changes in mean and extreme precipitation indices over India using PRECIS. *Global and Planetary change*, *113*, 77-90.

Rao, P. K. (1972). Remote sensing of urban heat islands from an environmental satellite. *Bulletin of the American meteorological society*,*53*(7), 647.Pielke Sr., R. A., M arland, G., Betts, R.

Ray, K. S., & Srivastava, A. K. (2000). Is there any change in extreme events like heavy rainfall?. *Current Science*, *79*(2), 155-158.

Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, *17*(1), 182-190.

Rigo, G., & Parlow, E. (2007). Modelling the ground heat flux of an urban area using remote sensing data. *Theoretical and Applied Climatology*, *90*(3-4), 185-199.

Root, B., Knight, P., Young, G., Greybush, S., Grumm, R., Holmes, R., & Ross, J. (2007). A fingerprinting technique for major weather events. *Journal of applied meteorology and climatology*, *46*(7), 1053-1066.

Roth, M., Oke, T. R., & Emery, W. J. (1989). Satellite-derived urban heat islands from three coastal cities and the utilization of such data in urban climatology. *International Journal of Remote Sensing*, *10*(11), 1699-1720.

Roy Bhowmik, S. K., & Durai, V. R. (2010). Application of multimodel ensemble techniques for real time district level rainfall forecasts in short range time scale over Indian region. *Meteorology and Atmospheric Physics*,*106*(1), 19-35.

Rummukainen, M. (1997). Methods for statistical downscaling of GCM simulations. *SMHI Rapporter. Meteorologi och Klimatologi (Sweden). no. 80.*

Sahai, A. K., Soman, M. K., & Satyan, V. (2000). All India summer monsoon rainfall prediction using an artificial neural network. *Climate dynamics*, *16*(4), 291-302.

Sailor, D. J. (2011). A review of methods for estimating anthropogenic heat and moisture emissions in the urban environment. *International Journal of Climatology*, *31*(2), 189-199.

Salathe Jr, E. P., Leung, L. R., Qian, Y., & Zhang, Y. (2010). Regional climate model projections for the State of Washington. *Climatic Change*,*102*(1-2), 51-75.

Sanderson, M., & Gorski, R. (1978). The effect of metropolitan Detroit-Windsor on precipitation. *Journal of Applied Meteorology*, *17*(4), 423-427.

Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Selvam, A. M. (2003). The dynamics of deterministic chaos in numerical weather prediction models. *arXiv preprint physics/0310034*.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, *63*(324), 1379-1389.

Shepherd, J. M., & Burian, S. J. (2003). Detection of urban-induced rainfall anomalies in a major coastal city. *Earth Interactions*, *7*(4), 1-17.

Sinha, S. K., Narkhedkar, S. G., & Mitra, A. K. (2006). Barnes objective analysis scheme of daily rainfall over Maharashtra (India) on a mesoscale grid. *Atmósfera*, *19*(2), 109-126.

Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for weather research and forecasting applications.*Journal of Computational Physics*, *227*(7), 3465-3485.

Small, C. (2003). High spatial resolution spectral mixture analysis of urban reflectance. *Remote Sensing of Environment*, *88*(1), 170-186.

Stathopoulou, M., & Cartalis, C. (2007). Daytime urban heat islands from Landsat ETM+ and Corine land cover data: An application to major cities in Greece. *Solar Energy*, *81*(3), 358-368.

Soille, P., & Pesaresi, M. (2002). Advances in mathematical morphology applied to geoscience and remote sensing. *Geoscience and Remote Sensing, IEEE Transactions on*, *40*(9), 2042-2055.

Solomon, S. (Ed.). (2007). *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC*(Vol. 4). Cambridge University Press.

Strikwerda, J. C. (2004). *Finite difference schemes and partial differential equations*. Siam.

Sudhira, H. S., Ramachandra, T. V., & Jagadish, K. S. (2004). Urban sprawl: metrics, dynamics and modelling using GIS. *International Journal of Applied Earth Observation and Geoinformation*, *5*(1), 29-39.

Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., & Dech, S. (2012). Monitoring urbanization in mega cities from space. *Remote sensing of Environment*, *117*, 162-176.

Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, *136*(12), 5095-5115.

Tran, Q. A., Zhang, Q. L., & Li, X. (2003, November). Reduce the number of support vectors by using clustering techniques. In *Machine Learning and Cybernetics, 2003 International Conference on* (Vol. 2, pp. 1245-1248). IEEE.

Treitz, P. M., Howarth, P. J., & Gong, P. (1992). Application of satellite and GIS technologies for land-cover and land-use mapping at the rural-urban fringe: a case study. *Photogrammetric engineering and remote sensing*,*58*(4), 439-448.

Trenberth, K. E., Smith, L., Qian, T., Dai, A., & Fasullo, J. (2007). Estimates of the global water budget and its annual cycle using observational and model data. *Journal of Hydrometeorology*, *8*(4), 758-769.

Trenberth, K. E. (2011). Changes in precipitation with climate change.*Climate Research*, *47*(1), 123.

Trenberth, K. (2005). Uncertainty in hurricanes and global warming. *Science,308*(5729), 1753-1754.

Trusilova, K., Jung, M., Churkina, G., Karstens, U., Heimann, M., & Claussen, M. (2008). Urbanization impacts on the climate in Europe: Numerical experiments by the PSU-NCAR Mesoscale Model (MM5). *Journal of Applied Meteorology and Climatology*, *47*(5), 1442-1455.

Tymvios, F., Savvidou, K., & Michaelides, S. C. (2010). Association of geopotential height patterns with heavy rainfall events in Cyprus. *Advances in Geosciences*, *23*, 73-78.

United Nations (2003). World urbanization prospects, The 2003 Revision, New York.

United Nations (2005). World Urbanization Prospects, The 2005 Revision, New York.

United Nations (2007). World Urbanization Prospects: The 2007 Revision.

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer Science & Business Media.

Vitousek, P. M. (1994). Beyond global warming: ecology and global change.*Ecology*, *75*(7), 1861-1876.

Vittal, H., Karmakar, S., & Ghosh, S. (2013). Diametric changes in trends and patterns of extreme rainfall over India from pre-1950 to post-1950.*Geophysical Research Letters*, *40*(12), 3253-3258.

von Storch, H. (1999). The global and regional climate system. In*Anthropogenic climate change* (pp. 3-36). Springer Berlin Heidelberg.

Voogt, J. A., & Oke, T. R. (2003). Thermal remote sensing of urban climates. *Remote sensing of environment*, *86*(3), 370-384.

Wan, Z., Zhang, Y., Zhang, Q., & Li, Z. L. (2004). Quality assessment and validation of the MODIS global land surface temperature. *International Journal of Remote Sensing*, *25*(1), 261-274.

Wang, J., Feng, J., Yan, Z., Hu, Y., & Jia, G. (2012). Nested high-resolution modeling of the impact of urbanization on regional climate in three vast urban agglomerations in China. *Journal of Geophysical Research: Atmospheres*,*117*(D21).

Weber, C., & Puissant, A. (2003). Urbanization pressure and modeling of urban growth: example of the Tunis Metropolitan Area. *Remote sensing of environment*, *86*(3), 341-352.

Webster, P. J., Holland, G. J., Curry, J. A., & Chang, H. R. (2005). Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science*, *309*(5742), 1844-1846.

Wei, M., & Toth, Z. (2003). A New Measure of Ensemble Performance; Perturbation versus Error Correlation Analysis (PECA). *Monthly weather review*, *131*(8), 1549-1565.

Welch, R., Remillard, M. M., & Slack, R. B. (1988). Remote sensing and geographic information system techniques for aquatic resource evaluation. *Photogrammetric engineering and remote sensing (USA)*.

WESTORELAND, S., & Stow, D. A. (1992). Category identification of changes land-use polygons in an integrated image processing geographic information system. *Photogrammetric Engineering and Remote Sensing*,*58*(11), 1593-1599.

Westra, S., Alexander, L. V., & Zwiers, F. W. (2013). Global increasing trends in annual maximum daily precipitation. *Journal of Climate*, *26*(11), 3904-3918.

Wheater, H. S. (2002). Progress in and prospects for fluvial flood modelling. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *360*(1796), 1409-1431.

Wheeler, S. G., Misra, P. N., & Holmes, Q. A. (1976). Linear dimensionality of Landsat agricultural data with implications for classification.

Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P., & Mearns, L. O. (2004). Guidelines for use of climate scenarios developed from statistical downscaling methods.

Wilby, R. L., & Wigley, T. M. L. (1997). Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography, 21*(4), 530-548.

Wilks, D. S., & Wilby, R. L. (1999). The weather generation game: a review of stochastic weather models. *Progress in Physical Geography, 23*(3), 329-357.

Wu, J., Jelinski, D. E., Luck, M., & Tueller, P. T. (2000). Multiscale analysis of landscape heterogeneity: scale variance and pattern metrics. *Geographic Information Sciences, 6*(1), 6-19.

Xian, G., & Crane, M. (2005). Assessments of urban growth in the Tampa Bay watershed using remote sensing data. *Remote Sensing of Environment,97*(2), 203-215.

Yatagai, A., Xie, P., & Alpert, P. (2008). Development of a daily gridded precipitation data set for the Middle East. *Advances in Geosciences, 12*, 165-170.

Young, P. C. (2002). Advances in real–time flood forecasting. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *360*(1796), 1433-1450.

Yuan, F., & Bauer, M. E. (2007). Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery. *Remote Sensing of Environment, 106*(3), 375-386.

Yuan, H., Van Der Wiele, C. F., & Khorram, S. (2009). An automated artificial neural network system for land use/land cover classification from Landsat TM imagery. *Remote Sensing, 1*(3), 243-265.

Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, *24*(3), 583-594.

Zhan, Y., & Shen, D. (2005). Design efficient support vector machine for fast classification. *Pattern Recognition*, *38*(1), 157-161.

Zhang, Q., Wang, J., Peng, X., Gong, P., & Shi, P. (2002). Urban built-up land change detection with road density and spectral information from multi-temporal Landsat TM data. *International Journal of Remote Sensing*, *23*(15), 3057-3078.