FUZZY CLUSTERING OF HYDROCLIMATIC REGIONS OF NORTHEAST INDIA AND CLIMATE CHANGE ANALYSIS

Submitted in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

by

JAYSHREE HAZARIKA

(Roll No: 136104016)

under the supervision of

Prof. ARUP KUMAR SARMA

DEPARTMENT OF CIVIL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI-781039, ASSAM, INDIA

December 2020

To my parents Sri. Puspa Hazarika and Smt. Hiranya Hazarika,

To my sister Smt. Shyamoli Hazarika

And to all my teachers

CERTIFICATE

This is to certify that the thesis entitled "**Fuzzy Clustering of Hydroclimatic Regions of Northeast India and Climate Change Analysis**", submitted by Jayshree Hazarika (136104016), a research scholar in the Department of Civil Engineering, Indian Institute of Technology Guwahati, for the award of the degree of **Doctor of Philosophy**, is a record of an original research work carried out by her under my supervision and guidance. The thesis has fulfilled all requirements as per the regulations of the institute and in my opinion has reached the standard needed for submission. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Date: 8th Dec 2020

Place: Guwahati

Prof. Arup Kumar Sarma

Dept. of Civil Engineering

Indian Institute of Technology Guwahati,

Guwahati - 781039, Assam, India.

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Jayshree Hazarika

Roll No.: 136104016

Date:

Place: Guwahati

ABSTRACT

In today's era, the most important issue that humanity has ever faced is considered to be climate change. In vast countries like India, where climate conditions vary from region to region, climate change is of major concern. The north-eastern part of India which receives heavier rainfall than other parts of the subcontinent suffers from high intensity rainfall of short duration and longer dry spells due to impact of climate change. It also affects river morphology due to flood or drought. Since water is one of the major sectors being affected by climate change, planning and management of water resources related issues should definitely include the impacts of climate change.

The primary goal of the study was to determine homogeneous hydroclimatic regions in northeast India and to analyze impacts of climate change on temperature and rainfall projections, with inclusion of orographic influence. At first, the region was delineated into homogeneous clusters using fuzzy clustering approach. A hybrid approach with GCM grid data (interpolated to station) along with rainfall data and location parameters (latitude, longitude and elevation) gave the best result among all other approaches. However, this may not be applicable for small scale regions. Hence, further sub-clustering was done with the use of rainfall data recorded by the tea gardens of Assam. The best result was exhibited by the approach, in which total annual rainfall, standard deviation of total annual rainfall and location parameters (latitude, longitude and elevation) were used as attributes. In the next step, five stations were selected to develop statistical downscaling models for predicting temperature and rainfall. Main objective was to understand the applicability of a specific downscaling model within a cluster. Three different GCM models were used for the study. It was observed that, although the same GCM variables were used for homogeneous clustering as well for downscaling, the correlation coefficients do not show any consistency within a cluster. This indicates that orography plays a very important role in rainfall variations at a local scale. However, in case of daily maximum and minimum temperature, local scale orographic arrangement is not required to be used in developing downscaling model for temperature. Therefore, a new downscaling approach was attempted with inclusion of an orographic factor, formulated by using various factors. The variable was then included in the downscaling models

and its efficiency was assessed. It was found that, for stations located on a hill the orographic factor could not improve the R^2 value. However, for stations at an identifiable distance from a hill, the orographic factor has exhibited increase in the R^2 value to a decent extent. Finally, uncertainty analysis was done and the future rainfall series were projected for respective model. The results exhibited increase in both rainfall and temperature in most of the cases for RCP8.5. In case of RCP2.6, much increment is not exhibited by both the models. The increase in projected annual rainfall is found to be more in case of new projections where orographic factor is incorporated. Overall, the projections for rainfall have higher uncertainty compared to temperature, which can be attributed to the complexity of rainfall process. It was found from the study that inclusion of orographic factor has been able to reduce the uncertainty of the downscaling process to a decent extent.

ACKNOWLEDGEMENT

I feel privileged in extending my earnest obligation, deep sense of gratitude and honour towards my supervisor, Prof. Arup Kumar Sarma, whose benevolent guidance, constant supervision, invaluable suggestions and generous help throughout the years have inspired me in successful completion of this research work. His wide knowledge in the field of water resources engineering has been of great benefit for me. It was my pleasure working under his supervision.

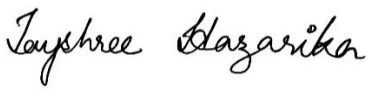
Sincere gratitude is extended towards all the members of my doctoral committee, Prof. Subashisa Dutta, Prof. Rajib Kumar Bhattcharjya and Prof. Bimlesh Kumar, for their co-operation at different stages of my research work. Their invaluable ideas, recommendations and constructive criticism have helped me in carrying out continuous improvement in my research.

I express sincere thanks to all the Civil Engineering faculties and other staff members for their suggestions and support to my work. I am extremely grateful to the department and central library for providing books and journals to successfully complete the thesis.

I would like to thank my friends L.N.V. Satish, Ashutosh Sharma and Shivam Gupta for their constant support throughout the course of my work, helping me to learn new and relevant concepts and for lending me a hand whenever I was in need. I will always be indebted to you, for the care and support that I have received. Thank you.

Completing this research journey would not have been possible for me without the presence of my friends. I would like to thank Vishal Deshpande sir, Mahesh Patel sir, Mayank Agarwal sir, Isha Vishan, Smrati Jain and Jaideep for always being there for me as mentors, as seniors, and as friends. Sincere thanks to Needhi Kotoky, Priyanka Talukdar, Poulomi Dey and Sayanti Ghosh for associating with me in a non-academic way and helping me in keeping my passion alive. It would not be complete without the mention of my friends Bandita Barman, Uddipana Dowerah, Nayantara Kotoky and Abhishek Kumar whose presence have made this journey quite easy for me. Thank you for always supporting me, encouraging me, helping me and believing in me.

Finally, I express my heartiest gratitude to my family for their support, encouragement and advices during the course of the project, which has always been a constant source of inspiration and strength for me. Also, I would like to acknowledge the kind cooperation of all those who helped me directly or indirectly in successful completion of this work.



Jayshree Hazarika

Roll No.: 136104016

Date:

Place: Guwahati

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABT	Aggregated Boosted Trees
AGCM	Atmospheric General Circulation Model
ANFIS	Adaptive-Network-based Fuzzy Inference System
ANN	Artificial Neural Network
AOGCM	Atmosphere-Ocean (or coupled) General Circulation Model
APHRODITE	Asian Precipitation – Highly Resolved Observational Data Integration Towards Evaluation of Water Resources
AR4	Fourth Assessment Report
AR5	Fifth Assessment Report
ASD	Automated Statistical Downscaling
BFSC	Barcelona Field Studies Centre
BMA	Bayesian Model Averaging
BMP	Best Management Practice
CART	Classification and Regression Tree
CCCSN	Canadian Climate Change Scenarios Network
CDF	Cumulative distribution Function
CFSR	Climate Forecast System Reanalysis
CGCM	Atmosphere-Ocean (or coupled) General Circulation Model
CMIP	Coupled Model Intercomparison Project
CORDEX	Coordinated Regional Climate Downscaling Experiment
CV	Cluster Validity index
DECK	Diagnostic, Evaluation and Characterization of Klima
DEM	Digital Elevation Model
DS	Dual Simplex
DTR	Diurnal Temperature Range
ECA	European Climate Assessment
EC-EARTH	European Consortium Earth System Model
EMP	Ecological Management Practice
EOF	Empirical Orthogonal Functions

ESM	Earth System Model
FCM	Fuzzy C- Means
FES	Fuzzy Expert System
FFA	Flood Frequency Analysis
GA	Genetic Algorithm
GAM	Generalized Additive Model
GCM	Global Climate Model or General Circulation Model
GEV	Generalized Extreme Value
GFDL	Geophysical Fluid Dynamics Laboratory
GHG	Greenhouse Gas
GIS	Geographic Information System
GLM	Generalized Linear Model
GNO	Generalized Normal
GOF	Goodness-Of-Fit
GOLD	Generalized Ocean Layer Dynamics
GPL	General Public License
GUI	Graphical User Interface
HCA	Hierarchical Cluster Analysis
ICA	Independent Component Analysis
IDF	Intensity–Duration–Frequency
IMD	Indian Meteorological Department
IPCC	Intergovernmental Panel on Climate Change
IQR	Inter Quartile Range
ISO	Integrated Simulation-Optimization
KANN	Kohonen Artificial Neural Networks
KNN	K-nearest-neighbour
LHS	Latin Hypercube Sampling
LSAV	Large Scale Atmospheric Variable
MK	Mann-Kendall
MLR	Multiple Linear Regressions
MME	Multi-model Ensemble
MOM	Method of Moments
MOU	Memorandum of Understanding

NARCCAP	North American Regional Climate Change Assessment Program
NCEP	National Centers for Environmental Prediction
NEM	Northeast Monsoon
NERC	Natural Environmental Research Council
NNHMM	Nonparametric Nonhomogeneous Hidden Markov Model
NOAA	National Oceanic and Atmospheric Administration
OGCM	Oceanic General Circulation Model
OLS	Ordinary Least Squares
OSGeo	Open Source Geospatial
PC	Principal Component
PCA	Principal Component Analysis
PCMDI	The Program for Climate Model Diagnosis and Intercomparison
PE3	Pearson type III
PRMS	Precipitation Runoff Modelling System
PRP	Partial Correlation
PRUDENCE	Prediction of Regional scenarios and Uncertainties for Defining European Climate change risks and Effects
RCM	Regional Climate Model
RCP	Representative Concentration Pathways
RMSE	Root Mean Square Error
ROI	Region of Influence
RRC	Representative Regional Catchments
RSM	Regional Spectral Model
RVM	Relevance Vector Machine
RWH	Rainwater Harvesting
SAF	Severity-Area-Frequency
SDSM	Statistical Downscaling Model
SOFM	Self-Organization Feature Map
SOM	Self-Organizing Maps
SP	Super Predictor
SPI	Slope Steepness Index
SRES	Special Report on Emission Scenarios
SREV	Square Root Error Variance

SRTM	Shuttle Radar Topography Mission
SVM	Support Vector Machine
SWAT	Soil and Water Assessment Tool
SWM	Southwest Monsoon
TAR	IPCC's Third Assessment Report
TRI	Terrain Ruggedness Index
UKMO	United Kingdom Met Office
WCRP	World Climate Research Programme
WEPP	Water Erosion Prediction Project
WGCM	Working Group on Coupled Modelling
WMO	World Meteorological Organization

Chapter 1

INTRODUCTION

Climate change, in the recent years, is gaining attention of the climatologists as well as environmental and water-resources engineers because of its expected impacts on the environment as well as on human beings. The global impact of climate change and its potential effects on various fields of sustainable development has threatened the whole world. Various researchers are working on both global as well as local impact of climate change on hydrology and water resources. A number of climate models are available that provides us with a useful dataset of atmospheric variables for climate change impact studies. However, the level of uncertainty associated with the climate change projections is always very high. It has been seen from previous studies that, for climatic parameters like rainfall, the correlation of the parameter with atmospheric variables during the base period (or observed period) itself is quite low, which eventually results in poorer performance of models developed for future climate projections. Therefore, it is felt that the efforts need to be directed towards reducing the uncertainty in the result that is obtained from these climate change projections. Furthermore, the uncertainty level varies from model to model. The model that is found to be best fit for one region may not be applicable to a different region with different hydrological characteristics than the previous one. Besides, these models are available at coarse resolution at a global scale, which indicates the need of downscaling these models to the region of study. The uncertainty level also depends on the downscaling approach used for the analysis. Hence, clustering of a region with homogeneous hydrological characteristics might be done before proceeding with the climate change impact analysis, which will help in reducing these uncertainties to an extent and may enhance the results of climate projections obtained from the developed model.

In vast countries like India, where climate conditions vary from region to region, climate change is of major concern, because the economic performance and social progress of the country are dependent on rainfall. The north-eastern part of India which receives heavier rainfall than other parts of the subcontinent suffers from high intensity rainfall of short duration and longer dry spells due to impact of climate change. It also affects river morphology due to flood or drought. Hence, assessment of climate change impacts in the region has gained a lot of attention from the research community in the recent years. However, the peculiar orography

of the region indicates that the orographic arrangement of the region also impacts the results of climate projections. Hence, there is a need of inclusion of orographic influence on climate change impact studies and to assess its effects on the overall result.

This chapter introduces the causes of climate change and its impacts on various fields of water resources with emphasis on the Indian subcontinent and the northeast Indian region. In the later part of the chapter, the scope and objective of the study has been described.

1.1 CLIMATE CHANGE

Climate Change is defined as the change in the statistical properties of the climate system of a region when considered over a long span of time. The cause of the change may be due to human influence or due to nature. As per the definition given by Canadian Climate Change Scenarios Network, "*Climate change is defined as a difference over a period of time (with respect to a baseline or a reference period) and corresponds to a statistical significant trend of mean climate or its variability, persistent over a long period of time (e.g. decades or more)*". As mentioned in the Fifth Assessment Report (AR5) of Intergovernmental Panel on Climate Change (IPCC), "*Climate change may be due to natural internal processes or external forcing such as modulation of the solar cycles, volcanic eruptions, and persistent anthropogenic changes in the composition of the atmosphere or in land use*."

The IPCC has been constantly delivering synthesis reports on the growing impacts of the changing climate. Earlier, in Synthesis report-2007, IPCC confirmed that climate change due to human activities and the consequences are likely to be serious in the near future. Rapid changes in climate, mostly caused by global warming, will enhance the temperature rise, sea level rise and precipitation change. According to the latest Synthesis report-2014, IPCC is now 95 percent confident that humans are the prime cause of current global warming. The more human activities disturb the climate, the greater the risks of severe, irreversible and long-lasting impacts. Furthermore, the longer we wait to take necessary actions, the more will be the cost and greater the economic, social, technological and institutional challenges we will face. The long-term impacts of the changing climate are manifold. It is, therefore, important to assess the climate change impacts on key climate parameters, such as temperature, precipitation and wind, which are important for climate sensitive sectors such as, energy, water management, agriculture and health.

1.2 GLOBAL WARMING AND THE CHANGING CLIMATE

Global Warming is considered as a major aspect of climate change. It is the long-term increase of Earth's average surface temperature due to effect of greenhouse gases, which trap the heat in the Earth's atmosphere that would otherwise escape. Greenhouse gases, as defined by the IPCC, are "*The gaseous constituents of the atmosphere, both natural and anthropogenic, that absorb and emit radiation at specific wavelengths within the spectrum of thermal infrared radiation emitted by the Earth's surface, the atmosphere itself and by clouds*." Greenhouse gases increase due to human activities such as burning of fossil fuel, deforestation, industrialization as well as rapid urbanization. Greenhouse gases include carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O) and ozone (O_3). Moreover, there are a number of entirely human-made greenhouse gases in the atmosphere, such as the halocarbons, sulphur hexafluoride (SF_6), chlorofluorocarbons (CFCs), hydrofluorocarbons (includes HCFCs and HFCs), perfluorocarbons (PFCs) and other chlorine- and bromine-containing substances. The major contributing greenhouse gas is carbon dioxide. The global annual mean concentration of CO_2 in the atmosphere has increased by more than 45% from the pre-industrial value (since the start of Industrial revolution, around 1750) of about 280 ppm to some 406 ppm in early 2017. This value has now reached to approximately 414 ppm as of June 2019, as per the studies done by Earth System Research Laboratory, National Oceanic and Atmospheric Administration (NOAA) of United States. It will take centuries to reduce this concentration, and perhaps more than a century even to stop it from increasing. Meanwhile, this increase has started to show impacts on the climate and has committed us to further think about reducing the impact of climate change.

Impact of global warming is evident everywhere. The World Meteorological Organization (WMO) in a recent press release (February, 2019) declared that 2015, 2016, 2017 and 2018 are the four warmest years on record. The long-term temperature trend has always been found to be upward, the last two decades being the warmest. However, the degree of warming during the past four years (2015-2018) has been exceptional. The year 2016 remains as the warmest year on record with a temperature increase of approximately 1.2°C above preindustrial baseline (1850-1900). A consolidated analysis of five leading international datasets by WMO exposed that the global average surface temperature in the last year i.e. 2018 was almost 1.0° Celsius (with an error of ±0.13°C) above the pre-industrial baseline. It ranks as the fourth warmest year on record.

1.3 IMPACTS OF CLIMATE CHANGE

According to the Working Group II's contribution to the Fifth Assessment Report (AR5) of IPCC, the areas mainly affected by climate change are divided into three major categories (Figure 1.1): Physical systems, Biological systems and Human and managed impacts.

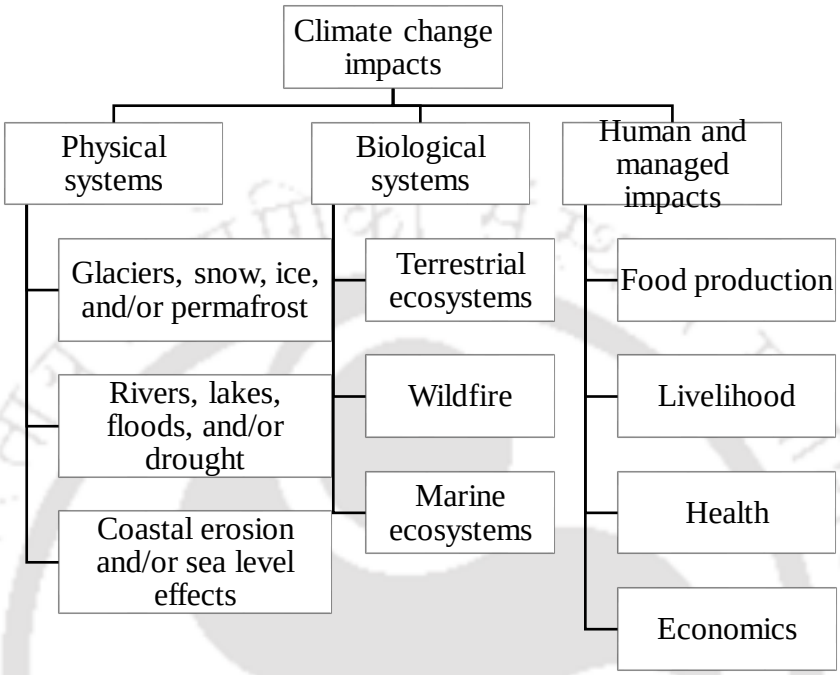


Figure 1. 1 Categorization of impacts attributed to climate change

Glaciers are shrinking, snow and ice caps are melting in many regions, which is affecting the runoff and downstream water resources such as rivers, lakes, and eventually altering the hydrological systems of the regions. Climate change is also causing permafrost warming and thawing in high-latitude, high-elevation regions. The number of extreme precipitation events has increased in many regions, exposing those catchments to greater risk of flooding. Whereas in some areas, changes in the rainfall patterns have caused droughts and wildfires. Rise in the mean sea level has increased the risk of coastal erosion, inundation of low-lying cities and islands with seawater. Several freshwater, terrestrial and marine species have started to shift their seasonal activities and migration pattern in response to ongoing climate change, although only a few species extinctions have been ascribed to climate change with high confidence. Melting of the glaciers and ice caps near the poles have caused loss of habitat of polar bears, that are now thought to be greatly endangered. Climate change has contributed to ocean acidification, which refers to decrease in the pH level of the ocean that makes the ocean more acidic. It occurs because of long-term change in seawater chemistry due to the absorption of carbon dioxide from the atmosphere. According to NOAA, the global average pH of the surface

of ocean has reduced by 0.11 since the Industrial Revolution, which corresponds to around 30% increase in the hydrogen-ion concentration. Ocean acidification imposes significant menaces to marine ecosystems, especially polar ecosystems, shellfish and coral reefs.

Climate change has shown negative impact on the production of crops like wheat, maize, rice, soybean etc. Although some has claimed that the production has increased in some high-latitude places, majority of the regions has shown decrease in the crop production, indicating food insecurity and rapid rise in food price in those regions. These impacts affect the livelihood of people directly or indirectly, especially of those living in poverty. Climate-related extremes, such as floods, droughts, cyclones, heat waves, and wildfires, exposes some ecosystems and many human systems to vulnerable environment, like decreased water supply, morbidity and mortality, outbreak of waterborne diseases, damage of settlements and infrastructure, and consequences for mental health and human well-being.

1.3.1 Impacts of climate change on global hydrologic cycle

The increasing global warming affects the hydrologic cycle, altering the rainfall and eventually the runoff. Warm air can hold more moisture and increases the evaporation of surface moisture which tends to increase the precipitation. However, increased evaporation and evapotranspiration leads to reduction in soil moisture. Precipitation pattern will considerably be changed, with visible variations in the seasonal distribution and amount of precipitation. The consequences of these changes will be seen in the form of flood in some regions, while other regions being hit by drought. Mountainous regions may experience changes in the balance between snowfall and rainfall. Melting of glacial ice has already been seen in most of the mountainous regions, which has reduced the size of the glaciers to a great extent. This leads to increase in the downstream flow, subsequently increasing the discharge in rivers and streams. However, long term shrinking of glaciers would eventually reduce their contribution to annual runoff, affecting water availability in those waterbodies. Due to the changes in temperature and precipitation pattern, vegetation cover will also face seasonal changes. Risk of wildfire has also increased in many regions. Coastal erosion and inundation have increased due to rise in the sea level. All these changes will ultimately lead to change in the management of land resources, along with water resources.

IPCC-AR4 found that due to climate change, annual precipitation increases in tropics and at high latitudes and decrease in sub-tropics. More precipitation will increase a region's

susceptibility to hazards, depending on variety of factors including flooding, rate of soil erosion, soil moisture availability and mass movement of land.

1.3.2 Future risks associated with impacts of climate change

IPCC's AR5-Working Group II report explains the followings as key risk areas, which are susceptible to climate change related hazards:

- Coastal zones in low-lying areas and small island have high risk of death, ill-health, injury and disrupted livelihoods, owing to coastal flooding and storm surges as consequences of sea level rise.
- Urban populations, residing in regions affected by inland flooding, have risk of severe ill-health and disturbed livelihoods.
- Risks associated with extreme weather events will lead to infrastructural breakdown, causing interruptions in critical services like water supply, electricity, health and emergency services.
- Extreme heat conditions will cause high risk of illness and mortality, especially for vulnerable urban and rural populations, working outdoors.
- Warming climate, drought and flood events, variation in precipitation and extreme events will lead to high risk of food insecurity, along with breakdown of food system. This will particularly hit the poorer population of urban and rural areas.
- Poorer farmers with nominal assets in semi-arid regions have risk of loss of income as well as livelihood, owing to reduced production in the agricultural sector and insufficient access to irrigation and drinking water.
- The fishing community in the tropical areas and the Arctic region will be affected due to risks associated with marine and coastal ecosystem loss, changes in biodiversity, and due to risk of loss of ecosystem services related to coastal livelihood.
- Terrestrial and inland water ecosystems have risk of loss, biodiversity changes, and will affect the associated functions and services of ecosystem required for livelihoods.

These risks associated with climate change need to be addressed with proper adaptation and mitigation strategies with implications for future generations, environments and economies. These strategies will not only reduce the current and future risks, but will also have additional benefits, such as improvement of livelihoods, biodiversity conservation, and human well-being with improved health conditions.

1.3.3 Impacts of climate change on water resources: Indian context

Water is one of the key areas affected largely by climate change. All the risks mentioned in the previous section are directly or indirectly related to water resources. Some of the regions are at high risk of abundant water availably causing extreme events like heavy precipitation, floods, cyclones, cloud bursts etc. In contrast, some other regions are at risk of water scarcity, extreme dry condition, groundwater depletion, drought, wildfire etc. In both cases, proper water resources management is necessary to adapt and mitigate the impacts.

Water resources in India are no exception. They are also affected by the changing climate. Mehrotra and Mehrotra (1995) mentioned that, the major impacts of climate change in India would be on the hydrology, water resources and agriculture of the country. In Synthesis report-2007, IPCC stated that if the current trend of glacial melt continues like this, then there is a possibility that the snow fed rivers of India such as Ganga, Brahmaputra could become seasonal rivers in the near future, affecting the lives of millions of people. Continuous loss of ice from the glaciers has changed their behavior. Shift of annual runoff peak from summer to spring season is expected in the eastern Himalayan region in the near future, as the contribution of glacial ice would be greatly reduced because of continuous shrinking. As the Himalayan region is glacier-dependent, livelihoods of the people will be threatened because of excessive runoff and flood events (Xu et al., 2009). Substantial growth in glacial lake areas has been evident in the eastern Himalaya region between years 1990 and 2009, which was mainly contributed by climate change (Gardelle et al., 2011).

Regarding extreme wind events, IPCC's AR5-Working Group II report states that India suffer 86% of mortality from tropical cyclones, which have major impacts on society and ecosystems. These tropical cyclones occur mainly and most commonly in coastal settlements. Increase in the concentrations of people, growth of infrastructure and industrial development along the coasts could increase the vulnerability in the near future, if adaptation measures are not being taken. The report also states that climate change will affect the water resources of Asian region in a major way. Although different river basins might experience different level of impacts, the consequences of climate change on water availability and food security is evident. Climate change affects the agricultural sector by affecting crop production. Emission of high concentration of CO_2 indirectly affects crop production by increasing tropospheric ozone (O_3). Severe impacts on crop yields could be seen in Indian region as production of major crops like maize, rice, wheat, soybean have seen losses in the recent years (Van Dingenen et al., 2009).

1.3.4 Importance of climate change impacted water resources planning and management in Northeast India

The water resources of the northeast Indian region are primarily dependent on the Brahmaputra basin. Other basins affecting the hydrology of the region are the Barak basin and part of the Ganga basin. Studies have shown that, climate change has major impact on the water availability of climate-susceptible basins such as the Ganga and Brahmaputra basin (Immerzeel et al., 2010). Climate change may disturb the dynamics of the river system by changing precipitation pattern and runoff quantity, eventually altering the seasonal water flows. The northeast Indian region receives heavier rainfall than other parts of the subcontinent, hence suffers from high intensity rainfall for short duration and longer dry spells. It also affects river morphology due to flood or drought.

Agriculture in the north-eastern region is mainly rainfed and more than half of the north-eastern population is dependent on agriculture, either directly or indirectly. Hence, economic performance and social progress of the region are hugely dependent on precipitation. Therefore, the consequences of climate change impacted variations have potentiality of affecting the food security of millions of people in the region. Furthermore, since Brahmaputra basin is a transboundary water basin, the climate-induced impacts on water availability of the basin have potential of affecting the geopolitical concerns also owing to the high international interdependency. Therefore, planning and management of water resources including the impacts of climate change is necessary for northeast Indian regions.

1.4 USAGE OF CLIMATE MODELS AND DOWNSCALING TECHNIQUES FOR CLIMATE RESEARCH

1.4.1 General Circulation Model or Global Climate Model (GCM)

GCM is a type of climate model that represents the general circulation of the planetary atmosphere or ocean as a mathematical model. These models are generated based on the Navier–Stokes equations on a rotating sphere, where various energy sources (radiation, latent heat) are expressed with thermodynamic terms. To develop a GCM, the planet is divided into 3-dimensional grids. The atmospheric variables like winds, relative humidity, radiation, heat transfer, surface hydrology etc. are calculated within each grid and the GCM datasets are then made available for those grid points. These GCM datasets are widely used for understanding

present climate, weather forecasting and to project the climate conditions into the future. The Intergovernmental Panel on Climate Change (IPCC) provides a platform for downloading these model datasets, with access to various GCMs developed by different climate centres and organizations all over the world.

The GCMs can be broadly classified into two key components: atmospheric GCMs (AGCM) and oceanic GCMs (OGCM). They also combine other components such as land-surface and sea ice components. AGCMs and OGCMs can be coupled together to form a coupled GCM, named as atmosphere-ocean (or coupled) general circulation model (CGCM or AOGCM). With the addition of sea ice component and a model for evapotranspiration over land, AOGCMs become a full climate model. Recently a new model has been introduced, named as Earth System Models (ESM). These ESMs are based on a coupled atmospheric and oceanic circulation model, with representations of sea ice, land and iceberg dynamics and they incorporate interactive biogeochemistry, including the carbon cycle. Hence, ESM is also a coupled global climate model that can explicitly model the movement of carbon through the earth system. ESM has the added capability to represent biogeochemical processes that interact with the physical climate and so alter its response to forcing such as that associated with human-caused emissions of greenhouse gases (Flato, 2011).

The GCMs have been in extensive use for water resources related studies. Simulations with GCMs have significantly advanced with substantial improvements in computational power. GCM can be expediently used for regional studies with large spatial coverage. However, they are unable to efficiently represent the small-scale features at sub-grid levels, as the local hydrological features and dynamics are not exploited explicitly by the GCMs (Wigley and barnett, 1990). This may the reason due to which they can accurately simulate unconditional (or independent) climate variables such as temperature, humidity etc. However, simulation of conditional variables like precipitation, evaporation, runoff etc. is not satisfactory, specially at a local scale.

1.4.2 Coupled Model Intercomparison Project (CMIP)

The Coupled Model Intercomparison Project (CMIP) is a collaborative framework, formed by the Working Group on Coupled Modelling (WGCM) of the World Climate Research Programme (WCRP). It was formed in 1995 with an aim to improve the understanding of climate change related knowledge. CMIP provide a global platform to a diverse community of

climate scientists to analyse GCMs for validation, intercomparison as well as documentation. A number of international climate modelling communities have been contributing to the CMIP project since its commencement.

Main purpose of CMIP is to improve understanding of the past, present and future climate changes that may arise due to natural or anthropogenic influences. This includes assessing the performance of various GCMs during the historical period as well as quantification of the magnitude of long-term impacts of the changing climate by projecting into the future. CMIP provides data access to these multi-model outputs in standardized formats, that can be used by the research community for assessment of climate change.

The CMIP assessments and documentation are developed in phases. The Program for Climate Model Diagnosis and Intercomparison (PCMDI) at Lawrence Livermore National Laboratory has been supporting the CMIP in all the phases and maintains the project's data base. The first and second phases (CMIP1 and CMIP2) involved comparing the model response to an idealized forcing, where a constant rate of CO_2 increase of 1% per year was integrated. Around 18 global coupled models participated in the project development. Phase three of CMIP (CMIP3) came during 2005-2006, that included simulation of past, present future climate scenarios. The concept of emission scenario came with CMIP3. IPCC's Special Report on Emissions Scenarios (SRES), that was published in 2000, describes the projection of various possible future climatic conditions based on different greenhouse gas emissions scenarios. The SRES scenarios are used in the IPCC's Third Assessment Report (TAR) and the Fourth Assessment Report (AR4).

CMIP5 is the latest development of CMIP (2010-2014). The new set of experiments of CMIP5 were developed with an objective to address the scientific queries arose from IPCC AR4 assessments. These CMIP5 models are expected to further improve understanding of climate. IPCC's Fifth Assessment Report (AR5) summarizes the information related to CMIP5 experiments. Most of the CMIP5 simulations were performed with prescribed CO_2 concentrations. The concept of Representative Concentration Pathways (RCPs) came along with CMIP5. The RCPs represent greenhouse gas concentrations (instead of emission scenarios like CMIP3) for climate modelling.

Currently, the sixth generation of CMIP (CMIP6) is under progress. The models in CMIP6 will run at high resolutions and will include some additional processes that were not incorporated

or simulated in the previous models. Also, the infrastructure and documentation will be improved for better access to the models by the scientific community. More than 30 climate research groups around the globe have participated in the CMIP6 project.

CMIP6 consists of three major components (Eyring et al., 2016) which are given below:

- Identification of common experiments, the DECK (Diagnostic, Evaluation and Characterization of Klima) experiments and historical simulations of CMIP (1850 – near-present), that will help in establishing the model characteristics.
- Common standards, coordination, infrastructure and documentation that will facilitate the distribution of model outputs and the characterization of the model ensemble.
- Adoption of a more federated structure, building on more autonomous CMIP-Endorsed MIPs.

1.4.3 Features of CMIP3 and CMIP5

1.4.3.1 Emission Scenarios

The concept of emission scenario came with CMIP3. Scenarios are images of the future, or alternative futures. They are neither predictions nor forecasts. As per Special Report on Emissions Scenarios (SRES) by IPCC, "Scenarios are alternative images of how the future might unfold and are an appropriate tool with which to analyze how driving forces may influence future emission outcomes and to assess the associated uncertainties. They assist in climate change analysis, including climate modelling and the assessment of impacts, adaptation, and mitigation."

Emissions scenarios are a central component of any assessment of climate change. The Greenhouse Gases (GHGs) and other gases covered in the emission scenario are CO_2, CO, Hydrochlorofluorocarbons (HCFCs), Hydrofluorocarbons (HFCs), CH_4, Nitrous Oxide (N_2O), Nitrogen Oxides (NOx), Non-Methane Volatile Organic Compounds (NMVOCs), Perfluorocarbons (PFCs), Sulfur Dioxide (SO_2), Sulfur Hexafluoride (SF_6). Among those CO_2, CH_4 and SO_2 emissions are the basic input for determining future climate patterns with simple climate models, as well as with complex general circulation models (GCMs).

The set of scenario consists of four Scenario families viz. A1, A2, B1 and B2. A1 is further divided into three scenario groups namely A1F1, A1T and A1B. 40 emission scenarios are

developed from these six modelling groups (Fig 1.2). The four scenario groups are described below:

- **A1 scenario**: - The Al scenario describes a future world with very rapid economic growth, low population growth, rapid introduction of new and efficient technologies, increased cultural and social interactions, substantial reduction in regional differences in per capita income. The three Al sub-groups are described as fossil fuel intensive (AlF1), non-fossil energy sources i.e. technologically advanced (A1T), and a balance across all sources (mix of fossil and non-fossil fuel) (A1B).
- **A2 scenario**: - The A2 scenario describes a very heterogeneous world, with more emphasis on family values and local traditions, self-reliance and preservation of regional and cultural identities, continuous global population growth, regionally oriented economic development. Technological changes are uneven and slower than other scenarios.
- **B1 scenario**: - The B1 scenario describes a convergent world with the same global population as in the A1 scenario, but with rapid changes in economic structures, reductions in material intensity, introduction of clean and resource-efficient technologies. Main emphasis is given on global solutions to economic, social, and environmental sustainability that will improve equity, but without additional climate initiatives.
- **B2 scenario**: - The B2 scenario describes a world, in which main emphasis is given on local solutions to economic, social, and environmental sustainability. It describes a heterogeneous world with continuous global population growth (at a rate lower than A2), intermediate economic development, diverse and slower technological change than other scenarios. While the scenario is also oriented toward environmental protection and social equity, it focuses on local and regional levels rather than global solutions.

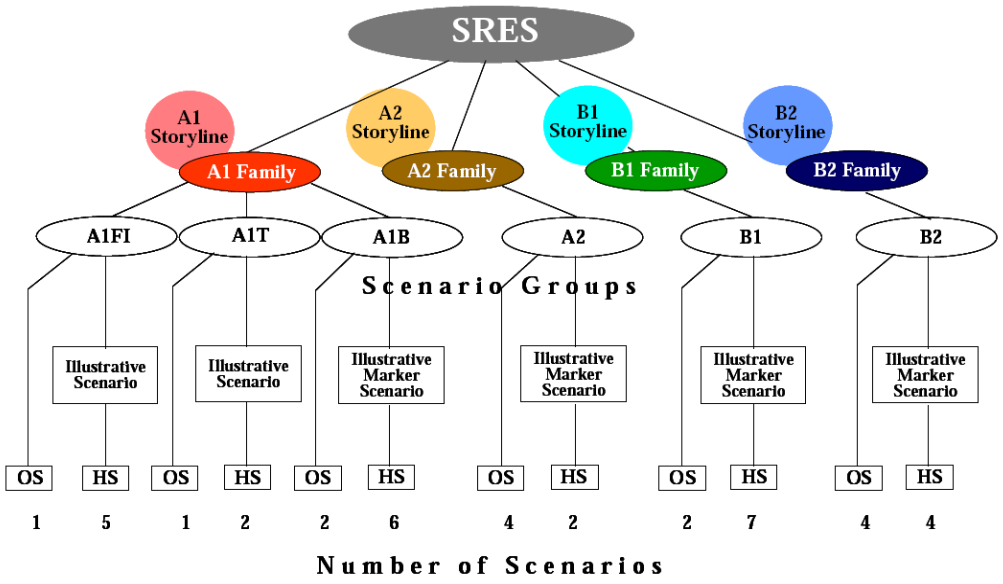


Figure 1. 2 Emission scenarios (source: IPCC website)

1.4.3.2 Representative Concentration Pathways (RCPs)

A Representative Concentration Pathway (RCP) characterizes greenhouse gas concentration (instead of emission scenarios) that has been used for climate modelling. The concept of Representative Concentration Pathways (RCPs) came along with CMIP5. IPCC's Fifth Assessment Report (AR5) summarizes the simulations of CMIP5, which are based on four different pathways. Each of these four RCPs describe a different future under climate change, which is considered to be possible. Each RCP represents a different volume of future GHG concentration, possible under a wind range of anthropogenic changes. The four RCPs are explained below:

- **RCP2.6 (radiative forcing in the year 2100 is 2.6 W/m^2)**: - RCP2.6 considers that CO_2 emissions start decreasing by 2020 and reach zero by the end of 2100. Other GHGs are also expected to show decreasing trend, with methane emissions (CH_4) reaching approximately half of those in 2020, sulphur dioxide (SO_2) emissions to reach about 10% of those of 1980-1990. RCP2.6 requires CO_2 emissions to show negative trend (such as absorption of CO_2 by plants). It is also expected to retain global temperature rise below 2 °C by the year 2100.
- **RCP4.5 (radiative forcing in the year 2100 is 4.5 W/m^2)**: - RCP4.5 is an intermediate scenario, where the GHG emissions peak around 2040, and then decline. CO_2 emissions start decreasing by the year 2045 and reach approximately half of the CO_2 levels of

2050 by the year 2100. CH_4 emissions start decreasing by 2050 and reach about 75% of the levels of 2040, and SO_2 emissions decline to approximately 20% of those of 1980-1990. RCP4.5 also requires negative CO_2 emissions and is expected to retain global temperature rise between 2 °C and 3 °C, by the year 2100. Mean sea level rise will be approximately 35% higher than that of RCP2.6. Many plant and animal species are believed to be unable to adapt to the impacts caused by RCP4.5 and higher RCPs.

- **RCP6.0 (radiative forcing in the year 2100 is 6 W/m^2)**: - RCP6.0 is also an intermediate scenario, where the GHG emissions peak around 2080, and then decline.
- **RCP8.5 (radiative forcing in the year 2100 is 8.5 W/m^2)**: - RCP8.5 considers continuous rise of GHG emissions throughout the 21st century. As per IPCC, this condition is thought to be very unlikely, as this is based on overestimation of projected coal outputs. RCP8.5 is considered as the basis for a worst-case climate change scenario.

1.4.4 Downscaling techniques

Downscaling is a method for obtaining high-resolution climate change information from relatively coarse-resolution global climate models (GCMs). Typically, GCMs have coarse resolution, covering large geographical areas. Many impact assessment models require information at finer scales of 50 km or less. To overcome this problem downscaling methods are developed. Primary assumption of these downscaling approaches is that the large-scale climate has influence on local scale climate to a great extent. Two main forms of downscaling technique exist. They are:

- **Dynamic downscaling**
- **Statistical downscaling**

1.4.4.1 Dynamic downscaling

Here output from the GCM is used to derive a regional, numerical model in higher spatial resolution, which therefore is able to simulate local conditions in greater detail. Dynamic downscaling represents the use of high-resolution Regional Climate Models (RCMs) which are nested with GCMs. The RCMs are similar to GCMs, but RCM generally improves with the higher-order statistics of the meteorological variables. It is a step by step process. A drawback of RCM is large demand of computer resources and the complexity of their operation, which requires trained persons.

1.4.4.2 *Statistical downscaling*

Here a statistical relationship is established from observations between large scale (GCM) variables, like atmospheric surface pressure, and a local (often station level) variable, like the wind speed at a particular site. The relationship is then subsequently used on the GCM data to obtain the local variables from the GCM output. This technique is broadly divided into three categories, namely

- Weather typing
- Weather generators
- Regression-based downscaling.

In weather typing local meteorological data are grouped in relation to existing atmospheric circulation pattern. Future scenarios are generated either by resampling observed datasets or by first generating synthetic series, using Monte Carlo technique, and then resampling the observed data. Although the technique is applicable for a wide variety of variables and has multisite applicability, it has some limitations. Specially, in case of precipitation downscaling, the series simulated by this method is sometimes found to be inconsistent with the GCM models used.

The weather generator is one of the popular techniques, that demands less computational efforts. It has the advantage of producing exact observed climate statistics. This model simulates precipitation using two–state, first order Markov chains: precipitation amounts on wet days using a gamma distribution; temperature and radiation components using first–order trivariate autoregression that is conditional on precipitation occurrence (Wilby and Dawson, 2007). It has the limitation of arbitrary adjustment of precipitation parameters in future simulations.

Among these approaches, regression methods are preferred because of its ease of implementation and low computation requirements. It establishes an empirical relationship between the predictand and the predictor variables.

Regression analysis can be done in two ways:

- Linear regression, which analyzes two variables – one predictand variable (called the criterion) and one predictor variable.
- Multiple linear regression, which analyzes three or more variables – one criterion and two or more predictor variables.

1.4.5 Projection of Future Climate using GCM and their Downscaling

Many researches and projects on climate change are already going on that emphasize on the use of various GCM models (Mehrotra and Mehrotra, 1995; Heyen et al., 1996; Wilby et al., 1998a,1998b; Gosain et al., 2006; Rahman et al., 2010). Eliasson (2000) suggested that the urbanization is always related to climate, i.e., urbanization induces climate change and the climate change causes more problem in urbanization. Due to climate change, severe water scarcity in one region and flood hazards in other regions occur (Sharma and Shakya, 2006; Mujumdar, 2008). Several studies have also been done on impact of climate change on water resources planning in the northeast Indian region. Studies have been conducted to investigate the extent of climatic variation and scope of utilizing rainwater harvesting (RWH) system for irrigating small agricultural field in Assam (Kumar and Sarma, 2010). Analysis of climatic parameters has revealed the increase in dry spell, increase in temperature range, and decrease in total annual rainfall. Studies on impact of climate change on the precipitation characteristics of Brahmaputra Basin shows that temperature may go up to 1.8˚C rise in the future (Deka and Sarma, 2011). Sarma et al. (2012) has carried out studies on the impact of climate change on water resources of Brahmaputra basin and adaptation strategies with regard to planning, design and management of water resources systems of the basin. The study indicated significant changes in temperature and rainfall pattern of this basin in future. Sarma et al. (2013) developed an optimization model for optimal Ecological Management Practice (EMP) to minimize impact of climate change and degradation of watershed due to urbanization. The study provides an idea for enhancing sustainability of EMPs by utilizing the carbon sequestration potential of vegetation, which can be adopted by industries to earn carbon credits, thus compensating their greenhouse gas (GHG) emissions.

General Circulation Models (GCMs) are widely used in climate change studies; but GCMs are relatively coarse in resolution. Therefore, some form of downscaling is required to produce scenario of finer spatial resolution than currently delivered by raw GCM output. Number of downscaling techniques and GCM combinations were experimented by various researchers. Wilby et al. (1998b) made a comparison of different Statistical downscaling methods using GCM output. Weather generator methods are better than methods like ANN for climate change porjections. Zorita and von Storch (1999) suggested a simple analog method for statistical downscaling and compared the same with other available complex methods. Misra et al. (2003) found that Regional Spectral Model (RSM) performs better than GCM models. Dibike and Coulibaly (2005) have made comparisons of hydrologic as well as downscaling models under

climate change impacts. They predicted future variation in river flow and reservoir inflow using downscaled data from SDSM (regression-based approach) and LARS-WG (Weather Generator). Zhang (2005) developed a simple method for statistical downscaling of GCM output to predict soil erosion and crop production. Mehrotra and Sharma (2005, 2006) attempted to develop nonparametric stochastic spatial downscaling methods for multisite daily rainfall projections. Ghose and Mujumdar (2006) developed a methodology to forecast future rainfall scenario over Orissa using GCM projections by statistical downscaling. The rainfall projection shows that there is a possibility of increase in hydrologic extremes in Orissa in future. Kang et al. (2007) have done statistical downscaling using six dynamic models and predicted the station scale precipitation in the Philippines and Thailand. They found that in some location, the statistical downscaling is not suitable because precipitation is governed mainly by local complicated terrain other than large-scale process. Serrat-Capdevila et al. (2007) have done statistical downscaling of GCM models to quantify the climate change impacts on hydrology. Anandhi et al. (2008) have presented a methodology for precipitation downscaling to river basins in India, with the use of support vector machine. Hessami et al. (2008) have done statistical downscaling of temperature and precipitation using automated statistical downscaling (ASD). Mujumdar and Ghosh (2008) have described a methodology of statistical downscaling based on Support Vector Machine (SVM) and Relevance Vector Machine (RVM). Tisseuil et al. (2010) compared four statistical downscaling models namely Generalized Linear Model (GLM), Generalized Additive Model (GAM), Artificial Neural Network (ANN) and Aggregated Boosted Trees (ABT). They used these models for predicting stream flow. Among these models, ABT was found to perform better. Kannan and Ghosh (2011) attempted to address the limitation of statistical downscaling in capturing correlation between multiple sites by representing the rainfall pattern in a river basin using rainfall state. Najafi et al. (2011a) have done statistical downscaling of precipitation employing independent component analysis (ICA) for predictor selection that determines spatially independent GCM variables. By doing uncertainty ana lysis of the downscaling methods, they found that model structures and input predictors are important factors in the downscaling process and the main sources of uncertainties. Deka and Sarma (2011) have carried out a study to quantify the impact of the climate change on the precipitation characteristics of Brahmaputra Basin. The future series generated by temperature model has produced up to 1.8˚C rise in the maximum temperature averaged over month. In the process of model development, it was experienced that the selection of predictors plays an important role in future series generation through

downscaling of GCM. Vinnarasi (2012), in her project work, has focused on simulating the impact of climate change with the use of downscaling methods, on the precipitation characteristics and stream flow behaviour of Dhansiri River. Salvi et al. (2013) have projected daily rainfall of multiple sites in India using statistical downscaling at a resolution of 0.5^0, using Classification and Regression Tree (CART) and nonparametric kernel regression technique. Forsythe et al. (2014) have applied stochastic weather generator to assess climate change impacts in a semi-arid region. They stated that sophisticated downscaling methods were needed to evaluate the climate change impacts and the interannual variabilities for better understanding of the future conditions. Fatichi et al. (2015) have done a high-resolution fully distributed hydrological analysis and future climate projections. They found that there is an elevation dependence of climate change impacts in the catchment under consideration. Shamir et al. (2015) have analysed climate change projections of precipitation using eight dynamically downscaled GCM models. They stated that climate change projections increase the uncertainty and aggravates the complexity of water resources management task.

1.5 USAGE OF HOMOGENEOUS CLUSTERING FOR CLIMATE RESEARCH

1.5.1 Clustering techniques

1.5.1.1 General Overview of clustering methods

Since clusters can formally be seen as subsets of the data set, one possible classification of clustering methods can be according to whether the subsets are fuzzy (soft) or crisp (hard). Thus, the clustering methods are mainly of following types:

- **Hard Clustering**

Methods, which are based on classical set theory, and require that an object either does or does not belong to a cluster, are referred to as hard clustering methods This means, partitioning the data into a specified number of mutually exclusive subsets. The discrete nature of the hard-partitioning causes difficulties with algorithms based on analytic functionals, since these functionals are not differentiable.

The objective of clustering is to partition the data set Z into c clusters (groups, classes). Generally, c is assumed to be known, based on prior knowledge. Using classical sets, a hard

partition of Z can be defined as a family of subsets $\{A_i \mid 1 \leq i \leq c\} \subset P(Z)1$ with the following properties (Bezdek, 1981):

$$U_{i=1}^{c} A_i = Z \tag{1.1}$$

$$A_i \cap A_j = \emptyset, 1 \leq i \neq j \leq c \tag{1.2}$$

$$\emptyset \subset A_i \subset Z, 1 \leq i \leq c \tag{1.3}$$

In terms of membership (characteristic) functions, a partition can be conveniently represented by the partition matrix $U = [\mu_{ik}]_{c \times N}$. The i^{th} row of this matrix contains values of the membership function μ_i of the i^{th} subset A_i of Z. The elements of U must satisfy the following conditions:

$$\mu_{ik} \epsilon \{0,1\}, 1 \leq i \leq c, 1 \leq k \leq N \tag{1.4}$$

$$\sum_{i=1}^{c} \mu_{ik} = 1, 1 \leq k \leq N \tag{1.5}$$

$$0 < \sum_{k=1}^{N} \mu_{ik} < N, 1 \leq i \leq c \tag{1.6}$$

- **Fuzzy Clustering**

These methods allow the objects to belong to several clusters simultaneously, with different degrees of membership. In many situations, fuzzy clustering is more natural than hard clustering. Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership.

Generalization of the hard partition to the fuzzy case follows directly by allowing μ_{ik} to attain real values in [0, 1]. Conditions for a fuzzy partition matrix are given by:

$$\mu_{ik} \epsilon [0,1], 1 \leq i \leq c, 1 \leq k \leq N \tag{1.7}$$

$$\sum_{i=1}^{c} \mu_{ik} = 1, 1 \leq k \leq N \tag{1.8}$$

$$0 < \sum_{k=1}^{N} \mu_{ik} < N, 1 \leq i \leq c \tag{1.9}$$

The i^{th} row of the fuzzy partition matrix U contains values of the i^{th} membership function of the fuzzy subset A_i of Z. Second equation constrains the sum of each column to 1, and thus the total membership of each z_k in Z equals one.

- **Possibilistic Partition**

A more general form of fuzzy partition, the possibilistic partition, can be obtained by relaxing the second constraint i.e. the total membership should be unity. This constraint, however, cannot be completely removed, in order to ensure that each point is assigned to at least one of the fuzzy subsets with a membership greater than zero. Thus the 2nd equation can be replaced by a less restrictive constraint ∀k, ∃i, $\mu_{ik} > 0$. The conditions for a possibilistic fuzzy partition matrix are:

$$\mu_{ik} \epsilon [0,1], 1 \leq i \leq c, 1 \leq k \leq N \quad (1.10)$$

$$\exists_i, \mu_{ik} > 0, \forall k \quad (1.11)$$

$$0 < \sum_{k=1}^{N} \mu_{ik} < N, 1 \leq i \leq c \quad (1.12)$$

1.5.1.2 Usage of Fuzzy C-means Clustering (FCM) for homogeneous clustering

The FCM approach is basically optimization of fuzzy c-means objective function. It was initially developed by Dunn (1973) and afterwards modified by Bezdek et al. (1984). The fuzzy c-means function is explained in section 3.2.1 in chapter 3. Moreover, the FCM algorithm that can be used for delineation of homogeneous rainfall region is discussed in detail in section.

1.5.2 Development of Homogeneous Climatic Regions using Clustering Techniques

Modelling of hydrological processes (e.g. precipitation, temperature etc.) is immensely complicated since majority of descriptive variables such as soil, topography, land-use and rainfall vary in space under different scales (Chavoshi et al., 2013). Therefore, to overcome the difficulties involved, regionalization of the area of study can be done. Various clustering methods are being used for this purpose. Principal component analysis has been used by many researchers for determining homogeneous regions (Bedi and Bindra, 1980; Bärring, 1987; Sumner and Bonell, 1988; Iyengar and Basak, 1994; Singh and Singh, 1996; Owen et al., 2006; Darand and Mansouri Daneshvar, 2014). Kulkarni et al. (1992) have used k-means clustering. Guttman (1993) has used L-moment test in determination of regional precipitation climates in continental United States. Problems were encountered mainly in mountainous and arid regions, which were resolved in all but three regions by examining orography and/or the data. Fill and Stedinger (1995) performed homogeneity tests based upon Gumbel distribution and had done a critical appraisal of Dalrymple's test. L-moment X-10 test was always found to be more

powerful. Hall and Minns (1999) have done hydrologically homogeneous clustering for flood data of the southwest of England and Wales, using Kohonen Network method and Fuzzy c-means method. Both methods were found to be effective. Wotling et al. (2000) have done regionalization of extreme precipitation intensities in the volcanic island of Tahiti using principal components of the topographical environment. They finally concluded that, this automatic and objective method could be applied in any mountainous area, where topography has a major influence on the precipitation features. Rao and Srinivas (2006a, 2006b) have done regionalization of watersheds in Indiana, USA by using both fuzzy clustering and k-means clustering combined with hierarchical clustering algorithms. Fuzzy clustering improves the homogeneity of clusters, although other methods are also found effective. Abida and Ellouze (2006) have discussed an innovative method for hydrological delineation of homogeneous regions in Tunisia, based on the shape of the empirical Cumulative Distribution Function (CDF) and similarities of physiographic and climatic characteristics. Viglione et al. (2007) made comparison among different homogeneity tests for regional frequency analysis. Pelczer et al. (2007) applied cluster analysis to achieve a regionalisation of the Sonora River Basin in the Sonora State, Mexico, into homogeneous zones. Two methods were compared, the hierarchical Ward's method and the non-hierarchical K-means method. By comparing results, it is showed that the K-means algorithm was an effective method. Plain et al. (2008) have proposed a spatially explicit seasonal forecasting model by using fuzzy spatiotemporal clustering of long-term daily rainfall and temperature data, which is found to be performing well.

Recently, a modern hydro-informatic tool, named as Self-Organization feature Map (SOFM) has been applied in several studies for clustering watersheds (Srinivas et al., 2008; Farsadnia et al., 2014; Mannan et al., 2018). However, this particular method is not alone sufficient to achieve reliable homogeneous regions. Recent studies show that application of fuzzy clustering approach has increased among the research community. Sadri and Burn (2011) adopted fuzzy c-means approach for regionalization. Satyanarayana and Srinivas (2011) have also done an experiment on regionalization of precipitation in data sparse areas based on fuzzy clustering approach. Chen et al. (2011) has done risk analysis of flood disaster in China, on the basis of Fuzzy Clustering Method. Dikbas et al. (2012) have also used fuzzy cluster method to classify precipitation series and to identify homogeneous regions. Chavoshi et al. (2013) has applied fuzzy expert system (FES) approach, optimized by genetic algorithm to perform hydrological regionalization. Asong et al. (2015) have applied combination of multivariate approaches such

as PCA, canonical correlation analysis and fuzzy C-means clustering. Bharath and Srinivas (2015) have proposed a new approach for regionalization using wavelet-based global fuzzy cluster analysis. Irwin et al. (2017) have used location and atmospheric attributes in FCM algorithm for delineation of homogeneous precipitation regions in two diverse Canadian climate regions. Wang et al. (2017) have done regional frequency analysis of precipitation extremes with the use of fuzzy c-means and L-moments approach, in Mainland China.

1.6 UNCERTAINTY ASSOCIATED WITH CLIMATE PROJECTION

1.6.1 Sources of Uncertainty

A unique definition of uncertainty is hard to find in the literature. It may be defined as a degree of ignorance, a state of incomplete knowledge, insufficient information or a departure from unattainable state of complete determinism. Uncertainty may be thought of as a measure of the incompleteness of one's knowledge or information about an unknown quantity to be measured or a situation to be forecast.

Various researchers have proposed their own kind of classification for uncertainty depending upon the purpose of their work. Broadly it can be classified into two types: inherent uncertainty, and epistemic uncertainty. Inherent uncertainty is the variability or randomness of the system in space and time and the epistemic uncertainty is due to the lack of knowledge of fundamental phenomena. While epistemic uncertainty is predictable and controllable, inherent uncertainty is unpredictable and uncontrollable (Faccioli et al., 2014).

Uncertainty can also be referred to as quantitative and qualitative. The former type refers to the uncertainty when it is represented using numerical values, typically by statistical inference. If uncertainty is represented using linguistic variables then it is referred to as qualitative uncertainty.

Sources of uncertainty related to water resources planning and precipitation forecasting can be classified as:

- Model uncertainty (due to assumption in model equations, model building and other forms of incompleteness in conceptualising the real system).
- Input uncertainty (due to imprecise forecasts of uncertainty in model inputs, such as future precipitation, temperature etc.).

- Parameter uncertainty (due to imperfect assessment of model parameters).
- Natural and operational uncertainty (due to erroneous and missing data, human errors and mistakes etc.)

The uncertainties mentioned above can be further classified based on the purpose of the study and requirements, as follows:

- Uncertainty due to GCM or RCM model (inter-modal uncertainty)
- Uncertainty due to different downscaling methods (downscaling uncertainty)
- Uncertainty due to hydrological model parameters (parameter uncertainty)
- Uncertainty due to different realization of the same GCM (intra-model uncertainty)

1.6.2 Uncertainty analysis in climate change projection

Hydrological forecasting systems are always dependent upon the dataset used, parameters considered as well as the hydrological models used for simulation and projection. However, studies show that all these quantities are associated with some kind of uncertainties. Physical processes like cloud formation, ocean water circulation, ice and snow albedo, water vapor formation and warming are complicated processes and hence difficult to model at a smaller scale. Therefore, use of downscaled data from GCM based simulations brings uncertainties in the future projections.

Detailed analysis of the associated uncertainties should be done for understanding the reliability of the results. Wilby (2005) has done an uncertainty analysis of water resources model parameters used for climate change impact assessment. It was recommended that routine sensitivity analyses should be done for climate-change impact assessments to quantify uncertainties related to parameter instability, identifiability and non-uniqueness, while using conceptual water balance models. Buytaert et al. (2009) analysed the impact of GCM model uncertainties on precipitation, temperature and evapotranspiration and assessed its propagation through hydroclimatic model to impact stream flow. They found that these uncertainties widely affected the future projection of stream flow, exhibiting diverse results for different GCM model. Najafi et al. (2011b) assessed the uncertainties associated with GCM models and hydrologic model selection in climate change impact studies and indicated their importance. Jung et al. (2012) have done an uncertainty assessment of climate change impacts for two hydrologically distinct river basins, one dominated by snowfall in winter and the other

dominated by rainfall in all seasons. The results indicated the need of more careful interpretation of runoff projections for snow-dominated regions. Chen et al. (2013) has evaluated the uncertainties related to empirical downscaling methods in quantifying climate change impacts on hydrology in two North American river basins. They emphasized on using multi-model climate projections and downscaling approaches for better assessment of uncertainty. Kumar et al. (2013) have presented a high-resolution multi-model projection for climate change over India and assessed the uncertainties associated with the approach. The results indicated that the RCMs simulated precipitation and temperature changes are largely influenced by the driving models and the impact of the GCM on projected precipitation change is as strong as the variability among the RCMs. Wang et al. (2013) have assessed the climate change impacts and associated uncertainty of extreme rainfall events in the Apalachicola River basin, Florida. They found that rainfall of the base period has been significantly overestimated by some models. Chen et al. (2014) have assessed the influence of climate change on mean annual rainfall in monsoon areas of Asia. It was found that, different basin responds differently to the same climate change scenario indicating uncertainty. Lespinas et al. (2014) have assessed uncertainties associated with regional climate models and hydrological models. They stated that uncertainties may arise due to the choice of RCM scenario, choice of hydrological model, choice of formula used for PE estimation, choice of calibration period considered and the dynamical parameterization of the hydrological model. Milne et al. (2014) have analysed uncertainties in the estimates of nitrous oxide and methane emissions in the UK's greenhouse gas inventory for agriculture, using Monte Carlo simulation. They suggested that firm guidance should be provided by IPCC on this issue, as several countries are moving towards inventory models with disaggregation. Najafi and Moradkhani (2015) have analysed the impacts of climate change on extreme values of runoff by applying multi-model ensemble analysis and assessed the uncertainties associated with the process. Results indicated that maximum temperature, minimum temperature and wind speed had less variations among different model simulations, whereas precipitation and runoff showed significant variations indicating high spatial variability within individual RCMs. Neupane et al. (2015) have studied the changes in hydrological projections due to climate change and deforestation. They found uncertainties in the climate variables, that might have occurred due to extreme topographical variations of the Himalayan region. Remesan and Holman (2015) evaluated the effect of meteorological data selection for baseline period and found existence of substantial influence of baseline data selection on climate change impacted future hydrological simulations. (Woldemeskel et al. (2016) have mentioned that there are three main sources of uncertainty associated with GCMs

in climate projections: model structure, the emission scenarios and natural variability. Joseph et al. (2018) have compared the parameter uncertainty (uncertainty arising from hydrological parameterization in climate change projections) with climate model uncertainty in hydrologic climate change impact assessments. It was found that, climate model uncertainties are significant in all seasons of a year and for all hydrological variables. Whereas, parameter uncertainty is almost negligible in comparison to huge climate model uncertainty. Shen et al. (2018) have mentioned that uncertainty associated with GCM models is the largest contributor of uncertainty in climate change impacted hydrological predictions. Soraisam et al. (2018) have assessed the uncertainties associated with different types of observed as well as reanalysed or downscaled datasets and their future projections. Gao et al. (2019) assessed the uncertainty associated with bias-corrected climate change projections and its influence on streamflow and watershed hydrology. The study exhibited that the streamflow sensitivity to selection of bias-correction of data is less than GCM selection, but more than sensitivity to RCM selection.

1.7 IDENTIFICATION OF THE PROBLEM

After observing the recent studies thoroughly, it has been established that the impact of climate change on water resources is a major issue throughout the world. More emphasis is given on various downscaling techniques that incorporates the use of Global Climate model (GCM). Application of fuzzy clustering has gained lot of interests from researchers as it helps in identification of hydrologically homogeneous climatic regions. However, there are a number of downscaling techniques that can be used. Furthermore, there are various types of climate models available for different emission scenarios/RCPs, provided by various organizations and climate centers. In addition to that, there are various hydrologic models available for hydrological analyses. The selection of appropriate hydrologic model, GCM as well as downscaling technique is not very clear from any of those works. They are basically dependent on the characteristics of the study area and hence varies from region to region. Also, there are a lot of uncertainties related to these model applications. Although some researchers have identified few approaches for selection of these models and techniques for a particular region, yet there is a need for an acceptable methodology for assessing a climate change impacted simulation model. Hence, if it becomes possible to delineate a region into some homogeneous clusters, in which same downscaling technique can be applied without having much variation, then it will be very useful in further studies related to water resources planning and management.

It was also noticed that, orographic disposition of a region affects the overall results. Few studies could be found that have incorporated the topography of the study area for analyses, yet enormous scope and need for study is there on projecting future impacts of climate change considering the effects of the orographic arrangements. Hence there is a need for developing methodologies for climate change impact studies that will take into account for the overall orographic arrangement, including locational parameters (latitude, longitude and elevation) and the complexity of the terrain.

1.8 OBJECTIVE AND SCOPE OF THE PROPOSED RESEARCH WORK

The primary goal of the study was to determine homogeneous hydroclimatic regions in northeast India and to analyse impacts of climate change on temperature and rainfall projections, with inclusion of orographic influence. Key objectives are defined as:

1. Delineation of the northeast Indian region into homogeneous clusters using fuzzy clustering approach.
2. Development of statistical downscaling models for predicting temperature and rainfall from GCM outputs and to compare the results of different models within a cluster.
3. Assessment of orographic effect on model simulations by incorporating orographic factor.
4. Comparison of different model simulations and assessment of associated uncertainties.

1.9 ORGANIZATION OF THE REPORT

Chapter 1 Begins with introduction to the research topic. GCMs and their usage are explained and need of downscaling and clustering in climate change studies are discussed. It also highlights the influence of orography in climate related studies.

Finally, the objectives of the study are defined.

Chapter 2 Presents the previous literatures related to climate change studies.

Chapter 3 Discusses the general methodology for delineation of a region into homogeneous clusters. Study area for the present work is described. Results

of fuzzy clustering are presented and comparison of the results are made with similar studies done previously.

Chapter 4 Discusses the methodology of downscaling for climate projections. Statistical downscaling models are developed for predicting temperature and rainfall from GCM outputs and results are compared for different models within a cluster.

Chapter 5 Assesses the orographic effect on model simulations. Topographical characteristics and wind flow pattern of the area is rigorously studied. Orographic factor is developed and incorporated in model simulation.

Chapter 6 Compares the different model simulations with and without inclusion of orographic factor. Also, associated uncertainties are assessed.

Chapter 7 Presents the summary, conclusion and future scope of the research work,

Chapter 2

LITERATURE REVIEW

This chapter begins with an outline of some of the literatures available on the importance of climate change impact studies and how it is affecting the water sector, with special emphasis on India. In the next section, regionalization, homogeneity tests and clustering approaches are reviewed to understand their applicability in the field of water resources. Application of various downscaling techniques in climate change impact analysis as well as uncertainty related to various sources that affects the impacts of climate change, which are reviewed in detail in the following sections. A critical appraisal of the reviewed literature has been presented, based on which the scope of the present research work has been defined.

2.1. IMPORTANCE OF CLIMATE CHANGE IMPACT STUDIES

For the past few decades, many researchers have attempted to present by means of their studies, that the climate is changing at an alarming rate and this issue needs to be addressed both at global as well as at regional level. Number of studies can be found which explain the importance of climate change knowledge in hydrology, water resources, agriculture, urban planning, ecosystem, land use change, environmental studies and so on. Some of these literatures are summarized below:

Eliasson (2000) stated that urbanization is inevitably related to climate change. In one hand urbanization influences climate change by affecting air quality and energy consumption, while on the other hand, change in the climate conditions causes more problems in urbanization. The study was mainly focused on finding out the importance of climate knowledge in urban planning process. It exposed that while there was an interest among urban planners to incorporate climate knowledge, unsystematic use of climatic information eventually resulted in low impact of the knowledge on urban planning. This was due to several constraints such as conceptual, technical, organisational, policy making etc. Hence, there is a need to address these constraints, if climate change is to be incorporated in urban planning.

Sharma and Shakya (2006) stated that the impacts of climate change on hydrological resources of a country should be assessed well in advance, before any long-term planning and water resources management program is carried out. They analysed the hydrological changes in

Bagmati watershed in Nepal. This watershed has the country's second largest hydroelectric plant in upper region and agricultural fertile land in the lower region. As the climate is changing, the trend of monsoon precipitation was showing degreasing tendency. They noticed a drift of hydrograph pattern also which affects water availability, eventually affecting the hydropower production, rice and wheat production. Furthermore, frequency and duration of flood exhibited increasing trend, though the overall magnitude of flood was decreasing. Water volume reduction in the downstream side towards India may kindle water conflicts between India and Nepal. It has been suggested that appropriate water policy should be formulated to address these changing scenarios.

Aguilera and Murillo (2009) explained the importance of determining the effect of climate change on natural ground water resources. They used a simple model to understand the impacts on natural water recharge to aquifers. The model termed as Estimation of Recharge in Over-exploited Aquifers (ERAS) simulates monthly water recharge to an aquifer with the use of basic data such as precipitation, temperature, storage coefficients and groundwater extraction. The results based on their study revealed a logarithmic decreasing trend in recharge of aquifers in the future. Furthermore, the percentage of effective precipitation and recharge to ground water storage with respect to actual precipitation is found to be progressively reducing.

Rahman et al., (2010) studied the impacts on morphological behaviour of the major river systems in Bangladesh, with the use of mathematical models MIKE 11 and MIKE 21. Existing morphological conditions were compared with predicted conditions. They stated that due to the effect of climate change, there is a possibility of rise in precipitation, temperature and sea level in near future.

Karamouz et al. (2011) stated that incorporation of climate change effects on urban water studies might help in achieving reliable results, that could be helpful in urban development planning. They proposed an algorithm for selection of the Best Management Practices (BMPs) to improve performance of urban drainage system, integrating anthropogenic and climate change effects. The algorithm was applied in Tehran metropolitan area. The results showed that there might be significant increase in high intensity rainfall events, eventually leading to more flash floods.

Estrela et al. (2012) did a detailed review on increase of hydrological stress as a consequence of climate change in Spain. Their study revealed that climate change will cause water shortages in the semi-arid zones, which will imbalance the resource-demand ratio. There might be 10%-

30% runoff reductions for the whole country throughout 21st century, impacting the water management sector. They specified the necessity of improved assessment and requirement of adaptation strategies in river basin management plans as well as in water related decision-making policies.

Hoyer and Chang (2014) assessed the impact of climate change and urbanization on freshwater ecosystem in Tualatin and Yamhill basins of Oregon. These basins are the fastest growing metropolitan areas of Oregon, where hotter and drier climate had been predicted as a result of climate change. The study revealed that estimates of water yield are highly sensitive to changes in the climate. Hence, they suggested inclusion of estimated changes under different climate conditions in land management decisions.

Runting et al. (2017) stated the importance of incorporating climate change and its uncertainty assessment in ecosystem services and decision making. They systematically reviewed literatures on climate change, which indicate significant impact on ecological services. They suggested that, for efficient allocation of resources for climate adaptation, effective management outcomes and decision-making, there is a need of integrated approaches with inclusion of multiple factors of change and sources of uncertainty.

Nkhonjera (2017) as well as Nkhonjera and Dinka (2017) did an extensive review on significant studies available on direct and indirect impact of the changing climate on groundwater resources in Africa. They considered the Olifants River basin as their area of interest. They stated that, since the number of studies available is inadequate, hence carrying out these kinds of climate change related scientific studies is essential to ensure quality and sustainability of water resources in Africa in the future.

2.2. CLIMATE CHANGE EFFECTS IN INDIA

Many research works and projects on climate change impact studies in Indian context are already going on. Some of the literatures are summarised below:

Mehrotra and Mehrotra (1995) have studied the climate change and hydrology with emphasis on the Indian subcontinent. They mentioned that, the major impacts of climate change in India would be on the hydrology, water resources and agriculture of the country. A sensitivity analysis was done by them on three basins situated in different agro-climatic zones, by using hypothetical scenarios for changes in temperature and precipitation. They stated that non-

availability of data and base record is the major restriction to draw a definite conclusion. Downscaling of GCM output to regional and sub-regional levels, by using suitable models for analysis, could be a solution for incorporating inter-annual variations in temporal and spatial scale. In case of water availability in storage reservoirs, they noticed significant impact of the changing climate. Hence, they suggested inclusion of impacts of climate variations in design and operation of water resources projects. Moreover, impact analysis of climate change on agriculture should include combined effects of CO_2 and other greenhouse gases, change in climatic conditions on regional scale, along with impacts on weeds, pests, diseases and other stresses, as suggested by them.

Gosain et al. (2006) have developed a hydrologic model for assessment and quantification of climate change impacts on various river basins of India. The model has been developed using Soil and Water Assessment Tool (SWAT) and HadRM2 daily weather data that determines spatial-temporal water availability in various river systems. The study indicated deterioration of the present conditions in terms of intensity of floods and severity of droughts in various parts of India. Two river basins predicted to be worst affected are assessed in detail: one affected by flood (Mahanadi) and the other affected by drought (Krishna). Based on the results they suggested that there is a need of several adaptation strategies ranging from change in land use, flood warning systems, cropping pattern to water conservation etc. Furthermore, there is a need of rigorous integrated analysis before taking any policy decision.

Mujumdar (2008) has given an overview of the Indian water resources scenario and assessments done on impact of climate change in various aspects of water resources. He stated that there is a need to understand the impact of global scale climate change on the local terrestrial hydrology. Furthermore, the changes in hydrologic extremes, irrigation demands, water availability, storm intensities, flood control and urban drainage design, saltwater intrusion etc. need to be addressed by utilizing the facilities of newly develops modelling and assessment tools. Also, to attain sustainable solutions for water resources management in the country, robust development in operating as well as development policies will be needed,

Kumar and Sarma (2010) conducted a case study in the Sonitpur district of Assam, India, to investigate the extent of climatic variation and scope of utilizing rainwater harvesting (RWH) system for irrigating small agricultural field. Analysis of climatic parameters has revealed the increase in dry spell, increase in temperature range, and decrease in total annual rainfall, which will affect the agriculture. Scope of irrigating agricultural field through RWH system has been

analysed considering cultivation of the most critical crops (maize, potato and tobacco) from water requirement point of view. It was suggested that successful implementation of RWH system might act as a beneficial source of irrigation and would help in meeting the agricultural water requirements of the region.

Deka and Sarma (2011) have carried out a study to quantify the climate change impacts on precipitation characteristics of Brahmaputra Basin. Two stations in the basin were selected based on their contrasting features and data availability: (1) Gerukamukh rainfall station, located on the upper region and (2) Rainfall station at Guwahati, situated at the lower region of the basin. A Statistical downscaling model (ASD-Automated Statistical Downscaling model) is developed by using the National Centres for Environmental Prediction (NCEP) reanalysis data, along with observed meteorological station data. The model was used to generate future scenarios for temperature and precipitation by utilizing large-scale predictor variables from Global Climate Model (GCM) outputs (HadCM3 daily weather data under A2 scenario). The results exhibited decrease in precipitation in early monsoon period, while precipitation in late monsoon was found to be increasing in near future. In case of temperature, there might be up to 1.8˚C rise in the maximum temperature averaged over month. It was experienced during the model development that selection of predictors plays a vital role in downscaling of GCM.

Sarma et al. (2012) under the network project taken by B. P. Chaliha Chair for Water Resources, carried out studies on the impact of climate change on water resources of Brahmaputra basin and adaptation strategies with regard to planning, design and management of water resources systems of the basin. The study indicated significant changes in temperature and rainfall pattern of this basin in future. High intensity rainfall for short duration and longer dry spells will deteriorate the flood and drought situation of the region. A climate change study has been carried out in Golaghat district which is considered as drought affected area. In this study, Dhansiri river (southern region) has been selected for the study. Statistical downscaling was done using Multiple Linear Regression. Predictor selection has been based on the correlation and Stepwise Regression. From the study, it was found that, the rainfall intensity (mm/day) will increase in future causing more erosion and flood. At the same time number of dry day will also increase, as such leading to drought situation.

Sarma et al. (2013) developed an optimization model for optimal Ecological Management Practice (EMP) to minimize impact of climate change and degradation of watershed due to

urbanization. This model was applied to a watershed of 0.17 km2 to explore its applicability in real condition. To determine the sediment yield and peak flow, RUSLE model and Rational Formula were applied respectively. Topographic parameters were extracted from DEM and other thematic layers prepared in GIS. Optimization tool box of MATLAB was utilized to formulate as well as solve the problem. The study provides an idea for enhancing sustainability of EMPs by utilizing the carbon sequestration potential of vegetation, which can be adopted by industries to earn carbon credits, thus compensating their greenhouse gas (GHG) emissions.

Dutta and Sarma (2020) has done a rigorous study on Brahmaputra river basin, to develop a hydrological model that can be used even in data-scarce regions of the basin. Brahmaputra basin, being a transboundary river basin and due to its complex hydroclimatic characteristics, imposes a lot of challenges on researchers. The hydro-climatological information available for the region is limited. To overcome this situation, various combinations of observed as well as generated datasets were utilized to develop a reliable SWAT model. The results showed that the observed weather stations data (IMD gridded data + China weather data) along with Climate Forecast System Reanalysis (CFSR) data provides acceptable hydrologic assessment for the Brahmaputra basin. Hence, this combination of dataset can be used for estimation of outflow at any desired outlet on the basin. Validation of the above statement is examined for discharge at Dhansiri river basin outlet in Gologhat, Assam. They have also mentioned that, this dataset can be utilized for GCM downscaling and climate change projection in the Brahmaputra basin, where data-scarcity creates huge uncertainty.

2.3. REGIONALIZATION, HOMOGENEITY TESTS AND CLUSTERING APPROACHES

Modelling of hydrological processes like precipitation, temperature etc. is very complex as most of the descriptive variables (soil, topography, land-use, rainfall) have different spatial and temporal variations. Therefore, to overcome the difficulties involved, regionalization of the region or area of study can be done. Regionalization is the process of finding out homogeneous groups of watersheds with similar hydrologic response, which helps in hydrologic design, planning and management of water resources systems. Literatures indicate that, in the past, political and geographical boundaries are used as a basis of forming homogeneous regions (Thomas and Benson, 1970; Chew and Heiler, 1974; NERC, 1975; Beable and McKerchar, 1982). However, use of political and geographical boundaries is found to be not very

convincing while forming hydrologically homogeneous regions. Even for two catchments which are found to be similar in hydrological responses, may only partly resemble to one another. Hence, it is not justifiable to assign those to only one group (Rao and Srinivas, 2006a). In recent years, various soft computing methods like clustering techniques are being used to model complex hydrological processes. These methods can efficiently simulate the real-life hydrological processes without prior knowledge of the exact relationship between variables (Chavoshi et al., 2013).

Many researchers have applied principal component analysis (PCA) for finding out hydrologically homogeneous regions. Bedi and Bindra (1980) used PCA for explaining the correlation among the monsoon rainfall over various parts of the Indian subcontinent. Bärring (1987) applied PCA, along with common factor analysis and spatial correlation, to define the spatial patterns of daily rainfall in central Kenya. Bonell and Sumner (1992) used S-mode PCA and Cluster Analysis to find out atmospheric circulation and daily precipitation pattern in Wales.

Kulkarni et al. (1992) have used map-to-map correlation method and k-means clustering method for classification of summer monsoon rainfall patterns over India, using 120-year period data (1871–1990). They used empirical orthogonal functions (EOFs), which is similar to principal components analysis except that the EOF method can find both spatial patterns and time series. Six distinct types were obtained from both EOF and k-means method. It was observed from the study that the monsoon rainfall for the country as a whole may have no relationship with the spatial patterns.

Guttman (1993) has used L-moment test in determination of regional precipitation climates in continental United States. He established a methodology or algorithm for defining regions of similar precipitation climates that are homogeneous with respect to the statistical distribution of annual precipitation. The algorithm consists of a two-step iterative procedure: the first step is clustering of sites based on seven variables on which precipitation depends, and the second step is the test of discordancy and homogeneity. SAS average linkage and Ward's minimum variance hierarchical clustering software are used for the initial physically based clustering. Then homogeneity of annual precipitation amounts within a region was evaluated by using L-moment techniques. The methodology resulted into 104 precipitation regions within the continental United States. Problems were encountered mainly in mountainous and arid regions, which were resolved in all but three regions by examining orography and/or the data.

Gadgil et al. (1993) delineated the Indian region into rainfall zones that are coherent with respect to the variations of summer monsoon rainfall. They found that for every pair of stations, trend of summer monsoon rainfall was positively correlated; and for each zone the mean inter-series correlation was high. An objective method was specifically developed for delineating the rainfall zones, using the inter-series correlation data. This method was suggested as appropriate for regionalization study of Indian region, based on rainfall variation of summer monsoon on interannual and larger scales.

Iyengar and Basak (1994) used principal component analysis (PCA) for regionalization of Indian monsoon rainfall. Four principal regions were first demarcated, having four different interannual variability signals, from which ten homogeneous sequential regions were formed. The PCA approach was recommended for further subdivision of the region.

Fill and Stedinger (1995) have mentioned about the importance of homogeneity test for regional flood frequency analysis. They performed homogeneity tests based upon Gumbel distribution and had done a critical appraisal of Dalrymple's test. The study analyzed the relative performance of Dalrymple's test, a normalized quantile test based on L-moment parameter estimation (X-10 test), and a method of moment Cv test (MOM-Cv test). Dalrymple's original test showed error and hence a corrected version was developed, whereas L-moment X-10 test was always found to be more powerful than the other two methods. It was recommended that, Dalrymple's original test should not be used for homogeneity test.

Singh and Singh (1996) have analysed the space-time variation and regionalization of seasonal as well as monthly summer monsoon rainfall of the sub-Himalayan region and Gangetic plains of India, by using principal component analysis (PCA). For the purpose, rainfall data for a period of 114 years (1871-1984) were used, from 90 well distributed stations. The 4 leading PCs that explained 52 to 61% of the variance for seasonal and monthly rainfall respectively were found statistically significant. When rotated by the varimax-rotated Principal component method (RPC), four distinct homogeneous rainfall areas were formed for both time scales.

Burn (1997) gave a regionalization approach for flood frequency analysis, using seasonality measures. The approach was employed to a set of catchments in the Canadian prairies. Comparison between seasonality-based region of influence (ROI) and ROI formed through the use of a GA search technique was done, from which seasonality-based ROI was found to be better. The approach was found to be effective in regionalization.

Adelekan (1998) examined the spatio-temporal variations in thunderstorm rainfall over Nigeria, using daily rainfall data. They illustrated that orography and moisture content of the region play an important role in the occurrence of thunderstorms. By using elementary linkage analysis, 19 stations were grouped into six thunderstorm rainfall regions, which showed internal coherence in terms of temporal fluctuations of thunderstorm rainfall.

Hall and Minns (1999) have done hydrologically homogeneous clustering for flood data of the southwest of England and Wales, using Kohonen Network method and Fuzzy c-means method. They illustrated that homogeneous regions may be better defined by Representative Regional Catchments (RRCs), which have characteristics that are hydrologically more appealing than the regions formed by geographical boundaries. Both Kohonen network and fuzzy c-means method were found to be effective in identifying similar RRCs and hence recommended feasible for applying on a country-wide basis homogeneous clustering.

Wotling et al. (2000) have done regionalization of extreme precipitation intensities in the volcanic island of Tahiti using principal components of the topographical environment. They focused on how the method automatically takes into account the topographical relief features. Principal component analysis of a digital elevation model supplies a limited set of variables describing the topographical environment, called as descriptors, which are linked to the parameters of the rainfall intensity Gumbel distribution by using a stepwise regression adjusted on 20 point-rainfall records. The model is then applied on a regular 300-points grid node, and interpolated using a spline function. The relationship between the rainfall parameter distributions and the topographical descriptors is found to be very strong, and the method supplies a direct estimation, in space, of the rainfall statistics through the regression model. Validation of the results and comparison with simple kriging interpolation show the relevance of the approach. However, more data are needed for a better confidence in the parameter estimation in some areas of the island. They finally concluded that, this automatic and objective method could be applied in any mountainous area, where topography has a major influence on the precipitation features.

Dinpashoh et al. (2004) have given an approach of variable selection using multivariate methods, for regionalization of Iran's precipitation climate. Initially, those variables were selected which represent most of the information from all candidate variables, using Procrustes Analysis. The selected variables were then used in factor analysis and clustering techniques. 77 weather station datasets (1956-1998) were used in the study. The H and Z-statistics, based

on L-moment technique, were used for the homogeneity test of each region. Seven distinct regions were formed by the approach, out of which six regions were homogeneous and one was found to be heterogeneous.

Goktepe et al. (2005) used hard k-means and fuzzy c-means algorithms for clustering of fine-grained soils in terms of plasticity index and shear strength, for Antalya region in Turkey. Laboratory tests were performed on 120 undisturbed soil samples, collected from the region. 20 fine grained soil samples were also collected from other regions of Turkey, to determine whether the clustering analysis has generalization capability or not. Fuzzy c-means algorithm was found to be better than hard k-means method in the study, and was suggested for classification of fine-grained soils based on shear strength and plasticity index. Moreover, it showed the capability of handling uncertainty related to the soil parameters.

Rao and Srinivas (2006a) have done regionalization of watersheds in Indiana, USA by hybrid-cluster analysis, that use agglomerative hierarchical clustering algorithms for clustering and partitional clustering procedure (the K-means algorithm) for refining the identified clusters. Three hybrid clustering algorithms are considered to determine their efficiency in regionalization, those are single linkage, complete linkage and Ward's algorithms. To evaluate the effectiveness of the partitions obtained from the cluster algorithms, four cluster validity indices are used viz. cophenetic correlation coefficient, average silhouette width, Dunn's index and Davies–Bouldin index. The hybrid of Ward's algorithm and K-means algorithm is found to be effective in minimizing the effort needed for identification of homogeneous regions for flood frequency analysis.

Rao and Srinivas (2006b) have also applied fuzzy cluster analysis for regionalization of the watersheds in Indiana, USA, by using annual maximum flow data. When hard clustering algorithms were used for regionalization, considerable effort was needed to adjust the clusters so that they become homogeneous. Fuzzy clustering allows a catchment to have partial memberships to one or more regions. Hence, it reduced their effort in adjusting the clusters while improving their homogeneity. The analysis yielded seven acceptably homogeneous hydrologic regions in the Indiana region.

Owen et al. (2006) have used principal component analysis (PCA) and cluster analysis for urban land classification in the UK West Midlands. An area covering 900 km^2 of the UK West Midland metropolitan region was chosen for the study. PCA was done to reduce the

dimensionality of the input land-cover data, having a large number of attributes (e.g. urban, villages, open land, motorway, etc.) and the urban classification were derived by cluster analysis. The output of the study divided the area into eight urban land-cover classes of 1 km^2 pixel size having similar urban land morphology. Clustering analysis using agglomeration technique on the extracted components, is associated with a lot of uncertainties that were investigated in detail using fuzzy-type analyses. They stated that, the method can be readily adapted to other metropolitan areas where basic urban features (e.g. roads, housing density) are gridded and can become useful for environmental and ecological surveys.

Abida and Ellouze (2006) have discussed an innovative method for hydrological delineation of homogeneous regions in Tunisia, based on the shape of the empirical Cumulative Distribution Function (CDF) and similarities of physiographic and climatic characteristics. This resulted in dividing Tunisia into 10 homogeneous regions. Based on annual flood data of 49 hydrometric stations, with record lengths ranging from 10 to 100 years, spatial trends of the shape of the CDF were employed to identify mixed processes in the flood series. This resulted in dividing the study area into mixture and non-mixture zones. Hydrologic homogeneity of the delineated regions was tested by L (Linear) moment-based statistical tests. The exercise resulted in dividing Tunisia into 10 homogeneous regions, where mixed flood processes haven't been recognized.

Venkatesh and Jose (2007) have done a study to identify homogeneous rainfall regimes in parts of Western Ghats region of Karnataka. The study area covers one coastal district and its adjacent areas in the state of Karnataka. To evaluate the rainfall variation pattern, several bands were constructed across the ghats parallel to the latitudes and the statistical analyses (cluster analysis and analysis of variation) were conducted. For this purpose, they used mean annual rainfall data from 93 rain gauge stations for a period of 10–50 years. The study gave three distinct rainfall zones: Coastal zone, Transition zone and Malanad zone. Also, it revealed that maximum rainfall occurs on the windward side of the ghats ahead of the geographical peak. The study was found to be quite valid for the study area.

Viglione et al. (2007) made comparison among different homogeneity tests for regional frequency analysis. Homogeneity tests based on L moments ratios and two other rank tests were used for comparison and their performance were assessed by using Monte Carlo simulation. The evaluation was also done by varying the number of sites in the regions, length of the series, type of parent distributions as well as degree of heterogeneity. They suggested

that moments-based tests perform better for slightly skewed samples, while in case of high skewness the rank tests are better.

Pelczer et al. (2007) applied cluster analysis to achieve a regionalisation of the Sonora River Basin in the Sonora State, Mexico, into homogeneous zones. Two methods were compared, the hierarchical Ward's method and the non-hierarchical K-means method. By comparing results, it is showed that the K-means algorithm was an effective method to achieve climatically homogeneous zones in the Sonora River Basin. It was also observed that working with average values could mask maximum and minimum values that can influence the climatic variability. It is also important to know the conditions at the surface, to find out if the regionalisation obtained is reliable.

Plain et al. (2008) have proposed a spatially explicit seasonal forecasting model to create climate memberships over time and location, by the use of fuzzy spatiotemporal clustering of long-term (40 years) daily rainfall and temperature data. The model was applied on the data obtained from weather stations of south-east Australia, covering subtropical to arid climate zones. Hence, the model includes both local climate information and continental drivers. The 6-month final forecast of rainfall and maximum average temperature produced relative errors of 0.56 and 0.89, with Pearson correlation coefficients of 0.82 and 0.83, respectively. This indicated reasonable accuracy of the predictions from the proposed model.

Recently, a modern hydro-informatic tool, named as Self-Organization feature Map (SOFM) has been applied in several studies for clustering watersheds. However, this particular method is not alone sufficient to achieve reliable homogeneous regions. Srinivas et al. (2008) have done regional flood frequency analysis (FFA) for watersheds in Indiana, USA, by combining fuzzy clustering with self-organizing feature map. They demonstrated a two-level SOFM-based clustering approach, in which the SOFM is used to form a two-dimensional feature map at first level and the outputs of SOFM are clustered by the use of Fuzzy c-means algorithm in the second level. They found that SOFMs serve as a useful precursor to clustering algorithms, when used for regionalization of watersheds. Performance of the proposed approach is compared with the results from other methods based on regression analysis and canonical correlation analysis, and it was found better in estimating flood quantiles at ungauged sites.

Sadri and Burn (2011) have adopted a Fuzzy C-Means approach for regionalization, using bivariate homogeneity and discordancy approach. The regions are formed based on bivariate

criteria instead of univariate ones; the two variables being severity and duration. This approach is applied to the records of 36 unregulated flow monitoring sites in Canadian prairie provinces of Alberta, Saskatchewan and Manitoba. Instead of traditional univariate L-moment statistics-based tests for homogeneity and discordancy, they used multivariate L-moments i.e. L-comoments. The fuzzy clustering approach, called Fuzzy C-Means (FCM) was used to form the initial clusters and adjustments were done to form final clusters so as to meet the criteria of homogeneity, lack of discordancy, and sufficient size. Also return periods were estimated using a bivariate copula method.

Satyanarayana and Srinivas (2011) have done an experiment on regionalization of precipitation in data sparse areas based on fuzzy clustering approach. They mentioned that, in ungauged areas as well as in areas having sparse rain gauge density, conventional method of using direct precipitation data may not be useful. Hence, they used large scale atmospheric variables (LSAV), that influence the precipitation of the study area, instead of conventional data. Furthermore, this approach has the ability to form clusters in ungauged areas where no at-site precipitation data are available, due to the attributes that can be reliably estimated. The location parameters (latitude, longitude and altitude), LSAVs and seasonal precipitation data are used as features for regionalization. The approach was applied to delineate homogeneous annual rainfall regions in India, and results were compared with those obtained using rainfall statistics, regionalization based on hard cluster analysis, and meteorological sub-divisions in India.

Chen et al. (2011) has done risk analysis of flood disaster of 30 provinces in China in 2008, on the basis of Fuzzy Clustering Method. To carry out the test, four indexes namely affected area, housing collapse, number of deaths and direct economic losses were selected. The fuzzy equivalent matrix was extracted through transitive closure method and the cluster graph was obtained, which shows the classification of severity of floods. They found that fuzzy clustering analysis method is appropriate for grade division of flood disaster.

Dikbas et al. (2012) have used fuzzy cluster method (Fuzzy C-Means: FCM) to classify precipitation series and to identify hydrologically homogeneous regions. Total precipitation data of stations are used for the study, operated by National Meteorology Works (DMI) in Turkish basins. Based on a number of performance evaluation indexes, the optimal cluster number was determined as six. The identified clusters are then checked for homogeneity by applying regional homogeneity tests based on L-moments and found to be sufficiently

homogeneous. Hence, they recommended FCM method classifying precipitation series and for identifying hydrologically homogenous regions.

Mok et al. (2012) have suggested a robust adaptive clustering analysis method that automatically identifies the desired number of clusters, producing reliable clustering solutions. The proposed method first attains several clustering results from a specific algorithm (such as Fuzzy c-means or k-means) and then integrates them as a judgement matrix. To get the final results with desired cluster numbers, an iterative graph-partitioning process was implemented. Cluster validity analysis, spectral clustering and cluster ensemble methods were used to find out the applicability of the proposed method.

Chavoshi et al. (2013) has applied fuzzy expert system (FES) approach, optimized by genetic algorithm to perform hydrological regionalization in the southern strip of the Caspian Sea basin, north of Iran, that has been affected by severe floods in the recent past. 61 hydrometric stations and 31 weather stations having 44 years of record data (1961–2005) are used in the study. The FES approach incorporates physical characteristics, climatic characteristics, flood seasonality and geographic location and is used to investigate the homogeneity of the catchments. Genetic algorithm (GA) was then utilized to adjust parameters of FES and to optimize the system. This approach was found better than the conventional hydrologic regionalization methods possibly because it utilizes larger number of homogeneity parameters and produced lower heterogeneity.

Saikranthi et al. (2013) used correlation analysis for identifying homogeneous rainfall regions in India, on the basis of seasonal [southwest monsoon (SWM) and northeast monsoon (NEM)] and annual rainfall. They used 51 years (1951–2001) daily rainfall data obtained for 1025 rain gauges across the country for the analysis, which produced 26 homogeneous rainfall zones. Results were compared with previous studies, which exhibited better performance of the correlation analysis than other methods. The zones formed were also found to be homogeneous in terms of amount, frequency, and type of rainfall. However, the degree of homogeneity seemed to vary both in terms of amount of rainfall and the temporal scale, indicating plausible changes in the homogeneity of regions, in future.

Farsadnia et al. (2014) have applied the Self-Organization feature Map (SOFM) technique to cluster the Mazandaran's province watersheds in the north of Iran. To identify the clusters, two-level SOFM and three clustering methods (fuzzy c-mean, K-mean, and Ward's

Agglomerative hierarchical clustering) are used. The SOFM is used to form a two-dimensional feature map, whose output nodes are then clustered by means of unified distance matrix algorithm and three clustering methods to form the clusters for flood frequency analysis. The combination of SOFM and Ward was found to be better than the combination of either SOFM and K-mean or SOFM and FCM.

Darand and Mansouri Daneshvar (2014) used principal component analysis (PCA) and hierarchical clustering analysis (HCA) for regionalization of precipitation in Iran. Daily gridded APHRODITE precipitation data (1951-2007) were used for the analysis. The S–mode of PCA was employed on seasonal-based precipitation correlation matrix with eight variables, and varimax rotation was utilized to improve interpretability of PCA scores. Then, HCA was applied, which exhibited nine distinct and homogenous regions. The Kolmogorov–Smirnov test helped in identifying the independent regimes, on the basis of seasonal percentage of precipitation distribution. Based on the results it was recommended that the method can be effectively used in arid and semi–arid regions of mid–latitudes, especially in Middle East countries.

Weiss et al. (2014) have used regional frequency analysis for homogeneous clustering of extreme significant wave heights. They proposed a physically based method to form homogeneous regions by detecting typical storms footprints, and applied it to delineate the North-East part of the Atlantic Ocean. To identify storms that generates marine extremes, a spatiotemporal procedure of de-clustering was applied. A clustering algorithm was then employed, that relies on storm propagation, to detect the typical storms footprints. The method proposed was found to be simple and effective, as it requires only the information about time of occurrence of the observed extremes. Six physically and statistically homogeneous regions were found in the study area.

Asong et al. (2015) have proposed an approach for regionalization of precipitation characteristics in the Canadian Prairie Provinces incorporating large-scale atmospheric covariates, teleconnection indices and geophysical attributes. This approach employed combination of multivariate approaches such as principal component analysis, canonical correlation analysis and fuzzy C-means clustering. Five homogeneous regions were formed after the analysis, which were validated independently using statistics of monthly and seasonal precipitation totals, and seasonal extremes.

Bharath and Srinivas (2015) have proposed a new approach for delineation of homogeneous hydrometeorological regions in India using wavelet-based global fuzzy cluster analysis, instead of conventional principal component analysis. This approach extracts information from wavelet transformations of the observed multivariate hydrometeorological time series, and utilizes the same for regionalization by global fuzzy c-means cluster analysis. They stated that, the approach can also explain the dynamic variability in time series and its non-stationarity. Results were used for formation of Drought severity-area-frequency (SAF) curves for regional drought analysis.

Agarwal et al. (2016) proposed a wavelet-based multiscale entropy method for regionalization. The method uses wavelet transform and k-means based hybrid approach for clustering of hydrologically homogeneous regions. Streamflow data from 1951 to 2002 at 530 selected catchments across the United States are utilized to assess the performance of the proposed method. The homogeneity tests performed after the analysis showed that the proposed method can be effectively used in regionalization.

Irwin et al. (2017) have used location and atmospheric attributes in FCM algorithm for delineation of homogeneous precipitation regions in two diverse Canadian climate regions: The Prairie and the Great Lakes–St Lawrence lowlands regions. The attributes selected are location parameters (latitude, longitude), site elevation, distance to major water bodies and atmospheric variables at different pressure levels. The ability of the attributes was assessed in terms of computational efficiency and regional homogeneity of precipitation statistics. Based on the outcome, it was recommended to use location parameters and distance to water bodies for regionalization of precipitation and to consider atmospheric variables for future climate change applications.

Komalasari et al. (2017) have used descriptive statistics and cluster analysis to explain regional pattern of extreme rainfall in Java Island. The analysis was performed to obtain centralization, variation and distribution of the Maximum daily rainfall data (1983-2012). The statistics used to describe the above information are mean, median, standard deviation, skewness, kurtosis and Inter Quartile Range (IQR). For clustering, squared Euclidean distance and ward method were applied. The analysis produced four clusters in the study region.

Kulkarni (2017) has used probability density function to divide the Indian subcontinent into homogeneous clusters, using daily summer monsoon (June-September) rainfall at 357 grids of

100 km × 100 km size. The analysis produced five clusters, out of which only one cluster covered contiguous region and all other clusters were dispersed indicating the irregular behaviour of daily rainfall pattern over India. In order to assess the possible changes in clustering pattern in recent period, the study was done by using two time periods 1901–1975 and 1976–2010. The resulting clusters were found to be extremely different in the two time periods, indicating more erroneous distribution of daily rainfall in the recent period.

Wang et al. (2017) have done regional frequency analysis of precipitation extremes with the use of fuzzy c-means and L-moments approach, in Mainland China. They used a high-resolution (0.50 × 0.50) daily precipitation dataset (1961-2013) to estimate spatial-temporal patterns of precipitation extremes. The results divided the region into 50 homogeneous regions based on location and mean annual precipitation characteristics. The goodness-of-fit (GOF) test results showed that Generalized Extreme Value (GEV), Generalized Normal (GNO) and Pearson type III (PE3) distributions fit well for most of the regions. They stated that the study results can be utilized for further study in Mainland China related to climate change as well as for prediction of extreme precipitation, flood prevention, control and management.

Mannan et al. (2018) have used climatic variables and self-organizing maps (SOM) to regionalize India. An artificial neural network algorithm is used to derive the clustering technique, along with four cluster validity indices to identify the optimal number of clusters. For assessing the technique gridded rainfall dataset (0.25° × 0.25°) from Indian Meteorological Department (IMD) for 34 years duration (1980–2013) was used. Climatic variables like specific humidity, geo-potential height, air temperature, surface pressure, etc., were derived from the Modern Era-Retrospective Analysis for Research and Applications reanalysis data-set. 10 homogeneous regions were formed when only rainfall data was used, wheras incorporation of climatic variables divided the region into 15 regions. All of the 15 delineated regions were found to be homogeneous when subjected to homogeneity test, and were capable of effectively capturing the spatial variability of rainfall.

Zaifoglu et al. (2018) have done regional frequency analysis using L-moments with time series clustering approaches, using dynamic data sets in Northern Cyprus. Also, comparisons were made among various clustering approaches such as shape-based, feature-based, and model-based approach. The cluster analysis gave five homogeneous regions, and as per the goodness-of-fit test, generalized logistic, generalized normal distributions and Pearson Type III exhibited

the best fit. Monte Carlo simulation was used to evaluate the accuracy of the estimated quantiles for different return periods, which exhibited spatial consistency.

2.4. DOWNSCALING APPROACHES

Now-a-days, General Circulation Models (GCMs) are widely used in climate change studies; but GCMs are relatively coarse in resolution and therefore unable to resolve significant sub-grid scale features, including topography and land use, as needed in hydrologic modelling and impact assessment. Therefore, some form of downscaling is required to produce scenario of finer spatial resolution than currently delivered by raw GCM output (Najafi *et al*., 2011a). Some literatures available on downscaling are discussed below:

Heyen et al. (1996) have done statistical downscaling of monthly mean air pressure to sea level anomalies. They developed a relationship between the anomalies of North Atlantic air pressure and the locally influenced sea level in the Baltic sea region, in winter. The results have shown good correlation during validation and the model was able to effectively reproduce decadal variations of sea level. However, the model was purely statistical, without any consideration of the plausible physical links between the variables.

Wilby et al., (1998a) have used statistical downscaling for simulation of hydrometeorological variables. They developed empirical relationships between daily climatic variables using GCM, for a catchment in Japan. It was found from their study that, NCEP reanalysis datasets have dissimilarity with GCM datasets (HadCM2SUL). Hence, a relation between the two datasets was made and the climatic variables have been predicted for future (2080-2099). Due to high resolution of the datasets, the results have shown large percentage of error.

Later, Wilby et al., (1998b) made a comparison of different statistical downscaling methods using GCM output. The downscaling methods used are WGEN, B-Circ, C-Circ, , SPEL and ANN1 and ANN2, and the GCM model used was HadCM2. From the analysis, substantial variations were found among the downscaling methods. The WGEN method yielded minimum variations between observed and simulated daily precipitation. On the other hand, ANN methods showed poor performance due to inadequate simulation of wet-day occurrence statistics.

Zorita and von Storch (1999) suggested a simple analog method for statistical downscaling and compared the same with other available complex methods. This method associates the large-

scale GCM simulations with the local variables observed simultaneously with the most similar large-scale circulation pattern in historical observations. Compared with other statistical methods like canonical correlation analysis, classification method, linear method, weather generators and neural network, the analog method was found to be the simplest one with results similar to the linear method.

Misra et al. (2003) have simulated seasonal values for a number of climatic variables using dynamic downscaling. Multiple atmospheric global circulation models (AGCM) have been used to develop the models for Amazon river basin, in South America. The Regional Spectral Model (RSM) simulations were produced at 80 km resolution for summer season (January-February-March). Comparison between the two models showed that RSM performs better than AGCM.

Dibike and Coulibaly (2005) have made comparisons of hydrologic as well as downscaling models under climate change impacts. Future variations of river flow and reservoir inflow have been predicted Chute-de Diable sub-basin in the Saguenay watershed, using downscaled data from LARS-WG (Weather Generator) and SDSM (regression-based method). Both daily precipitation and temperature showed increasing trend, although the downscaled data from the two models were not similar. The watershed analysis results exhibited increase in future river flow and reservoir inflow under the impacts of climate change.

Mehrotra and Sharma (2005) developed a nonparametric nonhomogeneous K-nearest-neighbour (KNN) based hidden Markov model, that was applied for spatial downscaling of multisite daily rainfall occurrences. Four atmospheric circulation variables were used for the study, over a 30-rain gauge-network near Sydney, Australia. The model developed can generate rainfall occurrences conditional to a continuous weather state and mean rainfall over previous day. It was further compared with the weather state–based parametric nonhomogeneous hidden Markov model (NHMM) and a standard KNN model. Results showed that the adopted continuous weather state in the proposed model successfully captured the daily rainfall characteristics, in comparison to discrete state NHMM.

Zhang (2005) have done statistical downscaling of GCM output to predict soil erosion and assessed crop production by using water erosion prediction project (WEPP) model. HadCM3 model was used to project monthly precipitation and temperature for 1900–1999 and 2070–2099, under the GGa emissions scenario. The results indicated that the proposed downscaling

technique could effectively assess the impacts of climate change on site-specific soil erosion and crop production, owing to its simplicity and soundness.

Ghose and Mujumdar (2006) developed a methodology to forecast future rainfall scenario over Orissa using GCM projections by statistical downscaling. The model developed was a linear regression model that used GCM outputs of mean sea-level pressure and geopotential height as explanatory variables. Fuzzy clustering technique was used to classify the PCs identified by the Principal Component Analysis (PCA). The fuzzy membership values were then used in the regression model with an assumption that different clusters have different effects of circulation patterns on precipitation, and later modified with an appropriate seasonality term. The proposed methodology was found to be computationally simple and it could model rainfall with a high R^2 value. The methodology was used to forecast monthly rainfall over Orissa, that showed possibility of increase in hydrologic extremes in future.

Mehrotra and Sharma (2006) have developed a framework of nonparametric stochastic downscaling for multisite rainfall projection. Station downscaling of rainfall was performed based on nonparametric nonhomogeneous hidden Markov model (NNHMM) that represents spatial variation of wetness fraction of rainfall occurrences. The framework was applied at 30 rain gauge stations around Sydney, Australia, using GCM data. The result exhibited that when separate treatment was provided for individual stations for rainfall amounts and occurrences, accuracy of the representation of characteristics (like rainfall spell patterns, spatial distribution of rainfall, low and high rainfall extremes at individual station, water balance and variability etc) has been enhanced.

Kang et al. (2007) have done statistical downscaling using six dynamic models from APEC Climate Center Multi-model Ensemble (MME) prediction system, in the Philippines and Thailand. Correlation analysis and singular value decomposition analysis were used to attain the atmosphere dynamic linkage, that provided a robust basis for the choice of predictor and its range in predicted fields. Cross validation of downscaled MME predictions with observed precipitation values has been done. The results indicated improvement in the prediction skill for most of the stations.

Serrat-Capdevila et al. (2007) have done statistical downscaling of GCM models to quantify the climate change impacts on hydrology of a riparian systems in the San Pedro basin, Arizona/Sonara. 17 GCMs under 4 different climate scenarios have been used in the study to achieve highest likelihood mean estimate and to quantify the uncertainty associated with the

results. The results indicated decrease in the ground water recharge in the basin, affecting the dynamics of the riparian area in future.

Anandhi et al. (2008) have presented a methodology for precipitation downscaling to river basin with the use of support vector machine, in Indian context. Predictor variable were extracted from NCEP reanalysis data, Canadian general circulation model (CGCM3) for emission scenarios A1B, A2, B1 and COMMIT for 1971-2100. K means clustering technique was used for seasonal stratification of past wet and dry season (using NCEP data), whereas nearest neighbour method was applied for the future wet and dry season. Both resulted into two clusters. A climatically sensitive region, the Malaprabha reservoir catchment is used as the study area. The downscaling results indicated increase in rainfall in the future for all scenarios, which highest projection for A2 and lowest projection for COMMIT.

Hessami et al. (2008) have done statistical downscaling of temperature and precipitation using automated statistical downscaling (ASD). The regression-based approach was inspired by SDSM model, developed by Wilby et al. (2002), and was based on backward stepwise regression. NCEP reanalysis data has been used for calibration and two GCM models, HadCM3 and CGCM1, were used for future projection. The results indicated that ASD performs better than SDSM for temperature prediction. However, for precipitation neither of the methods gave good result.

Mujumdar and Ghosh (2008) have presented a methodology for downscaling based on fuzzy clustering and Relevance Vector Machine (RVM). The methodology was applied to project monsoon streamflow from three GCMs under two emission scenarios, in the Mahanadi river, India. NCEP/NCAR reanalysis data was used to train the model. The results indicated decrease in streamflow due to increase in temperature. Limitation of the study was that they assumed stationarity condition for future in case of land use pattern.

Tisseuil et al. (2010) have done statistical downscaling of river flows using streamflow data of 51 stations located in southwest of France. Four statistical downscaling models were compared, namely Generalized Linear Model (GLM), Generalized Additive Model (GAM), Artificial Neural Network (ANN) and Aggregated Boosted Trees (ABT), for projected stream flow data. Among these models, ABT was found to perform better. GCM model predictions (under A2 and A1B scenarios) indicated decrease in stream flow rate in the future for both pluvial and nival regimes.

Kannan and Ghosh (2011) attempted to address the limitation of statistical downscaling in capturing correlation between multiple sites by representing the rainfall pattern in a river basin using rainfall state. The methodology proposed was tested for the Mahanadi river basin in India. The rainfall states were generated from GCM data by using k-means clustering technique coupled with a supervised data classification technique, namely Classification and Regression Tree (CART). Various cluster validity measures were applied to get the optimum number of clusters in the observed rainfall data. The results indicated possible increase in the occurrences of 'almost dry' and 'high' rainfall states and a decrease in the occurrences of 'medium' rainfall states.

Najafi et al. (2011a) have done statistical downscaling of precipitation employing independent component analysis (ICA) for predictor selection that determines spatially independent GCM variables. These climate variables and available precipitation data of the upper Willamette basin are used to calibrate three downscaling models: Multilinear Regression (MLR), Support Vector Machine (SVM) and Adaptive-Network-based Fuzzy Inference System (ANFIS). They found that the procedure given by them is successful in choosing the predictors for downscaling the GCM data both in monthly and seasonal timescales. MLR is found to be an efficient method for precipitation downscaling. By doing uncertainty analysis of the downscaling methods, they found that model structures and input predictors are important factors in the downscaling process and the main sources of uncertainties.

Vinnarasi (2012) in her project work, has focused on simulating the impact of climate change with the use of downscaling methods, on the precipitation characteristics and stream flow behaviour of Dhansiri River, a southern tributary of Brahmaputra basin. Downscaling has been done in two ways: statistical downscaling using regression analysis and Artificial Neural Network (ANN). The downscaling model uses CGCM3, HadCM3, and MRCGCM2 monthly weather data under A2 scenario. The downscaling result has been used to predict the future rainfall intensity. The result shows a 20 % increase in the average annual precipitation by 2100, with an indication that rainfall in the future during early months of present monsoon season would substantially decrease, whereas the rainfall is likely to increase in later part of the year.

Salvi et al. (2013) have projected daily rainfall of multiple sites in India using statistical downscaling. The projection has been done at a resolution of 0.5^0 over the whole Indian continent using Classification and Regression Tree (CART) and nonparametric kernel regression technique. GCM developed by the Canadian Centre for Climate Modeling and

Analysis is used for this study. To assess the uncertainty, five ensemble runs were performed. The model appears to effectively capture individual station means, the spatial patterns of the standard deviations, and the cross correlation between station rainfalls. The downscaling model seems to capture the orographic effect on rainfall in mountainous areas of the Western Ghats and northeast India.

Forsythe et al. (2014) have applied stochastic weather generator to assess climate change impacts in a semi-arid region. A stochastic rainfall model (RainSim) combined with a rainfall conditioned weather generator (CRU WG) have been validated with observed data and applied to upper Indus basin. Validation results have shown good performance under complex climatological behaviour of mountainous regions. Future projections showed year-round increases in precipitation (with increased intensity in the wettest months) and year-round increases in mean temperature. Relevant indices from the European Climate Assessment (ECA) were used to assess the productivity of natural resource-dependent systems, that indicated potential risk to water resources and local agriculture in the future. They stated that sophisticated downscaling methods were needed to evaluate the climate change impacts and the interannual variabilities for better understanding of the future conditions.

Fatichi et al. (2015) have done a high-resolution fully distributed hydrological analysis and future climate projections at the upper Rhone basin in the Alpine catchment. To downscale, they used a methodology that accounts for the uncertainty related to stochastic variability of climate variables like temperature and precipitation. The effects of climate change on propagation of streamflow from high elevation to the river in the valley has been analysed by applying various hydrological metrices at different temporal scales across 297 control sections. They found that there is an elevation dependence of climate change impacts in the catchment. Severe reduction in streamflow causing a dampened effect downstream, reduced ice cover and ice melt affecting hydropower production were observed from the results. However, no major changes in total runoff were observed for the next four decades.

Shamir et al. (2015) have analysed climate change projections of precipitation in the upper Santa Cruz river. Eight dynamically downscaled GCM models were used to assess the climate change impacts and its effect on regional water resources management. The climate analysis results were compared with GCMs and bias corrected statistically downscaled CMIP3 and CMIP5 projections, that are readily available for the contiguous U.S. A series of hydrologic models were developed to estimate the groundwater recharge and change in groundwater

storage. They stated that climate change projections increases the uncertainty and aggravates the complexity of water resources management task.

Moursi et al. (2017) have done a probabilistic assessment of climate change impacts on agricultural water scarcity in a semi-arid snowmelt dominated basin, the Sevier River Basin of Utah. 31 general circulation models (GCMs) were used to project and assess the vulnerability associated with water scarcity for the period 2000–2099. Results indicated that off-season precipitation affects water scarcity the most, followed by precipitation and temperature in the growing season. Although temperature and precipitation projections have indicated increase in water availability, a considerable risk probability of agricultural water scarcity was observed in 2025 through 2049 (RCP4.5).

2.5. UNCERTAINTY ANALYSES RELATED TO CLIMATE CHANGE IMPACTS

Hydrological forecasting systems are always dependent upon the dataset used, parameters considered as well as the hydrological models used for simulation and projection. However, studies show that all these quantities are associated with some kind of uncertainties. Physical processes like cloud formation, ocean water circulation, ice and snow albedo, water vapor formation and warming are complicated processes and hence difficult to model at a smaller scale. Therefore, use of downscaled data from GCM based simulations brings uncertainties in the future projections. Some literatures available on uncertainty assessment related to climate change are presented below:

Wilby (2005) has done an uncertainty analysis of water resources model parameters used for climate change impact assessment. He mentioned that although lumped conceptual models are used widely for climate-change impact assessments in spite of their limitations, it is important to understand the relative magnitude of uncertainties in water resource projections arising from the choice of model calibration period, model structure, and non-uniqueness of model parameter sets. He used a conceptual water balance model CATCHMOD to project changes in daily flows for the River Thames at Kingston. Monte Carlo sampling was also used to explore parameter stability and identifiability. Parameters were found to be highly sensitive to the training period. It was recommended that routine sensitivity analyses should be done for climate-change impact assessments to quantify uncertainties related to parameter instability, identifiability and non-uniqueness, while using conceptual water balance models.

Buytaert et al. (2009) analysed the impact of GCM model uncertainties on precipitation, temperature and evapotranspiration and assessed its propagation through hydroclimatic model to impact stream flow. They used entire GCM ensemble of IPCC's Fourth assessment report for climate change impact assessment in the tropical Andes. Four hydrological catchments were considered for the study, situated in the Paute river basin, south Ecuador. A1B scenario was used for GCM predictions of future period 2011-2030. They found that, temperature showed consistent changes for all GCM models. For, precipitation, the range of changes is relatively large and very high discrepancies between individual models could be found. Potential evapotranspiration was also affected by this divergent projection of precipitation and thus showed high range of change. It was found that, these uncertainties in precipitation and evapotranspiration propagated through the hydrological model and widely affected the future projection of stream flow, exhibiting diverse results for different GCM model. They also suggested that along with GCM model uncertainty, other uncertainties related to hydrological model and downscaling techniques should also be considered, which were neglected in the study.

Najafi et al. (2011b) assessed the uncertainties associated with GCM models and hydrologic model selection in climate change impact studies. The study was conducted over a rainfall-dominated basin, namely Tualatin River Basin (TRB) in Oregon, USA. Statistically downscaled outputs for eight GCM models and two emission scenarios were used to assess the multi-model uncertainties. These downscaled climate data were then used as inputs to four hydrologic models. The hydrologic models used in the study are: the Sacramento Soil Moisture Accounting (SAC-SMA) model, Conceptual Hydrologic Model (HYMOD), Thornthwaite-Mather model (TM) and the Precipitation Runoff Modelling System (PRMS). The results exhibited that the GCMs uncertainties diverge between different hydrologic models during summer; however, the discrepancies are not substantial in the winter. Therefore, GCM uncertainty is quite higher than hydrologic model uncertainty in wet season. However, hydrologic model uncertainties become important when analysing the dry season. This indicated importance of hydrologic model selection in climate change impact assessment.

Jung et al. (2012) have done an uncertainty assessment of climate change impacts for two hydrologically distinct river basins, one dominated by snowfall in winter and the other dominated by rainfall in all seasons. Two emission scenarios and eight GCMs are considered in the study. The parameter space of Precipitation Runoff Modelling System (PRMS) were

sampled using Latin Hypercube Sampling (LHS). Uncertainties associated with GCM models, emission scenario, and hydrologic model parameter identification were assessed for seasonal mean as well as for extreme flows for the two basins. It was found that the hydrological uncertainty significantly varies between the two basins. The uncertainty range for summer runoff changes were found to be relatively larger than those in winter. Seasonal runoff changes indicated GCM uncertainty to be greater than scenario uncertainty. The changes in winter runoff for the snow-dominated basin were more influenced by hydrologic model parameter uncertainty, while they are less affected in the rain-dominated basin. This indicated the need of more careful interpretation of runoff projections for snow-dominated regions.

Chen et al. (2013) has evaluated the uncertainties related to empirical downscaling methods in quantifying climate change impacts on hydrology in two North American river basins. Six different downscaling methods are used in the study: two change factor approaches (constant scaling and daily scaling) and four bias correction approaches (daily translation, local intensity scaling, daily bias correction and quantile mapping). To understand the uncertainty associated with climate simulations, four RCM simulations were considered. The future hydrological impacts on discharge (annual mean and extreme) from these simulated series were assessed by using an empirical lumped hydrological model (HSAMI), developed by Hydro-Québec. The results exhibited that the biases are significantly large in case of extreme values as compared to annual mean discharge values. This was also true for climate change projections (precipitation and temperature). In case of downscaling methods, it was found that uncertainty associated with choice of approach is more than within each type of approach. Further, the results indicated that uncertainty linked to the choice of downscaling method is similar to that of RCM simulations, later being slightly higher. They emphasized on using multi-model climate projections and downscaling approaches for better assessment of uncertainty.

Kumar et al. (2013) have presented a high-resolution multi-model projection for climate change over India and assessed the uncertainties associated with the approach. The simulated precipitation and temperature for the present and future time period by using ensembles of GCMs and RCMs. The results indicated that GCMs are suitable for assessment at coarse resolution as it they exhibit less biases, but regional scale orographic rainfall was not captured by GCMs. Whereas, RCMs has shown better reproduction of monsoon rainfall variations at regional scale. The two GCMs, ECHAM5 and HadCM3, that captured the mean monsoon rainfall and interannual variability to a good extent were used to force the RCM simulations to

simulate the future projections. The results showed increase in precipitation over Indo-Gangetic plane and NW India and decrease over the east coast of peninsular India by the end of 21st century, thus indicating significant spatial variability. In case of temperature, future projections suggest a wide spread warming by the end of 21st century. Uncertainty assessment indicated that the RCMs simulated precipitation and temperature changes are largely influenced by the driving models and the impact of the GCM on projected precipitation change is as strong as the variability among the RCMs.

Wang et al. (2013) have assessed the climate change impacts and associated uncertainty of extreme rainfall events in the Apalachicola River basin, Florida. Seven different RCM models have been used in the study, out of which two models, RCM3–GFDL and HRM3–HADCM3, were found to be suitable for assessing high intensity events during mid-afternoon. Using these two models, rainfall intensity–duration–frequency (IDF) curves were generated to understand the potential impact of climate change on IDF generation. Two methods were used for bias correction of future IDF curves of rainfall, (i) Maximum intensity percentile-based method, and (ii) Sequential monthly bias correction and maximum intensity percentile-based method. From the results, it was found that rainfall of the base period was significantly overestimated by RCM3-GFDL model, implying presence of high uncertainty. While HRM3–HADCM3 has shown no significant change in rainfall projection for upstream and midstream (but high intensity for downstream), RCM3-GFDL showed increased rainfall intensity from upstream to downstream. Temporal shift of extreme rainfall events has also been noticed, which may aggravate the flood magnitudes leading to increasing sediment load in the basin.

Chen et al. (2014) have assessed the influence of climate change on mean annual rainfall in monsoon areas of Asia. Two parameters are considered in the study, precipitation and temperature and their effect on runoff changes under the impact of climate change have been assessed in eight different basins in China. In order to check the elasticity of runoff change due to precipitation and temperature, Fu's two-parameter method of climate elasticity was used. It was further improved by Gardner function that enhances the prediction reliability of the climate elasticity index. The precipitation and temperature changes were predicted for the future decades under RCP2.6, RCP4.5 and RCP8.5 using 21 different GCMs. It was found that, both temperature and precipitation showed increasing trend in the future, but different basin responds differently to the same climate change scenario. Runoff was found to be positively related to precipitation but negatively affected by temperature. However, sensitivity of runoff

is more towards precipitation than temperature. Uncertainty associated with the future projection of the climate variables were assessed by using the innovative method proposed by Woldemeskel et al. (2012). Overall, decreasing trends were seen for the basins in future, with highest decrease under RCP8.5 scenario.

Lespinas et al. (2014) have assessed uncertainties associated with regional climate models and hydrological models. Six Mediterranean river basins situated in the southern coastal France region were selected for the study. A conceptual hydrological model, GR2M, was used for simulating the river discharge. GR2M model requires monthly precipitation (P) and potential evapotranspiration (PE) as inputs. Due to limited availability of observed PE data, PE values were estimated from temperature data by using an empirical formula developed by Folton and Lavabre (2014). Dynamically downscaled RCM data for the time period 2071-2100 was obtained from Prediction of Regional scenarios and Uncertainties for Defining European Climate change risks and Effects (PRUDENCE) project. Two climate scenarios (A2 and B2) were considered in the future projection. These projected datasets were used as input for the hydrological model, after implementing the classical perturbation method, to project water discharge for the period 2071-2100. Uncertainties associated with the model were assessed thereafter. They stated that uncertainties may arise due to the choice of RCM scenario, choice of hydrological model, choice of formula used for PE estimation, choice of calibration period considered and the dynamical parameterization of the hydrological model. The results indicated that the model is quite sensitive towards the choice of the PE formulation used. It was found that a simple correction method based on a dynamical parametrization of one model parameter with temperature data significantly reduces the model drift.

Milne et al. (2014) have analysed the uncertainties associated with nitrous oxide and methane emissions in the UK's greenhouse gas inventory for agriculture, using Monte Carlo simulation. They found that the uncertainties propagate through the models to the estimated emissions. A sensitivity analysis indicated that the uncertainty in the emission factor for emissions from N inputs affected uncertainty the most in England and Scotland, but in Northern Ireland and Wales, the emission factor for N leaching and runoff (EF5) influenced the uncertainty. They presented that if uncertainty associated with any of these emission factors can be reduced by 50%, it would reduce the uncertainty in emissions of nitrous oxide by 10%. They suggested that firm guidance should be provided by IPCC on this issue, as several countries are moving towards inventory models with disaggregation.

Najafi and Moradkhani (2015) have analysed the impacts of climate change on extreme values of runoff by applying multi-model ensemble analysis and assessed the uncertainties associated with the process. 8 GCM-RCM combinations were used for the study, provided by the North American Regional Climate Change Assessment Program (NARCCAP). The Pacific Northwest region that covers the Columbia river basin is considered for the study, which has both rain and snow dominated regions. The outputs from the RCMs were fed into the Variable Infiltration Capacity (VIC) hydrologic model to simulate the historical and future runoff, followed by hierarchical Bayesian modelling for evaluation of extreme runoff. Ensemble merging was done by using the Bayesian Model Averaging (BMA) method. Results indicated that maximum temperature, minimum temperature and wind speed had less variations among different model simulations and hence the RCMs were equally weighted. Whereas for precipitation and runoff, the variations were significant indicating high spatial variability within individual RCMs. Ensemble merging was compared for three different strategies: merging of climatological signals, merging after hydrological modelling and merging after extreme analysis. It was found that, first two cases underestimated the extreme runoff in all seasons for most of the regions. Overall, the multi-model ensemble results indicated increase in the future extreme runoffs in most regions, particularly over high elevated reigons.

Neupane et al. (2015) have studied the changes in hydrological projections due to climate change and deforestation, in monsoon-dominated Tamor and Seti river basins of Himalayan mountain range. The Soil and Water Assessment Tool (SWAT) was used to project the hydrological and land use changes. 16 GCM data were used for three different emission scenarios (A1B, A2 and B1) to project future temperature and precipitation changes. Land use changes were spatially modelled, including expansion of grassland, agricultural land and human settlement. Future predictions for all GCMs resulted in higher precipitation with subsequent increase in discharge. However, minimal changes were found in stream discharge for land use scenarios. Some uncertainties in precipitation predictions were found, while capturing the peak flows of daily simulations by the SWAT model. This indicated uncertainties in the climate variables that might have occurred due to extreme topographical variations of the Himalayan region.

Remesan and Holman (2015) evaluated the effect of meteorological data selection for baseline period on climate change impacted hydrological modelling. The study was done on Himalayan Beas river of the Indus basin, India. Six different combinations of baseline data, using two

gridded precipitation datasets (the Tropical Rainfall Measuring Mission (TRMM) and the Asian Precipitation – Highly Resolved Observational Data Integration Towards Evaluation of Water Resources (APHRODITE)) was used for calibration of the HySim model, along with three reference evapotranspiration equations of different complexity level (FAO Penman–Monteith, Hargreaves–Samani and Priestley–Taylor). For the historical period, all the six models showed similar performance. However, for the future period, significant variations in precipitation and temperature changes were found in the models' responses. This indicated the substantial influence of baseline data selection on climate change impacted future hydrological simulations.

Woldemeskel et al. (2016) have mentioned that there are three main sources of uncertainty associated with GCMs in climate projections: model structure, the emission scenarios and natural variability. They used 6 CMIP3 and 13 CMIP5 models, with all RCPs and scenarios, to investigate and to compare their capability in reducing uncertainties associated with them. The square root error variance (SREV) method, developed by Woldemeskel et al. (2012), was used to quantify the uncertainties in both set of models. Precipitation and temperature projections were considered in the study since they are the key input parameters of hydrological studies. From the study, it was found that total uncertainty in CMIP5 precipitation at global scale is of the same order as that of CMIP3 for the historical period. However, for the future projection, CMIP5 showed reduced uncertainty as compared to CMIP3 indicating decrease in the model structure uncertainty. Uncertainty due to natural variability was found to be contributing the least to the total uncertainty. No significant change, however, could be seen for temperature between the two sets of models. Regional scale uncertainty indicated larger uncertainties in precipitation in coastal and mountainous regions and areas with heavy rainfall. Whereas, for temperature, the uncertainties are more in lower elevation areas and in extratropical cold regions.

You et al. (2017) have investigated the diurnal temperature range (DTR) from 17 different CMIP5 GCMs and compared the results with observed data over the Tibetan plateau. While DTR range in observed data shows significant increasing trend, most of the GCMs results exhibited underestimated DTR values. Moreover, correlations of DTR with GCM variables were assessed, which indicated that model differences were possibly dominated by the variations in the values of radiation variables and total cloud fraction. Statistical indices (AE, RE and RMSE) were used to understand the uncertainties associated with the results.

Joseph et al. (2018) have compared the parameter uncertainty (uncertainty arising from hydrological parameterization in climate change projections) with climate model uncertainty in hydrologic climate change impact assessments. The study was implemented over the Ganga River Basin, India. Three different GCM models were used in the analysis: Institut Pierre Simon Laplace (IPSL), European Consortium Earth System Model (EC-EARTH) and MPI (Max Plank Institut) ESM (Earth System Model). For climate projections, statistical downscaling outputs of the GCMs and dynamic downscaling outputs from Coordinated Regional Climate Downscaling Experiment (CORDEX) have been used. For analysing the sensitivity of hydrological parameters (parameter uncertainty), Monte-Carlo Simulations (MCS) were performed. It was found that, climate model uncertainties are significant in all seasons of a year and for all hydrological variables. Whereas, parameter uncertainty is almost negligible in comparison to huge climate model uncertainty.

Uncertainties have also largely affected the water management and policy makers. Kundzewicz et al. (2018) presented a framework for assessing and reducing uncertainty associated with climate change impact assessments. Moreover, measures for uncertainty improvements have also been proposed. They suggested two alternative courses of action if uncertainty is irreducible: the precautionary principle and the adaptive management. Precautionary principle is a variation of the min-max concept, where the approach that minimises the worst outcome (such as threats of serious or irreversible damage) is chosen. When uncertainties are large for different climate impact scenarios, adaptive planning should be made based on ensembles and multi-model probabilistic approaches.

Shen et al. (2018) have mentioned that uncertainty associated with GCM models is the largest contributor of uncertainty in climate change impacted hydrological predictions. To examine the GCM uncertainty, 20 CMIP5 GCM models were selected under RCP4.5 and RCP8.5. The Hanjiang River watershed, situated in south-central China, was considered as the study area. The Daily Bias Correction (DBC) method was utilized for its simplicity and good performance to downscale the GCM outputs. A lumped conceptual rainfall-runoff model, known as HMETS model (Hydrological Model of École de technologie supérieure), developed at the École de technologie supérieure, University of Quebec, was used for flow simulations. Long term temporal variations associated with GCM uncertainty were examined, by considering fifty-one 30-year future periods, moving from 2021 to 2100 with one-year interval. Wavelet analysis was used for investigating the climatic and hydrological changes and their uncertainty, whereas

Mann-Kendall trend test and regression analysis were used for trend analysis of those predictions. It was found from the study that, precipitation and temperature are highly influenced by GCM uncertainty, which significantly increase over time. Furthermore, higher emission scenario (RCP8.5) has higher uncertainty than lower scenario (RCP4.5). Finally, they suggested that multiple GCM models should be utilized for climate change impacted hydrological study to better tackle the uncertainties.

Soraisam et al. (2018) have assessed the uncertainties associated with different types of observed as well as reanalysed or downscaled datasets and their future projections. The study was focused on northeast Indian region, as the region has highly varying orography compared to a smaller number of observed and modelled data. Seasonal cycle, mean and extreme value analysis as well as trend analysis have been done in the study using IMD observed dataset, APHRODITE and CRU reanalysis data and five RCMs from CORDEX data (LMDZ, MPI, GFDL, CNRM and ACCESS). Downscaled historical simulations were validated with observed datasets, and future scenarios (2011-2060) for RCP4.5 were compared among the different models. For the historical period, the models were found to be underestimating the hydrological parameters along with higher spatial distribution; however, seasonal cycle showed qualitatively good response, specifically for temperature. Future trends of temperature were found to be increasing, whereas precipitation showed decreasing or insignificant trend. They suggested that, these limitations of observed, downscaled and model results should be kept in mind while performing regional scale climate study over regions with complex orography.

Zhuang et al. (2018) have developed an integrated simulation-optimization (ISO) approach for analysing the impacts of climate change on water resources and uncertainty assessment. The study was done on Kaidu watershed in northwest China, a snowmelt-precipitation-driven watershed. The ISO was developed to incorporate meteorological projection from multiple GCMs, hydrological simulation for daily streamflow, and an interval multistage chance-constrained (IMC) optimization. The ISO also allows thorough analyses of various policy scenarios. The results indicated increasing trend for both temperature and precipitation. It was realized that the uncertainties in the system would significantly affect the allocation pattern of water resources, indicating shortage of water in the future.

Gao et al. (2019) assessed the uncertainty associated with bias-corrected climate change projections and its influence on streamflow and watershed hydrology. The assessment was

performed on the Smoky Hill River Watershed in West-Central Kansas, USA. Three climate datasets (1985-2004) were used as observed data for bias-correction of GCM model predictions: NCDC land-based weather stations, PRISM spatial grid and NEXRAD spatial grid. GCM data were collected for six different high-resolution models from CMIP5, for RCP2.6, RCP4.5 and RCP8.5. To simulate the streamflow, Soil and Water Assessment Tool (SWAT) model was used. The bias-correction method used in the study is the Daily Translation (DT) method. Result indicated that the hydrologic model developed in SWAT is more sensitive to precipitation than temperature. Uncertainty assessment indicated that GCM model selection imposes the largest uncertainty to simulated streamflow. Overall, the study exhibited that the streamflow sensitivity to selection of bias-correction of data is less than GCM selection, but more than sensitivity to RCM selection.

2.6. SUMMARY

The above literature survey indicated the existence of major impacts of changing climate on water resources, which could be seen throughout the world. Various downscaling techniques were utilized to study the regional impacts. Application of fuzzy clustering has gained lot of interests due to its vast applicability in identification of hydrologically homogeneous climatic regions. However, selection of appropriate hydrologic model, GCM as well as downscaling technique is still not very clear as they are largely dependent on the characteristics of the study area and hence varies from region to region. Moreover, there are a lot of uncertainties related to these model applications. It also came into notice that, orographic arrangement of a region affects the overall results. Few studies could be found that have incorporated the topography of the study area for analyses, yet enormous scope and need for study is there on projecting future impacts of climate change considering the effects of the orographic arrangements.

Chapter 3

DELINEATION OF THE NORTHEAST INDIAN REGION INTO HOMOGENEOUS CLUSTERS USING FUZZY CLUSTERING APPROACH

3.1. INTRODUCTION

Rainfall is one of the most important hydrological parameters that requires to be studied scrupulously, both in spatial and temporal scale. However, challenge comes in understanding the behaviour of rainfall distribution pattern when applied on a regional scale. Pattern of rainfall, its frequency and magnitude may drastically vary depending upon the orography of the region (Venkatesh and Jose, 2007). Scarcity of sufficient data at many sites of interest, may further complicate the investigation. To address this issue, the region may be classified into few homogeneous rainfall regions of similar rainfall distribution, also termed as regionalization. The northeast Indian region has its own peculiar topography. The variations in the orographic arrangements and altitude differences in the region give rise to irregular and complex rainfall patterns at a local scale, which eventually amplifies the need of regionalization.

Regionalization has found its way into many applications in water resources planning, agricultural planning, drainage design, and estimating magnitude and frequency of extreme events like flood and drought. Literatures indicate that, in the past, political and geographical boundaries are used as a basis of forming homogeneous regions (Thomas and Benson, 1970; Chew and Heiler, 1974; NERC, 1975; Beable and McKerchar, 1982). However, use of political and geographical boundaries is found to be not very convincing while forming hydrologically homogeneous regions (Bonell and Sumner, 1992; Burn, 1997; Rao and Srinivas, 2006b, 2006a; Satyanarayana and Srinivas, 2011; Dikbas et al., 2012). A considerable number of research investigations have been carried out in the recent years to identify homogeneous rainfall regions with various methods other than geographical divisions (Bedi and Bindra, 1980; Bärring, 1987; Sumner and Bonell, 1988; Kulkarni et al., 1992; Gadgil et al., 1993; Burn, 1997; Adelekan, 1998). Regionalization with the use of principal component analysis (PCA) was found to be of use (Singh and Singh, 1996; Wotling et al., 2000). When subjectivity involved with PCA came into notice, the concept of cluster analysis started getting attention (Bonell and Sumner, 1992; Guttman, 1993; Venkatesh and Jose, 2007). Cluster analysis refers to a varied

group of statistical procedures used to classify a multivariate dataset into some clusters or groups (Rao and Srinivas, 2006b; Srinivas et al., 2008; Dikbas et al., 2012). Studies have also been done where PCA was further associated with cluster analysis for homogeneous clustering (Dinpashoh et al., 2004; Satyanarayana and Srinivas, 2011; Darand and Mansouri Daneshvar, 2014). The k-means clustering is being vastly applied for regionalization (Rao and Srinivas, 2006b; Pelczer et al., 2007; Agarwal et al., 2016). K-means clustering splits a region into hard clusters, i.e. with a degree of membership 1 or 0. This means that a site can at most belong to only one cluster. However, this may not be valid in real world cases. To address this matter, the concept of fuzzy clustering was introduced to regionalization, which allows a site to fit into several clusters concurrently with a certain membership value. The fuzzy membership value of a site signifies the extent to which it fits into a particular group of sites (Rao and Srinivas, 2006a). A lot of studies have successfully applied fuzzy clustering technique for clustering of hydrologically homogeneous regions in the recent years (Hall and Minns, 1999; Owen et al., 2006; Plain et al., 2008; Sadri and Burn, 2011; Satyanarayana and Srinivas, 2011; Chen et al., 2011; Dikbas et al., 2012; Mok et al., 2012; Chavoshi et al., 2013; Asong et al., 2015; Bharath and Srinivas, 2015; Goyal and Sharma, 2016; Irwin et al., 2017; Wang et al., 2017). On the basis of critical reviews of earlier studies on regionalization, the objective of this chapter is defined as to identify homogeneous rainfall regions in the northeast Indian region by using fuzzy clustering approach. Initially, 10 IMD (Indian Meteorological Department) stations have been selected to make the clusters, situated in various regions of the northeast. On the Basis of IMD data available for those stations, homogeneous clustering has been done using both rainfall data as well as GCM data with the use of fuzzy clustering approach. Clustering has also been done considering gridded rainfall data (collected from IMD) within which all the 10 IMD stations come. Furthermore, sub-clustering of upper region of Brahmaputra valley has also been done using rainfall data recorded by the tea gardens of Assam, to understand the effect of local rainfall data in homogeneous clustering. To achieve the best possible partition from the fuzzy c-means (FCM) algorithm, seven different cluster validity indices (CVs) have been used. Three different combinations of feature vectors have been employed in FCM algorithm and the outputs have been compared for attaining best solutions to regionalization. The homogeneous rainfall regions (fuzzy clusters) thus formed by the use of FCM algorithm and validated with CVs have been assessed for statistical homogeneity by performing homogeneity tests using L-moment approach (Hosking and Wallis, 1997). The results are compared with some previous regionalization studies and finally concluding remarks are presented.

3.2. METHODOLOGY

In the subsections below, the FCM clustering approach as well as the FCM algorithm for delineation of homogeneous clusters is described at first, which is followed by a brief description of the CVs, homogeneity test methods and process of adjustment of heterogeneous clusters. The methodology used is shown in Fig. 3.1, by means of a flow chart.

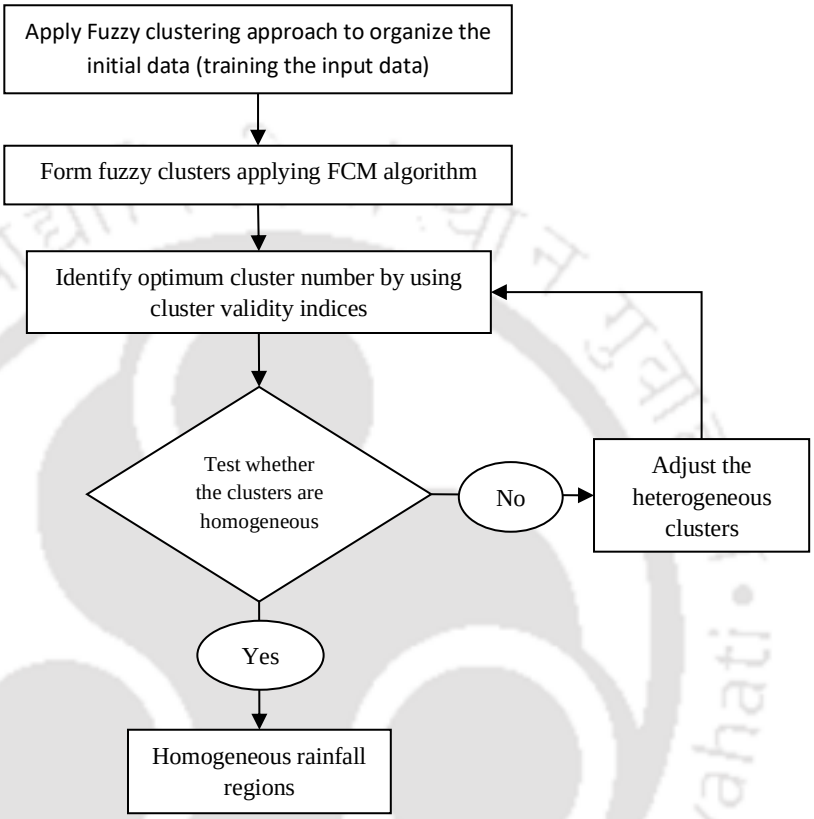


Figure 3. 1 Flow chart of the methodology proposed for identification of homogeneous rainfall regions

3.2.1. Fuzzy C-Means clustering (FCM)

The FCM approach is basically optimization of fuzzy c-means objective function. It was initially developed by Dunn (1973) and afterwards modified by Bezdek et al. (1984). The fuzzy c-means function to be minimized, is expressed as:

$$J(\boldsymbol{Z};\boldsymbol{U},\boldsymbol{V}) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^{m}\|\boldsymbol{z}_k-\boldsymbol{v}_i\|^2 \; A \tag{3.1}$$

where, $\boldsymbol{U} = [\mu_{ik}] \in M_{fc}$ is the fuzzy partition matrix of Z; $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_c]$, $\boldsymbol{v}_i \in R^n$ is vector of cluster centers to be determined; $d^2(\boldsymbol{z}_k, \boldsymbol{v}_i)A = D_{ikA}^2 = \|\boldsymbol{z}_k - \boldsymbol{v}_i\|^2 A = (\boldsymbol{z}_k - \boldsymbol{v}_i)^T A(\boldsymbol{z}_k - \boldsymbol{v}_i)$ is a squared distance norm; and m is the fuzziness parameter or fuzzifier,

where $m = [1, \infty)$. Usually, its value falls between 1 and 2.5 (Pal and Bezdek, 1995)

3.2.2. FCM algorithm for delineation of homogeneous rainfall regions

Suppose there are '*N*' number of sites in a study area. The '*n*' attributes, influencing rainfall at each site, have to be identified first. As mentioned by Satyanarayana and Srinivas (2011), the attributes may include large scale atmospheric variables (LSAVs) of Global Climate Models (GCMs) or their principal components (PCs), location parameters (latitude, longitude and altitude), and seasonality measures (maximum, minimum, standard deviation of rainfall etc.). Consequently, a feature vector is formed for each site using the identified attributes for the corresponding sites.

The i^{th} site, in n-dimensional attribute space, is denoted by the feature vector

$$\boldsymbol{y}_i = [y_{1i}, \dots, y_{ji}, \dots, y_{ni}]^T \in R^n, \tag{3.2}$$

where, y_{ji} is the value of j^{th} attribute in y_i. The attributes of y_i are rescaled using

$$x_{ji} = \frac{(y_{ji} - \overline{y_j})}{\sigma_j}, for\ 1 \le j \le n, 1 \le i \le n \tag{3.3}$$

where, x_{ji} denotes the rescaled value of y_{ji}, σ_j represents the standard deviation of attribute j, and $\overline{y_j}$ is the mean value of attribute j over all the N feature vectors.

The necessity of rescaling the attributes comes in order to nullify the differences in their variance, importance and relative magnitude. Otherwise, the attributes having greater variance and magnitude will influence the formation of clusters, which is not desired. Importance of rescaling in cluster analysis is discussed earlier, which can be found in Dunn and Landwehr (1980) and Hosking and Wallis (1997). If it is known that certain attributes are more important than others in influencing rainfall in that study area, then the rescaling should be done in such a way that the variances of rescaled values of those more important attributes are greater than those of the less important attributes.

Let $\boldsymbol{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_i, \dots, \boldsymbol{x}_N)$ represents matrix containing rescaled feature vectors, where x_i is rescaled feature vector for the i^{th} site. Next task is to partition $\boldsymbol{X}$ into c soft clusters using Fuzzy c-means (FCM) algorithm (Bezdek, 1981), to arrive at the optimum value of the following objective function:

Minimize, $$J(\boldsymbol{X}; \boldsymbol{U}, \boldsymbol{V}) = \sum_{k=1}^{c} \sum_{i=1}^{N} (\mu_{ki})^m \|\boldsymbol{x}_i - \boldsymbol{v}_k\|^2 \tag{3.4}$$

which can also be written as,

Minimize, $$J(\boldsymbol{X};\boldsymbol{U},\boldsymbol{V}) = \sum_{k=1}^{c}\sum_{i=1}^{N}(\mu_{ki})^{m}\ d^2(\boldsymbol{x}_i,\boldsymbol{v}_k) \quad (3.5)$$

Subject to the following constraints,

$$\sum_{k=1}^{c}\mu_{ki} = 1\ \forall i\ \in \{1,\ldots.,N\} \quad (3.6)$$

$$0 < \sum_{i=1}^{N}\mu_{ki} < N\ \forall k \in \{1,\ldots,c\} \quad (3.7)$$

where $\boldsymbol{V} = (\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k,\ldots,\boldsymbol{v}_c)$ represents a matrix of cluster centroids and $\boldsymbol{v}_k = [v_{1k},\ldots,v_{jk},\ldots,v_{nk}] \in R^n$ indicates centroid of k^{th} soft cluster,

$\mu_{ki} \in [0,1]$ is the membership of x_i in the k^{th} soft cluster; $\boldsymbol{U}$ is the fuzzy partition matrix which contains the membership of each rescaled feature vector in each soft cluster; The parameter $m \in [1,\infty)$, known as fuzzifier, refers to the weight exponent for each fuzzy membership; d($\boldsymbol{x}_i$, $\boldsymbol{v}_k$) is the distance from $\boldsymbol{x}_i$ to $\boldsymbol{v}_k$.

The iterative method of FCM algorithm (Bezdek, 1981) used to form the homogeneous rainfall regions is given below:

(i) The initial fuzzy partition matrix $\boldsymbol{U}$ is set.

(ii) Then, initial membership values μ_{ki}^{init} of x_i that belongs to cluster k, is adjusted by using equation:

$$\mu_{ki} = \frac{\mu_{ki}^{init}}{\sum_{j=1}^{c}\mu_{ji}^{init}}, for\ 1 \le k \le c, 1 \le i \le N \quad (3.8)$$

(iii) Fuzzy cluster centroid $\boldsymbol{v}_k$ is then calculated as

$$\mathbf{v}_{\mathrm{k}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mu_{\mathrm{ki}})^{\mathrm{m}}\mathbf{x}_{\mathrm{i}}}{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mu_{\mathrm{ki}})^{\mathrm{m}}}, \text{for } 1 \le \mathrm{k} \le \mathrm{c} \quad (3.9)$$

(iv) Fuzzy membership value μ_{ki} is updated as

$$\mu_{ki} = \frac{(\frac{1}{d^2(x_i v_k)})^{\frac{1}{m-1}}}{\sum_{k=1}^{c}(\frac{1}{d^2(x_i v_k)})^{\frac{1}{m-1}}}, for\ 1 \le k \le c, 1 \le i \le N \quad (3.10)$$

(v) The objective function is then calculated as

$$J(\boldsymbol{X};\boldsymbol{U},\boldsymbol{V}) = \sum_{k=1}^{c}\sum_{i=1}^{N}(\mu_{ki})^{m}\ d^{2}(\boldsymbol{x}_i,\boldsymbol{v}_k) \tag{3.11}$$

The above steps from (iii) to (v) are repeated till the difference in the objective function for two consecutive iterations becomes adequately small.

3.2.3. Parameters of the Fuzzy C-means (FCM) algorithm

Before using the FCM algorithm, the following parameters must be specified:

1. The Number of Clusters, *c*

The cluster number *c* is initially chosen as per the required number of regions. It is the most important parameter since the remaining parameters have less influence on the resulting partition. Rao and Srinivas (2006b) stated that optimal value of *c* decreases with increase in number of clusters. However, since the number of regions is not known a priori, one has to make assumptions about the number of clusters. The best way to do that is to choose a range of values for *c*, and subsequently to find out the best one.

2. The Fuzziness Exponent or Fuzzifier, *m*

The weighting exponent *m* is an important parameter because it significantly influences the fuzziness of the resulting partition. As *m* approaches 1($m \rightarrow 1$), the partition becomes hard ($\mu_{ik} \in \{0,1\}$). As $m \rightarrow \infty$, the partition becomes completely fuzzy ($\mu_{ik} = 1/c$). Usually, *m* = 1 to 2.5 is initially chosen (Pal and Bezdek, 1995).

3. The Termination Tolerance (Absolute Difference between Two Successive Iterations)

The FCM algorithm discontinues the iteration when the norm of the difference between $\boldsymbol{U}$ in two successive iterations becomes smaller than the termination tolerance. Usually 0.001 is taken, even though 0.01 works well in most cases, while drastically reducing the computing times.

4. The Fuzzy Partition Matrix, $\boldsymbol{U}$

The fuzzy partition matrix must be initialized at the beginning. However, taking random value for $\boldsymbol{U}$ is also acceptable as the algorithm is not affected by the initial value of $\boldsymbol{U}$.

3.2.4. Cluster Validity indices (CVs)

FCM algorithm divides the data into well separated and compact clusters, given the optimal values of *c* and *m*. Hence deciding the optimal values of these parameters is very crucial. Bezdek (1981) addressed this matter by stating the concept of validity indices. These indices essentially measure the goodness of the partitioned clusters. In hydrological studies several indices are used (Hall and Minns, 1999; Srinivas et al., 2008). In case of FCM algorithm, the following CVs are found to perform well:

1. Fuzzy partition coefficient (V_{PC})

$$V_{PC}(U) = \frac{1}{N}\sum_{k=1}^{c}\sum_{i=1}^{N}(\mu_{ki})^2 \tag{3.12}$$

2. Fuzzy partition entropy (V_{PE})

$$V_{PE}(U) = -\frac{1}{N}[\sum_{k=1}^{c}\sum_{i=1}^{N}\mu_{ki}log(\mu_{ki})] \tag{3.13}$$

3. Fuzziness performance index (V_{FPI})

$$V_{FPI}(U) = 1 - \frac{c \times V_{PC}(U) - 1}{c-1} \tag{3.14}$$

4. Normalized classification entropy (V_{NCE})

$$V_{NCE}(U) = \frac{V_{PE}(U)}{\log(c)} \tag{3.15}$$

Bezdek† (1973, 1974) formulated V_{PC} and V_{PE}, whereas V_{FPI} and V_{NCE} were proposed by Roubens (1982). The range for V_{PC} is [1/c, 1]; V_{PC} =1 /c indicates equal sharing of clusters i.e. equal membership values of a data in all clusters (i.e. μ_{ki}=1/c ∀i,k) and V_{PC} =1 indicates no sharing of membership among the clusters. This means that a data can fit into only a single cluster. Similarly, the range of V_{PE} is [0, log (c)]. V_{PE}=0 implies no sharing of membership among clusters and V_{PE}=log(c) implies equal sharing of clusters (i.e. μ_{ki} =1/c ∀ i, k). On the contrary, this range is [0, 1] for V_{FPI} and V_{NCE}; 0 implies no membership sharing between clusters and 1 implies equal sharing of clusters (i.e. μ_{ki} =1/c ∀ i, k). For a hard or crisp clustering, V_{PC} should be 1, whilst V_{PE}, V_{FPI} and V_{NCE} should be zero. As such, a maximum value for V_{PC} indicates optimum partition (i.e. minimum value for V_{PE}, V_{FPI} and V_{NCE}), that means least overlap among clusters.

Earlier studies have stated that these four CVs tend to display monotonous increasing or decreasing trend (Rao and Srinivas, 2006a; Srinivas et al., 2008). Hence, they are not very effective in obtaining optimum partition to delineate rainfall regions. Xie and Beni (1991) found no direct correlation of V_{PC} and V_{PE} with any property of the data. Furthermore, they are

found to be very sensitive to the fuzzifier value, *m* (Halkidi et al., 2001). Here, these indices are used mainly to validate their performances in detecting optimum cluster number. To eliminate this drawback the other validity indices are introduced, as explained below:

5. Extended Xie and beni index (V_{XB})

$$V_{XB,m}(U,V:X) = \frac{\sum_{k=1}^{c}\sum_{i=1}^{N}(\mu_{ki})^m \|\boldsymbol{x}_i - \boldsymbol{v}_k\|^2}{N \min_{l,l\neq k}\|\boldsymbol{v}_l - \boldsymbol{v}_k\|^2} \quad (3.16)$$

Xie and Beni (1991) proposed the cluster validity index $V_{XB,m}$. It quantifies the ratio of compactness within a fuzzy cluster to separation of clusters. Optimal partition of clusters should exhibit minimum value of $V_{XB,m}$.

6. Fukuyama and Sugeno index (V_{FS})

$$V_{FS}(U,V:X) = \sum_{i=1}^{N}\sum_{k=1}^{c}(\mu_{ki})^m \|\boldsymbol{x}_i - \boldsymbol{v}_k\|^2 A - \sum_{i=1}^{N}\sum_{k=1}^{c}(\mu_{ki})^m \|\boldsymbol{v}_i - \bar{\boldsymbol{v}}\|^2 A \quad (3.17)$$

Proposed by Fukuyama and Sugeno (1989), optimum partition is indicated by a minimum value of V_{FS}.

7. Kwon index (V_K)

$$V_K(U,V:X) = \frac{\sum_{k=1}^{c}\sum_{i=1}^{N}(\mu_{ki})^m \|\boldsymbol{x}_i - \boldsymbol{v}_k\|^2 + \frac{1}{c}\sum_{k=1}^{c}\|\boldsymbol{v}_i - \bar{\boldsymbol{v}}\|^2}{\min_{i\neq k}\|\boldsymbol{x}_i - \boldsymbol{v}_k\|^2} \quad (3.18)$$

The Extended Xie and beni index (V_{XB}) exhibits monotonous decreasing tendency as c→N. To address this problem Kwon (1998) proposed another index V_K that has an ad-hoc punishing function in numerator.

To determine the optimal values of *c* and *m*, a range of values for the two parameters are selected and subsequent partitioning results show different sets of clusters, along with their validity indices. To decide the optimal set of values for *c* and *m* among those sets, first the optimum selection criteria of each of the validity indices are examined. Then, the sites having greater membership value in the clusters are identified, based on a threshold value of fuzzy membership (T_i). Thus, a fuzzy cluster is made by allocating those sites to the cluster, whose membership values are found to be higher than or equal to the threshold fuzzy membership value (T_i). The selection of this threshold value is subjective (Satyanarayana and Srinivas, 2011). The most reasonable explanation would be to allocate the site to that group where its membership value is the highest. Yet, uncertainty arises when a site has low membership value in all the clusters or has equal memberships. To address this issue homogeneity test is done

which is followed by adjustment of heterogeneous clusters. The methodologies for both of these are explained in the following subsection.

3.2.5. Homogeneity test and adjustment of heterogeneous clusters

The fuzzy clusters thus formed by using FCM algorithm and validated with CVs are then required to be assessed for statistical homogeneity by performing homogeneity tests. Heterogeneity measure (L-moment based) proposed by Hosking and Wallis performs better when skewness is low (average L-skew < 0.23) for a sample set of data, whilst for higher skewness bootstrap Anderson-Darling test is recommended (Viglione et al., 2007). Previous studies have shown that Hosking and Wallis's homogeneity test is appropriate for delineation of homogeneous rainfall regions (Satyanarayana and Srinivas, 2011), hence is considered in this study.

3.2.5.1. L-moment of data samples

L-moment is a method of explaining the probability distribution shape and evaluating the distribution parameters, especially for small sample sizes of environmental data, since it is unbiased and has a nearly normal distribution (Hosking, 1990). Like usual moments, L-moments too determines the location, dispersion, peakedness, skewness and any other feature of shape of probability distribution. However, L-moments are derived from linear combination of data (Hosking, 1990). These statistics are established by modifying "probability weighted moments" (Greenwood et al., 1979), which explains L-moments by means of linear combinations. Sample probability weighted moments as explained by Greenwood et al. (1979) is given below:

$$b_0 = n^{-1} \sum_{j=1}^{n} x_j \tag{3.19}$$

$$b_r = n^{-1} \sum_{j=r+1}^{n} \frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)} x_j \tag{3.20}$$

The first few L-moments and L-moment ratios are defined as:

Location, mean (l_1): $l_1 = b_0$ (3.21)

Scale, L-Cv (t_2): $t_2 = \frac{l_2}{l_1}$, where $l_2 = 2b_1 - b_0$ (3.22)

L-Skewness (t_3): $t_3 = \frac{l_3}{l_2}$, where $l_3 = 6b_2 - 6b_1 + b_0$ (3.23)

L-Kurtosis (t_4): $t_4 = \frac{l_4}{l_2}$, where $l_4 = 20b_3 - 30b_2 + 12b_1 - b_0$ (3.24)

3.2.5.2. Discordancy measures, heterogeneity measures and adjustment of heterogeneous clusters

1. Discordancy measures

Discordancy measure (D_i) detects those sites which are unacceptably discordant with the designated cluster (Hosking and Wallis, 1993). This discordancy value for ith site (Hosking and Wallis, 1995) is given as,

$$D_i = \frac{N}{3(N-1)}(u_i - \bar{u})^T S^{-1}(u_i - \bar{u}) \quad (3.25)$$

Here, S is the sample covariance matrix expressed as:

$$S = (N-1)^{-1} \sum_{i=1}^{N}(u_i - \bar{u})(u_i - \bar{u})^T \quad (3.26)$$

where $u_i = \left[t^{(i)} t_3{}^{(i)} t_4{}^{(i)}\right]^T$ means a vector comprising of the values of t, t_3 and t_4 for ith site.

Hence, $$\bar{u} = N^{-1} \sum_{i=1}^{N} u_i \quad (3.27)$$

Large values of D_i indicate probable errors in the site data. Hosking and Wallis (1993) explained that a particular site is not considered to be homogeneous with the region if D_i is more than a certain critical value, than that. $D_i \geq 3$ is suggested as the criterion for affirming a site to be discordant, for regions with 15 or more sites. However, (Hosking and Wallis, 1993, 1995) have advised to scrutinize the dataset for the largest D_i values, irrespective of their magnitude.

2. Heterogeneity measures

Heterogeneity measures give the degree of heterogeneity existing within the region. It is estimated based on the extent of actual variability in L-moment ratios in relation to the expected variability in a homogeneous region. The heterogeneity measures to be estimated are H_1, H_2 and H_3. These measures are defined based on L-Cv, L-Skewness and L-Kurtosis. These three heterogeneity measures are given below.

i. Heterogeneity measures based on L-Cv

$$H_1 = \frac{V - \mu_V}{\sigma_V} \quad (3.28)$$

ii. Heterogeneity measures based on L-Cv and L-Skewness

$$H_2 = \frac{V_2 - \mu_{V2}}{\sigma_{V2}} \tag{3.29}$$

iii. Heterogeneity measures based on L-Cv and L-Kurtosis

$$H_3 = \frac{V_3 - \mu_{V3}}{\sigma_{V3}} \tag{3.30}$$

Here V is the weighted standard deviation or dispersion of the sample coefficients of L-variation (L-Cv); V_2 and V_3 denote weighted average distance from the site to group weighted mean in a 2-D space of L-Cv/L-Skewness and L-Skewness/L-Kurtosis, respectively; μ_V, μ_{V2} and μ_{V3} are the mean of V, V_2 and V_3 values, respectively, calculated from a large number of simulations (N_{sim}); σ_V, σ_{V2} and σ_{V3} are standard deviations of V, V_2 and V_3, respectively, calculated from a number of simulations. In this study N_{sim} is takes as 1000, simulated from kappa distribution and fitted using regional average L-moment ratios.

$$V = \left\{ \frac{\sum_{i=1}^{N} n_i \left(t^{(i)} - t^{(R)}\right)^2}{\sum_{i=1}^{N} n_i} \right\}^{1/2} \tag{3.31}$$

$$V_2 = \frac{\sum_{i=1}^{N} n_i \{ \left(t^{(i)} - t^{(R)}\right)^2 + \left(t_3^{(i)} - t_3^{(R)}\right)^2 \}^{1/2}}{\sum_{i=1}^{N} n_i} \tag{3.32}$$

$$V_3 = \frac{\sum_{i=1}^{N} n_i \{ \left(t_3^{(i)} - t_3^{(R)}\right)^2 + \left(t_4^{(i)} - t_4^{(R)}\right)^2 \}^{1/2}}{\sum_{i=1}^{N} n_i} \tag{3.33}$$

Here N indicates number of sites/stations in the clustered region; n_i indicates record length of site i; $t^{(i)}$, $t_3^{(i)}$ and $t_4^{(i)}$ indicate sample L-moment ratios of site i; $t^{(R)}$, $t_3^{(R)}$ and $t_4^{(R)}$ indicate regional average L-moment ratios (L-Cv, L-Skewness and L-Kurtosis respectively).

If H<1 for a region, it is described as ‘acceptably homogeneous’, 1≤H<2 implies ‘possibly heterogeneous’, and H≥2 implies ‘definitely heterogeneous’. For further details on heterogeneity measures, Hosking and Wallis (1997) can be referred.

3. Adjustment of heterogeneous clusters

Clusters formed by the use of clustering algorithms does not always exhibit statistical homogeneity. Even after performing homogeneity test, there is a need to adjust the possibly or definitely heterogeneous clusters to come to a definite conclusion. Furthermore, it is assumed that there is not any cross-correlation among the data. Nevertheless, in actual scenarios there

is gradual variation of rainfall across space, which implies that there exists cross-correlation among geographically contiguous sites (Satyanarayana and Srinivas, 2011). So further adjustments are essential to form homogeneous regions. Studies suggest that, to decrease heterogeneity measures values, the discordant sites of one region can be either removed or shifted to some other region, after confirming that the site has not exhibited high fuzzy membership value in that cluster and discordancy of that site is not because of sampling variability (Rao and Srinivas, 2006a; Satyanarayana and Srinivas, 2011). The heterogeneous regions can also be broken into two or more regions, and if required two or more small regions can be merged together.

3.3. DATA COLLECTION

Four types of data have been used in this study, namely

1. Observed rainfall data of 10 station, collected from IMD
2. Global Climate model (GCM) data collected from IPCC AR4 and AR5
3. Daily gridded rainfall data of size 0.5˚× 0.5˚, collected from IMD, Pune
4. Rainfall data, collected from various tea gardens of Assam.

3.3.1. Observed rainfall data

Observed daily rainfall data have been collected from Indian Meteorological Department (IMD) under an MOU between IIT Guwahati and IMD. Observed data from 01-01-1969 to 31-01-2012 were provided. However, data from 2001 to 2005 could not be obtained.

The 10 IMD stations selected, along with their latitude-longitude-elevation and rainfall data, are given in Table 3.1 below. The locations of the stations are shown in the Fig. 3.2. The study area is situated in the north-eastern region of India, covering mostly the Brahmaputra valley and the Barak valley region. It lies between 24˚45'0" N (24.75˚N) and 28˚6'0" N (28.1˚N) latitudes and 89˚59'0" E (89.98˚E) and 95˚23'0" E (95.38˚E) longitudes. This region of the valley is surrounded by Eastern Himalayas towards the north, the Patkai Bum in the east, the Naga Hills and Meghalaya plateau in the south. These valleys have great geographical as well as political significance. The orography of the north-eastern region is very unique and lots of geographical and meteorological variations can be seen in the region. Total annual rainfall (averaged over 30 years) of the stations under consideration varies from 1454.6 mm to 11759 mm and the elevation varies from as low as 21 m to as high as 1600 m. The north-eastern region of India has a subtropical climate which is influenced by northeast and southwest monsoon.

The Meghalaya plateau, the Himalayas and the surrounding hills of Arunachal Pradesh, Manipur, Nagaland and Mizoram influences the climate. The monsoon winds coming from Bay of Bengal move towards the northeast and hits these mountains causing heavy rainfall on the valley.

Table 3. 1 Latitude-longitude-elevation and rainfall data of the IMD stations

Index No.	Name of the stations	Region	Latitude	Longitude	Elevation	30-year annual average rainfall (in mm)
42220	Passighat	Arunachal	28 °6' 0" N (28.1 °)	95 °23' 0" E (95.3833 °)	157 m (515 ft)	4540.16
42309	North-Lakhimpur	Assam	27 °14' 0" N (27.2333 °)	94 °7' 0" E (94.1167 °)	101 m (331 ft)	3285.33
42314	Mohanbari	Assam	27 °29' 2" N (27.4839 °)	95 °1' 1" E (95.0169 °)	111 m (364 ft)	2654.87
42404	Dhubri	Assam	26 °1' 0" N (26.0167 °)	89 °59' 0" E (89.9833 °)	35 m (115 ft)	2400.21
42410	Guwahati	Assam	26 °6' 22" N (26.1061 °)	91 °35' 9" E (91.5859 °)	54 m (177 ft)	1811.8
42415	Tezpur	Assam	26 °37' 0" N (26.6167 °)	92 °47' 0" E (92.7833 °)	91 m (299 ft)	1986.35
42515	Cherrapunjee	Meghalaya	25 °15' 0" N (25.25 °)	91 °44' 0" E (91.7333 °)	1300 m (4265 ft)	11759
42516	Shillong	Meghalaya	25 °34' 0" N (25.5667 °)	91 °53' 0" E (91.8833 °)	1600 m (5249 ft)	2296
42619	Silchar	Assam	24 °45' 0" N (24.75 °)	92 °48' 0" E (92.8 °)	21 m (70 ft)	3302.1
42623	Imphal	Manipur	24 °45' 36" N (24.7599 °)	93 °53' 48" E (93.8967 °)	774 m (2539 ft)	1454.6

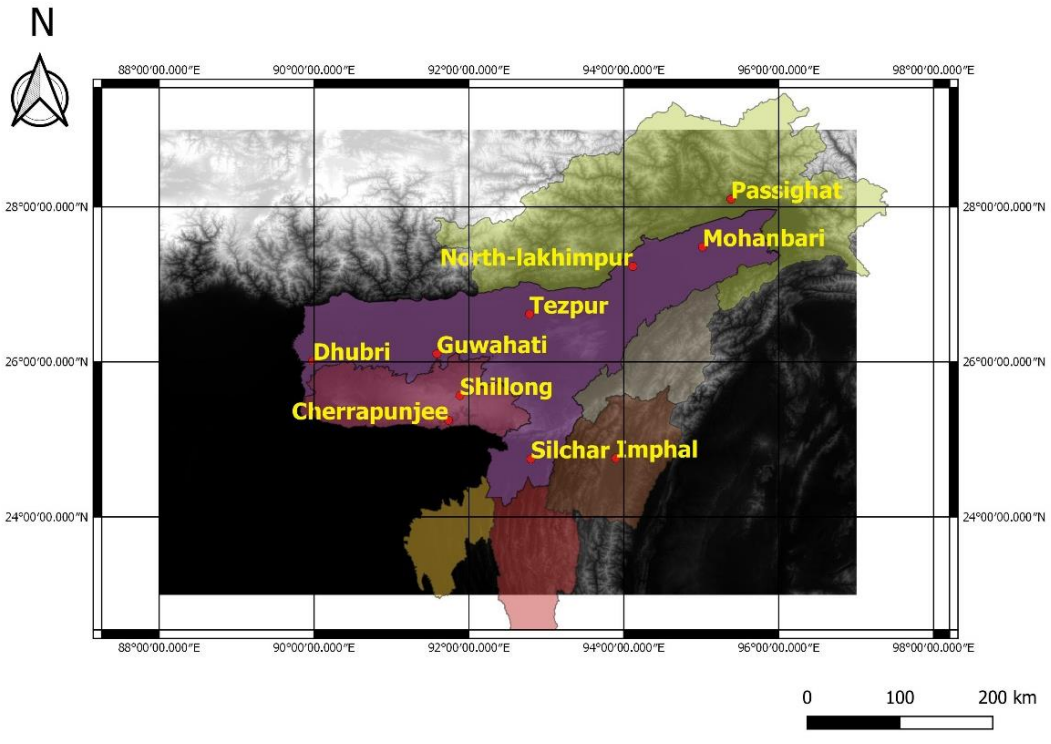


Figure 3. 2 Location of the IMD stations

3.3.2. Global Climate model (GCM) data

The GCM data are downloaded from Fourth Assessment Report (AR4) and Fifth Assessment Report (AR5) of Intergovernmental Panel on Climate Change (IPCC). Three GCM models are used in the study: HadCM3, ESM2G and ESM2M.

3.3.2.1. HadCM3

HadCM3 stands for Hadley Centre Coupled Model version 3. It is a coupled atmosphere-ocean general circulation model (AOGCM) developed at the Hadley Centre in the United Kingdom Met Office (UKMO). Unlike earlier AOGCMs at the Hadley Centre and elsewhere, this model does not require flux adjustment. HadCM3 is composed of two components: the atmospheric model HadAM3 and the ocean model (which includes a sea ice model). Simulations often use a 360-day calendar, where each month is 30 days. The atmospheric model, HadAM3, is a grid point model and has a horizontal resolution of 3.75˚in longitude × 2.5˚in latitude, which gives a global grid of 96 × 73 grid points. This gives a resolution of approximately 300 km. There are 19 levels in the vertical. The ocean model has a resolution of 1.25˚× 1.25˚with 20 vertical levels. This model has been selected based on previous literatures available on climate change studies. Among the general circulation models (GCMs), available under AR4, HadCM3 has been vastly utilized by researchers for future predictions of climatic parameters like temperature and rainfall (Zhang, 2005; Hessami et al., 2008; Vinnarasi, 2012; Kumar et al., 2013). HadCM3 is one of the major models, used in the Third Assessment report also. This model has been efficiently applied in several studies carried out in the Brahmaputra basin (Deka and Sarma, 2011; Vinnarasi and Sarma, 2011; Vinnarasi and Sarma, 2012).

HadCM3 monthly data are available from 2000 to 2100. The emission scenario considered in the present study for analysis is A2, which is described in IPCC's Special Report on Emission Scenarios (SRES) as a scenario representing a very heterogeneous world with continuously increasing global population and regionally oriented economic growth. It considers the forcing effect of greenhouse gases and sulphate aerosol direct effect, which are based on IPCC SRES-A2 (Special Report on Emission Scenario A2). The list of available predictors under this GCM model are given in Annexure A.

3.3.2.2. GFDL-ESM

The Geophysical Fluid Dynamics Laboratory (GFDL) is a laboratory in the National Oceanic and Atmospheric Administration's (NOAA) Office of Oceanic and Atmospheric Research (OAR). It is one of seven NOAA Research Laboratories. GFDL has constructed NOAA's first ESMs to explore the interaction of Earth's biogeochemical cycles (including human actions)

with the climate system. These ESMs are based on a coupled atmospheric-oceanic circulation model which incorporates interactive biochemistry and carbon cycle, and can represent sea ice, land and iceberg dynamics.

The atmospheric component of the ESMs includes physical features such as aerosols (both natural and anthropogenic), cloud physics, and rainfall. The land component includes rainfall and evaporation, streams, lakes, rivers, and runoff as well as a terrestrial ecology component to simulate dynamic reservoirs of carbon and other tracers. The oceanic component includes features such as free surface to capture wave processes; water fluxes, or flow; currents; sea ice dynamics; iceberg transport of freshwater; and a state-of-the-art representation of ocean mixing as well as marine biogeochemistry and ecology.

ESM2G and ESM2M, developed by the Geophysical Fluid Dynamics Laboratory (GFDL), are two mostly used models from the project. Lot of studies could be found which have successfully utilized these models for regional scale climate change studies in India (Singh and Goyal, 2016; Shivam et al., 2017a, 2017b; Singh et al., 2019; Sharma and Goyal, 2020). Jena et al. (2015) statistically attempted to select optimum models in the CMIP5 dataset for climate change projections of Indian monsoon rainfall. They stated that GFDL-ESM2G, along with GFDL-CM3 and two other models, best capture the pattern in Indian summer monsoon rainfall over the historical period (1871–2005).

ESM-2M and ESM-2G

GFDL has developed two new ESM models: ESM-2M and ESM-2G, under Intergovernmental Panel on Climate Change (IPCC)'s Coupled Model Intercomparison Project phase 5 (CMIP5). The models differ mainly in the physical ocean component. In one model, ESM2M, pressure-based vertical coordinates are used along the developmental path of GFDL's Modular Ocean Model version 4.1. In the other, ESM2G, an independently developed isopycnal model using the Generalized Ocean Layer Dynamics (GOLD) code base was used. Comparison between these two models allows us to assess the sensitivity of the coupled climate-carbon system to our assumptions about ocean formulation. While the models demonstrate similar overall scale fidelity, they have important differences in both their thermocline characteristics, deep circulation, ventilation patterns and El Nino variability that suggest critical roles for details of ocean configuration in the coupled carbon climate system.

These two models consider Representative Concentration Pathways (RCPs) to project the future climate change impacts. These RCPs represent four greenhouse gas concentrations

(instead of emissions) for climate modelling, as adopted by IPCC's Fifth Assessment Report (AR5). Both the ESMs have simulated datasets for historical data (up to 2005) and 4 future scenario datasets for 4 RCPs (2006-2100) under CMIP5 protocol.

These two models, ESM-2M and ESM-2G are used in the present study. The analysis has been done using both historical series and RCP8.5. The reason for considering RCP8.5 for the study is that, SRES A2 has similar trajectory to RCP8.5 with both reaching about 8 Wm-2 by 2100, as per IPCC AR5 report. Hence, it will be useful in comparing the resulting clusters obtained from both sets of data. The list of available predictors under ESM2G and ESM2M are given in Annexure A.

3.3.3. Daily gridded rainfall data

Daily gridded rainfall data are collected from IMD, Pune. The time period of data collection is from 01-01-1971 to 31-12-2005. The data are available for whole Indian sub-continent with resolution as high as $0.5^{\circ} \times 0.5^{\circ}$. Data of the north-eastern part, that is required for the analysis, are extracted from the whole data.

3.3.4. Rainfall data, collected from various tea gardens of Assam

A total of 83 raingauge stations with observed periods of varying number of years (from 5 years daily data to as long as 84 years and upto year 2016) were selected from various tea gardens of Assam. A total of 105 station rainfall data are found which are scattered all over Assam, out of which 83 stations are found to be situated in the upper Brahmaputra valley region. These rainfall data were collected from the selected 83 tea gardens. The location (longitude and latitude), total annual rainfall (in mm) and elevation (in m) of the raingauge stations are shown in the Table 3.2. Elevation of each station was determined from Shuttle Radar Topography Mission's (SRTM) digital elevation model (DEM) data of 30m resolution.

The locations of the raingauge stations are shown in the Fig. 3.3. It is situated in the upstream part of Brahmaputra River in the state of Assam, covering the central Brahmaputra valley and the eastern Brahmaputra valley region. It covers most of the upper and middle Assam districts. IT is surrounded by Bhutan in the north, Arunachal Pradesh in the north and east, Nagaland and Karbi-Anglong hills in the south. The area lies between 25.9210 N and 27.6190 N latitudes

and 91.8960 E and 95.7680 E longitudes. For this particular sub-region, total annual rainfall varies from 859.175 mm to 3411.875 mm and the elevation varies from 67 m to 427 m.

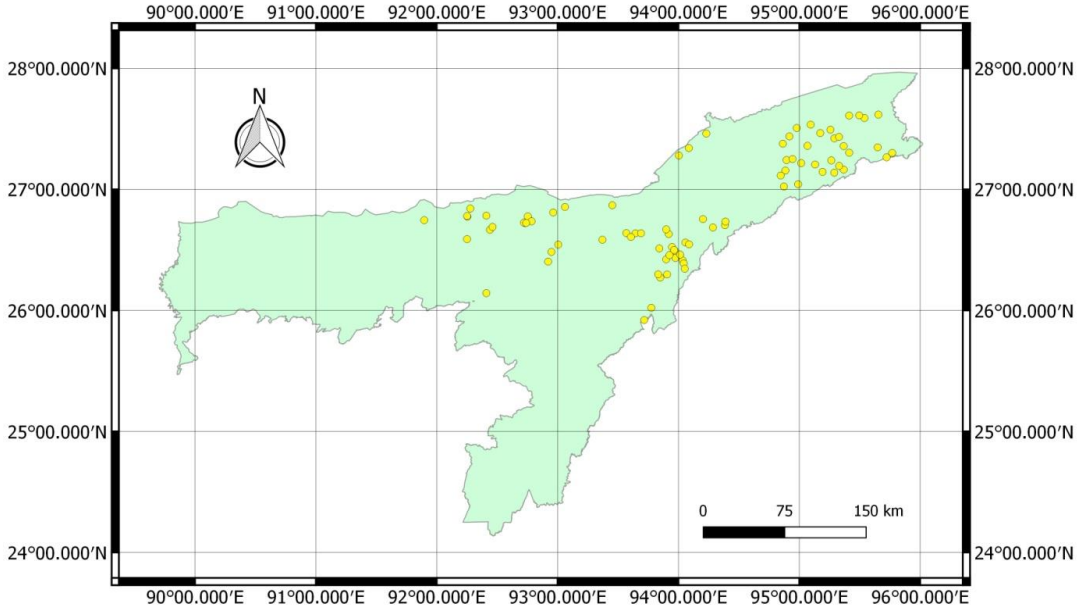


Figure 3. 3 Raingauge stations located in various tea gardens of the upper Brahmaputra valley region

Table 3. 2 Location details of the raingauge stations in various tea gardens of the upper Brahmaputra valley region with total annual rainfall

Sl. No.	Name of the station	Longitude (^{0}E)	Latitude (^{0}N)	Total annual rainfall in mm	Elevation in m	Sl. No.	Name of the station	Longitude (^{0}E)	Latitude (^{0}N)	Total annual rainfall in mm	Elevation in m
1	Abhoijan	93.899	26.423	2316.618	121	**43**	Keyhung	95.328	27.435	2457.185	127
2	Achabam	95.264	27.240	2311.642	123	**44**	Khowang	94.894	27.243	2095.938	103
3	Amsoi	92.410	26.143	1546.093	67	**45**	Koomsong	95.654	27.619	2692.581	144
4	Anand	94.230	27.462	3224.776	112	**46**	Kopati	92.251	26.590	1685.350	75
5	Arin	93.975	26.488	1679.050	100	**47**	Lakwa	94.872	27.024	2415.756	100
6	Arun	92.440	26.667	1563.424	80	**48**	Lamabari	92.278	26.843	2166.459	134
7	Athabari	94.035	26.414	996.060	104	**49**	Ledo	95.768	27.302	1333.175	153
8	Azizbagh	95.130	27.206	2642.110	113	**50**	Lengeree	93.717	25.921	1354.273	154
9	Bahani	94.203	26.756	1984.807	93	**51**	Lepetkata	94.864	27.378	2402.118	104
10	Basmatia	95.067	27.361	2252.881	113	**52**	Madhuting	95.367	27.358	896.679	129
11	Bateli	92.254	26.776	1711.969	112	**53**	Mahalakshmi	93.061	26.856	1473.345	88

Sl. No.	Name of the station	Longitude (^{0}E)	Latitude (^{0}N)	Total annual rainfall in mm	Elevation in m	Sl. No.	Name of the station	Longitude (^{0}E)	Latitude (^{0}N)	Total annual rainfall in mm	Elevation in m
12	Bhelaguri	94.385	26.705	1654.099	116	**54**	Maijan	94.978	27.508	2628.565	111
13	Bokajan	93.775	26.022	1084.235	136	**55**	Mancotta	94.918	27.439	3411.875	108
14	Bokakhat	93.644	26.638	1968.237	87	**56**	Mazbat	92.252	26.781	2160.181	113
15	Borahi	94.990	27.043	2488.364	104	**57**	Moran	94.885	27.156	2145.734	103
16	Borchapori	93.690	26.638	2174.268	92	**58**	Murphuloni	93.925	26.458	1694.360	106
17	Borhat	95.289	27.139	859.175	122	**59**	Nahorjan	93.607	26.608	1812.100	98
18	Borjan	94.058	26.562	1447.944	104	**60**	Namburnodi	93.832	26.299	1505.916	134
19	Borpathar	93.849	26.272	1430.350	127	**61**	Namdang	95.722	27.266	2929.631	427
20	Chubwa	95.173	27.466	2980.904	118	**62**	Namrup	95.329	27.195	1465.176	125
21	Cinnatolliah	94.086	27.342	3061.256	123	**63**	Nitinnagar	94.052	26.345	1112.845	107
22	Dalowjan	93.975	26.434	1555.215	105	**64**	Nonoi	92.922	26.404	1993.879	79
23	Deamoolie	95.539	27.591	2477.556	134	**65**	Ouphulia	95.015	27.218	2638.972	109
24	Dejoo	94.003	27.281	3001.704	123	**66**	Paneery	91.896	26.746	2254.247	122
25	Dekorai	92.964	26.810	2347.238	83	**67**	Panitola	95.257	27.494	2114.227	124
26	Deohall	95.289	27.422	2344.418	125	**68**	Pavoijan	93.906	26.298	1343.575	116
27	Dhekiajuli	92.462	26.690	2109.710	77	**69**	Powai	95.648	27.348	2328.983	158
28	Dholaguri	93.842	26.513	1622.052	99	**70**	Rungamatty	93.899	26.670	1824.846	90
29	Digulturrung	95.412	27.610	1903.000	126	**71**	Rupai	95.495	27.612	2608.085	132
30	Dilli	95.367	27.164	2876.955	132	**72**	Rupajuli	92.722	26.726	2301.322	78
31	Diphloo	93.569	26.639	1821.861	84	**73**	Sagmootea	93.005	26.544	1850.731	92
32	Dooria	93.919	26.632	1675.026	97	**74**	Santi	95.413	27.304	2881.667	127
33	Duklingia	94.285	26.686	2514.777	112	**75**	Sepoi	92.410	26.784	2124.710	110
34	Durrung	92.730	26.723	1533.854	76	**76**	Sepon	94.846	27.115	2195.201	103
35	Furkating	94.014	26.462	1535.197	106	**77**	Sockeiting	94.087	26.547	1692.910	98
36	Halem	93.452	26.870	1856.142	90	**78**	Sonabheel	92.785	26.737	1921.136	76
37	Halmira	93.947	26.524	1687.894	101	**79**	Sundarpur	95.192	27.145	3046.350	113
38	Harchurah	92.754	26.778	2282.480	94	**80**	Teloijan	94.944	27.251	1537.104	103
39	Hatigarh	94.043	26.389	1589.608	102	**81**	Tezpore and gogra	92.740	26.723	2019.596	79
40	Hatikhuli	93.371	26.585	1870.303	87	**82**	Thanai	95.094	27.536	2116.391	116
41	Kakojan	94.389	26.735	1517.938	107	**83**	Uday jyoti	93.964	26.500	1343.675	99
42	Kellyden	92.949	26.484	1700.084	86						

3.4. RESULTS AND DISCUSSION

It was realized from previous studies that various approaches can be used to perform homogeneous clustering. In most of the cases, researchers have used directly the rainfall (or rainfall) data for the purpose. Hence this has been considered as a conventional way of regionalization in the present study. Afterwards, a considerable number of studies have been carried out to identify homogeneous rainfall regions with various methods using parameters other than rainfall (Satyanarayana and Srinivas, 2011; Darand and Mansouri Daneshvar, 2014). Those are considered as non-conventional methods here.

To understand the variations among these conventional and non-conventional approaches and to assess their applicability in the study area considered for this particular study, a number of approaches have been tried and analyzed. After that, the results were compared to find out the best approach that can be used in the north-eastern region.

Various approaches that are taken into account in this study, while performing the regionalization using FCM analysis are:

- **Non-conventional regionalization, using interpolated GCM data**
- **Non-conventional regionalization, using nearby GCM grid point data**
- **Conventional regionalization, using rainfall data**
- **Regionalization, using gridded rainfall data**
- **Regionalization using both rainfall and GCM data along with location parameters (Latitude, longitude and elevation)**

During the non-monsoon period (November to February), very less to no rainfall occurs. Sometimes for a day or two, rainfall of very small magnitude occurs which doesn't have any significant effect over monthly rainfall values. However, the cumulative impact of these large quantity of small rainfall values reduces the peak values while averaged over a year, thus giving good correlation among the stations when annual data series is used for analysis, eventually affecting the homogeneous clustering. To take care of this drawback, regionalization was also done for monsoon data series (i.e. eliminating the data from November to February). Hence, the regionalization was done for both annual rainfall data and monsoon rainfall data.

Furthermore, for the initial three approaches, the clustering or regionalization was done mainly in two ways: (i) considering the effect of one station rainfall data, and (ii) considering the effects of three stations rainfall data (the station under consideration and two nearby stations).

The latter was considered because it was assumed that, the rainfall (and the LSAVs affecting the rainfall) of neighbouring stations may also influence the rainfall of a particular station to some extent. To find the influencing stations, geographical location of the stations was considered. However, this method was not found to be necessary for fourth approach (gridded rainfall data), hence not used.

The combinations of these approaches are shown in Table 3.3.

Table 3. 3 Various approaches while performing regionalization using FCM analysis

Sl No	Data used	Cases	Data series	
1	With GCM (HadCM3-A2) grid data (interpolated to station)	1 station effect	Annual	Monsoon
		3 station effect	Annual	Monsoon
2	With GCM (HadCM3-A2) grid data (taking nearest grid point)	1 station effect	Annual	Monsoon
		3 station effect	Annual	Monsoon
3	With IMD station rainfall data	1 station effect	Annual	Monsoon
		3 station effect	Annual	Monsoon
4	With IMD gridded rainfall data	1 station effect	Annual	Monsoon
5	With New RCP grid data (interpolated to station)	ESM2G (RCP8.5)	Annual	Monsoon
		ESM2M (RCP8.5)	Annual	Monsoon
6	With New RCP grid data (interpolated to station) along with location parameters (Latitude, longitude and elevation) and rainfall data	ESM2G (RCP8.5)	Annual	Monsoon
		ESM2M (RCP8.5)	Annual	Monsoon

3.4.1. Fuzzy clustering approach for initial data organization (training the input data)

Before applying the FCM algorithm, it is necessary to organize the input data. There are two major input matrices to the algorithm: data matrix $\boldsymbol{X}$ and the initial fuzzy partition matrix $\boldsymbol{U}$.

Output of the FCM algorithm depend on these two main inputs, hence proper training of input data is required. Again, there are two different sets of data that needs to be used as input for clustering: (i) rainfall data, and (ii) variables affecting the rainfall data. It was mentioned earlier (section 3.2.2) that these variables may include LSAVs of GCMs or their principal components (PCs), location parameters (latitude, longitude and altitude), and seasonality measures (maximum, minimum, standard deviation of rainfall etc.). Therefore, different strategies are needed to determine the input matrices for different FCM approaches mentioned above. The strategies are explained below:

3.4.1.1. Regionalization Using GCM Data

1. Selection of Attributes

The first work is to find out the LSAVs, which influence the rainfall of a particular station. For doing this, a relation has to be established between the observed rainfall data and the LSAVs. In the present work, Pearson Correlation is considered to be an effective way to find out the relation. Hence, for each station, Pearson co-efficients were calculated using the data available. It has been observed from the correlation results that rainfall data of different stations have different correlation with different LSAVs. Hence those LSAVs, which have good relations (more than 55%) with rainfall data of most of the stations, are selected as attributes. To reduce the curse of high dimensionality, mean monthly values of each of the selected LSAVs were computed at each of the stations. The mean monthly value of a variable denotes the average value of the variable computed for the month, for all years of the historical record. Thus, there are 12 mean monthly values for each LSAV at each station.

2. Principal Component Analysis

Since several of the atmospheric variables are correlated to each other, hence to avoid redundancy, Principal Component Analysis (PCA) has been done. Principal Component Analysis (PCA) is a method of dimensionality reduction without much sacrificing the accuracy. PCA aims to summarize data with many independent variables to a smaller set of derived variables, in such a way, that the first component has maximum variance, followed by the second component and so on. In the present study, the Principal Components (PCs) which preserved more than 98% of the variance were extracted for further analysis. In the present work, PCA has been done with the help of MATLAB program. The output matrix of PCA has been used as the input data matrix $\boldsymbol{X}$ in the FCM algorithm for further analysis.

3. Selection of Initial Partition Matrix

To proceed with the FCM algorithm, initial fuzzy partition matrix ***U*** has to be initialized at first. In this study, this has been done with the help of geographical location of the stations and cross-correlation among the stations. Cross-correlations for all the stations have been determined for the followings:

- Cross-correlation for mean monthly rainfall,
- Cross-correlation for monthly maximum rainfall,
- Cross-correlation for monthly minimum rainfall,
- Cross-correlation for monthly dry period.

With a close observation of the cross-correlations and the locations of the stations, the partition matrix ***U*** is initialized. However, in case of gridded rainfall data, the ***U*** value is taken as a random input, since it is not feasible and necessary to determine the cross-correlation among all the grid points. Furthermore, randomness of the initial partition matrix ***U*** does not affect the output of the FCM algorithm due to its iterative nature.

4. Regionalization using FCM Algorithm

Using the initial partition matrix ***U*** and the input data matrix ***X*** (output matrix of PCA), regionalization i.e. homogeneous clustering has been done with the help of FCM algorithm. A MATLAB program is made for the FCM algorithm where ***U*** and ***X*** were used as inputs. The algorithm gives the final partition matrix as output.

3.4.1.2. Regionalization Using rainfall Data

The conventional method of using directly the rainfall data for regionalization is a little different than the previous method. Selection of attributes is not required in this case. Hence, input matrix of PCA was composed of monthly rainfall data instead of LSAV data. Thereafter, same strategy for PCA, ***U*** and FCM analysis has been followed.

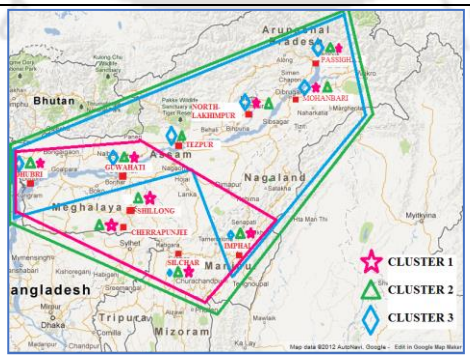
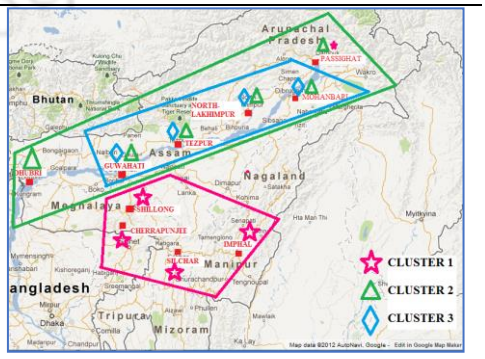
3.4.2. Regionalization using FCM algorithm: Results of various approaches used

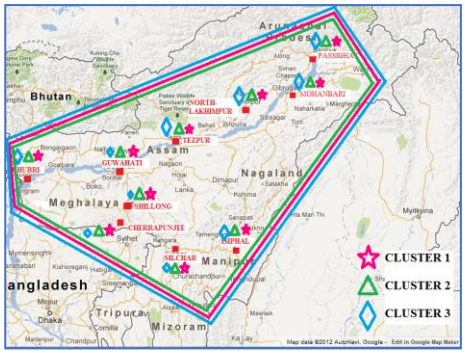
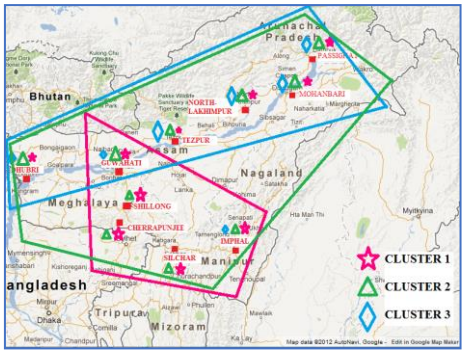
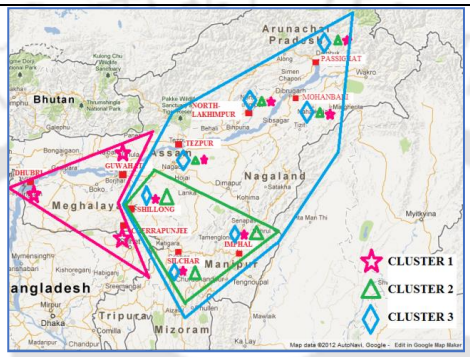
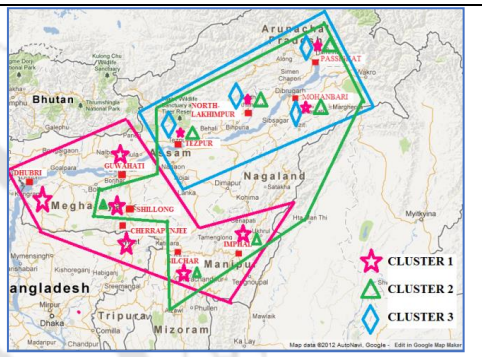
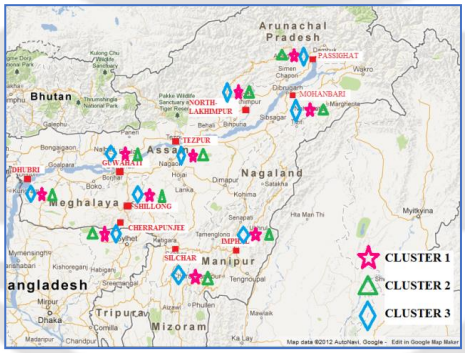
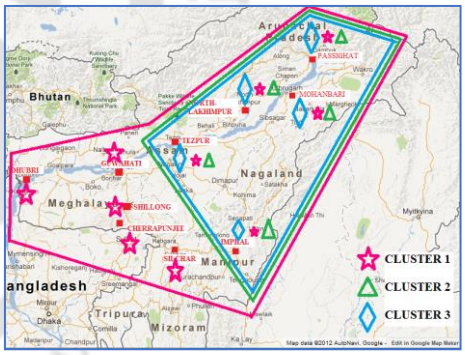
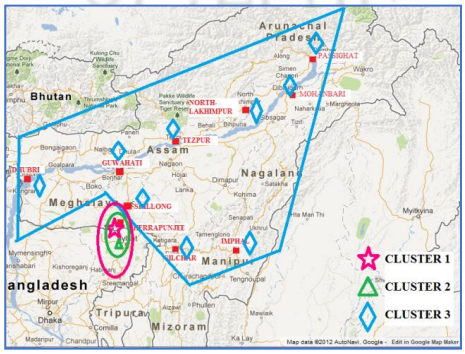
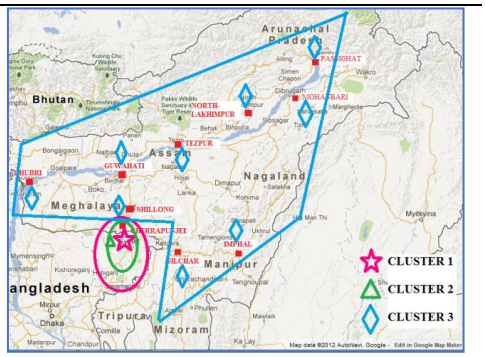
3.4.2.1. Regionalization of northeast Indian region

In the first four combination of approaches of regionalization (as mentioned in Table 3.3), GCM data (AR4) as well as observed rainfall data have been used. Pearson Coefficients were

calculated for all the stations to determine the most influencing LSAVs. The threshold values are taken as 55% and 50% for annual and monsoon data series respectively. In case of the first approach (interpolated GCM data), the 16 LSAVs selected for annual series are hur200, hur500, hur850, mrso, pr, prc, psl, ta200, ta500, ta850, tas, ts, ua200, ua500, ua850 and zg200. The 4 LSAVs selected for monsoon series are ta500, ua200, ua500 and zg200. In the second approach (nearest GCM grid data), the 13 LSAVs selected in the former case are hur200, hur500, hur850, mrso, psl, ta200, ta500, ta850, tas, ts, ua200, ua500 and zg200. The 4 selected LSAVs in the later case (monsoon series) are ta500, ua200, ua500 and zg200. PCA was done by using the GCM data of these selected LSAVs for respective cases. In the third and fourth approach (Table 3.3), observed rainfall data of each of the 10 stations are used for PCA. The Principal Components (PCs) which preserved more than 98% of the variance were extracted for further analysis.

Cluster number is another crucial parameter in FCM algorithm. In most of the cases, $c \leq N^{1/2}$ is suggested as the upper bound (Xie and Beni, 1991; Pal and Bezdek, 1995; Mok et al., 2012). Hence, with 10 stations, the cluster number comes out to be 3. Based on the reason that the total number of IMD stations was small and their locations are comparatively close to each other with respect to the grid size of GCM data, the cluster number is considered to be 3 for the analysis. In a similar way, optimum value of fuzzifier (***m***) also needs to be found out. Pal and Bezdek (1995) presented that FCM algorithm works well when ***m*** varies from 1–2.5. In the present cases, this value is taken as 2. The clusters formed after the analysis are shown in the Fig. 3.4.

Sl No.	Data used	Cases	Figures
1	With GCM (HadCM3-A2) grid data (interpolated to station)	1 station	 (a) Annual  (b) Monsoon

		3 station		
			(c) Annual	(d) Monsoon
2	With GCM (HadCM3-A2) grid data (taking nearest grid point)	1 station		
			(e) Annual	(f) Monsoon
		3 station		
			(g) Annual	(h) Monsoon
3	With IMD station rainfall data	1 station		

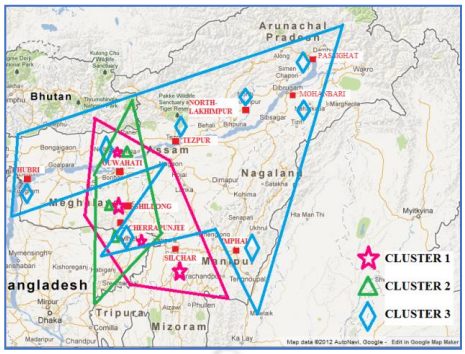
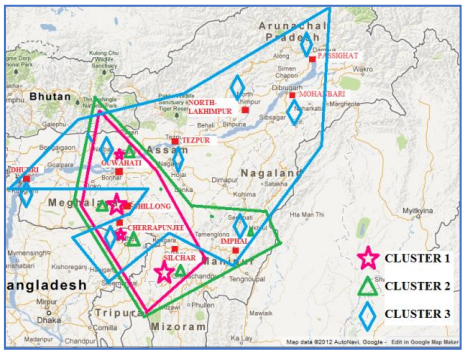
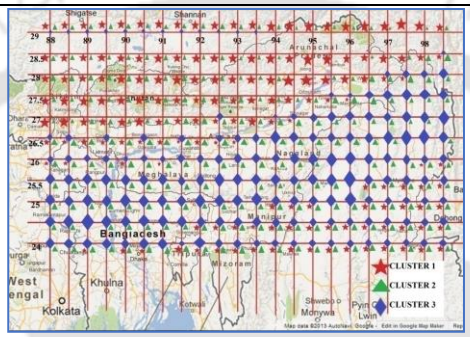
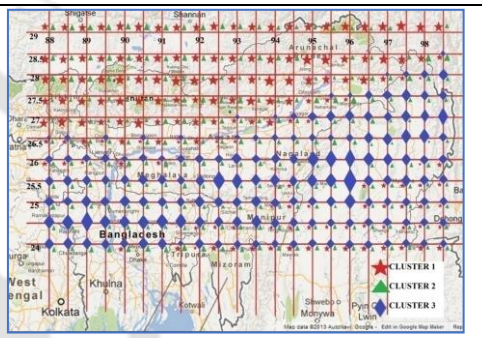
			(i) Annual	(j) Monsoon
		3 station		
			(k) Annual	(l) Monsoon
4	With IMD gridded rainfall data	1 station		
			(m) Annual	(n) Monsoon

Figure 3. 4 Clusters formed after Fuzzy Clustering Analysis in the northeast Indian region. Fig.3.4(a)-(h): Clusters obtained by using GCM data (IPCC, AR4); Fig.3.4(i)-(n): Clusters obtained by using rainfall data

From the results of the FCM analysis it was noticed that, different clustering has been found for different combinations. The clusters obtained for interpolated GCM data (Fig. 3.4(a),(b),(c),(d)) shows that annual data series gives a mixed clustering, whereas monsoon data shows that the stations in the Brahmaputra valley region are coming in one cluster and the remaining stations which are nearer to Barak valley region are coming under a separate cluster. When three stations interpolated GCM data are considered, annual data resulted into equal membership of all the stations to all the clusters. Whereas, monsoon data shows one visible cluster in Brahmaputra valley and one in the Barak valley region.

In the second approach (Fig. 3.4(e),(f),(g),(h)), where GCM data of nearest grid was considered, mainly two clusters are being formed: one on the east and another on the west side of the study area. However, the major impact of the selected grid point on clustering is clearly visible hear. The stations for which the nearest GCM grid point is same tend to show high membership in the same cluster. Even in case of monsoon data, the membership values of the

stations are affected by this. When three stations effect was considered, similar clustering pattern is attained for monsoon data. Whereas, annual data couldn't deliver any proper clustering result as all the stations showed equal membership to all the clusters.

In the conventional approach (Fig. 3.4(i),(j),(k),(l)), where rainfall data was used, it is seen that all the stations are coming under single cluster, except Cherrapunjee. This was true for both annual and monsoon data series. This can be attributed to the reason that, the average rainfall of Cherrapunjee is much higher than other 9 stations under consideration. This indicates the inability or limitation of conventional method of using rainfall data in clustering, since high magnitude of input data can hugely impact the clustering results. In case of three station effect, Cherrapunjee and its nearby stations are coming under one cluster, leaving the other station to form another cluster.

When gridded rainfall data was used for clustering (fourth approach), it was found that there are two clearly visible clusters (Fig. 3.4(m),(n)). The grid points on the Brahmaputra valley region are coming in the same cluster and the remaining grid points which are nearer to Barak valley region are coming under a different cluster. The results are same for both annual and monsoon data series. This is similar to the results obtained for interpolated GCM data for monsoon period.

Further analysis has been done using IPCC AR5 data (Table 3.3). Two models are considered here: ESM-2M and ESM-2G, under IPCC's Coupled Model Intercomparison Project phase 5 (CMIP5). The analysis has been done using RCP8.5. The reason for considering RCP8.5 is that in the previous cases, SRES A2 scenario has been considered for analysis. IPCC AR5 report suggests that, SRES A2 has similar trajectory to RCP8.5 with both reaching about 8 Wm^{-2} by the year 2100 (Fig. 3.5). Hence, it will be useful in comparing the resulting clusters obtained from both sets of data.

From the clustering results of previous approaches, it was evident that consideration of nearest stations' effect in the formation of the input data matrix is not very useful. In most of the cases, all the stations are found to be in the same cluster. Hence, it contradicts with the whole purpose of regionalization. Therefore, this condition was omitted in the rest two approaches. Furthermore, it was realized that rainfall alone is not effective if some stations have very high rainfall value compared to other stations. If the difference in rainfall magnitudes is very high, it affects the output clustering pattern. It was assumed that location parameters might be effective in getting satisfactory clustering results. Hence, in the sixth combination, a hybrid

approach is tried with new RCP grid data (interpolated to station) along with rainfall data and location parameters (Latitude, longitude and elevation). The clusters formed after the analysis are shown in the Fig. 3.6.

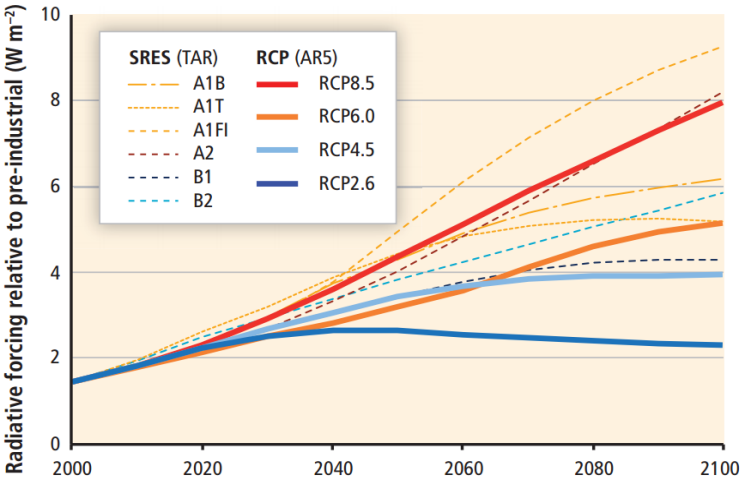
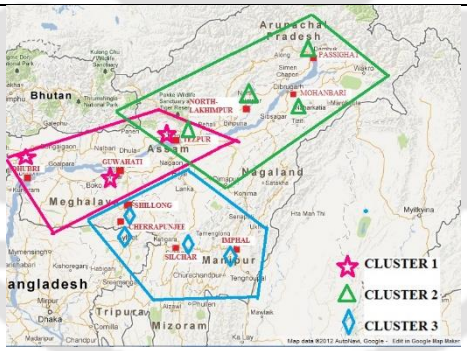
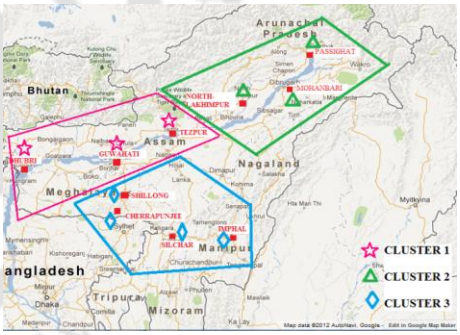
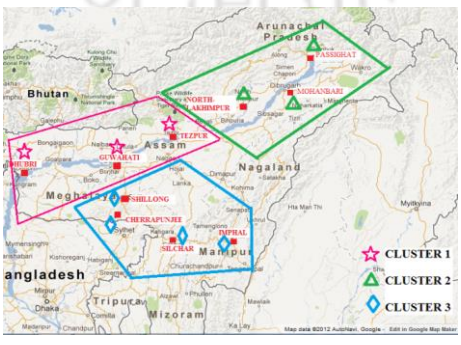
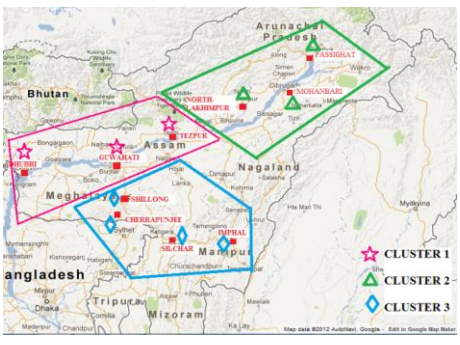


Figure 3. 5 Projected radiative forcing (RF, Wm^{-2}) over the 21st century using the Special Report on Emissions Scenarios (SRES) and Representative Concentration Pathway (RCP) scenarios [Reference: Fig. 1-4 (a) of IPCC AR5 WGII report, Chapter 1]

Sl No.	Data used	Cases	Figures
5	With New RCP grid data (interpolated to station)	ESM2G (RCP8.5)	 (a) Annual  (b) Monsoon
		ESM2M (RCP8.5)	 (c) Annual  (d) Monsoon

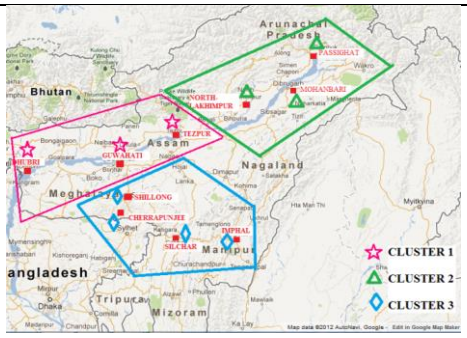
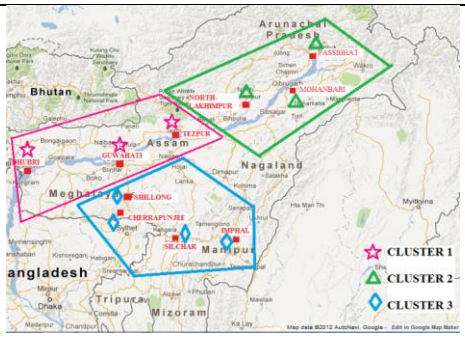
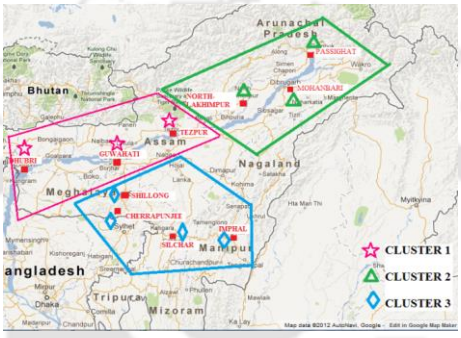
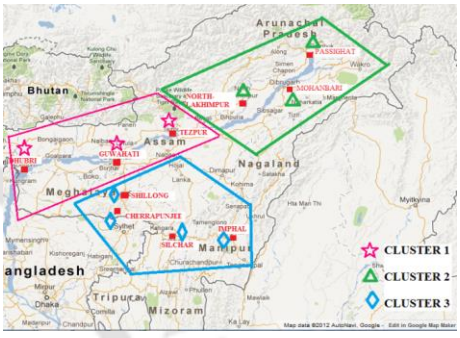
6	With New RCP grid data (interpolated to station) along with location parameters (Latitude, longitude and elevation) and rainfall data	ESM2G (RCP8.5)		
			(e) Annual	(f) Monsoon
		ESM2M (RCP8.5)		
			(g) Annual	(h) Monsoon

Figure 3. 6 Clusters formed after Fuzzy Clustering Analysis in the northeast Indian region, obtained by using GCM data (IPCC, AR5)

After the analysis it was found that, with RCP data of new GCM models, three clusters are being formed (Fig. 3.6(a),(b),(c),(d)). Guwahati, Dhubri and Tezpur are coming under the same cluster (**Cluster 1**). Passighat, North-Lakhimpur and Mohanbari are forming a different cluster (**Cluster 2**). Similarly, Shillong, Silchar, Imphal and Cherrapunjee are coming in a separate cluster (**Cluster 3**). Although, in case of ESM2G annual data series, Tezpur was showing almost equal membership to two clusters. In the sixth hybrid approach too, three clusters are being formed (Fig. 3.6(e),(f),(g),(h)). In this approach, all four cases are showing the same result.

The main difference of the results of these two approaches with the previous ones is that, previously two clusters could be confidently identified: one in Brahmaputra valley region and another in the Barak valley region. However, for some stations (Dhubri, Tezpur), it was difficult to come to a definite conclusion. This drawback was rectified in the sixth hybrid approach, where these stations are found to be having acceptable membership values with one

particular cluster. Hence, it verifies that use of rainfall data and location parameters i.e. Latitude, longitude and elevation, along with GCM data (interpolated to station) can give a proper clustering result, eventually providing an acceptable outcome for homogeneous regionalization.

3.4.2.2. Regionalization (sub-clustering) of Upper Brahmaputra valley region using rainfall data of the tea gardens

It was established in the previous section that hybrid method of using location parameters and LSAVs of GCM models, along with rainfall data can deliver satisfactory results for homogeneous regionalization. However, if further clustering needs to be done at a local scale, this approach again shows some limitations. There are two major drawbacks that can be found in the previous analysis as given below.

1. **Use of limited number of rainfall stations for analysis:** Due to the use of a smaller number of stations, the clusters formed after analysis are also less in number. As mentioned earlier, upper bound of cluster number c is $N^{1/2}$, N being the number of stations (Xie and Beni, 1991; Pal and Bezdek, 1995; Mok et al., 2012). Hence, with 10 stations, the maximum number of stations that we can get is 3.
2. **Coarse resolution of GCM grid data:** The grid size of GCM data is generally of coarse resolution. HadCM3 has a grid size of 3.75˚ in longitude × 2.5˚ in latitude, whereas the grid size of ESM2G and ESM2M is 2.02˚ in longitude × 2.5˚ in latitude. This size of grid covers a vast region. If the whole region falls inside only grid, interpolation of GCM data would be done from the same four grid points. Hence, for a small area it may not be very effective to use interpolated GCM data.

To fix this issue and to give a suitable regionalization approach for geographically small region, further analysis was done by utilizing rainfall data, collected from 83 tea gardens of Assam (Fig. 3.3). These stations are spread over the region covered by **Cluster 2** in the previous analyses (Fig. 3.6(a)). In this case, total annual rainfall, total monthly rainfall, standard deviation of total annual rainfall and all three location parameters (latitude, longitude and elevation) of 83 raingauge stations were included as attributes, instead of using only rainfall data. Three combinations were tried in this study to observe the effect of a particular attribute in the cluster formation. Those are:

Case 1: Input data matrix with total monthly rainfall as attributes,

Case 2: Input data matrix with standard deviation of total monthly rainfall as attributes,

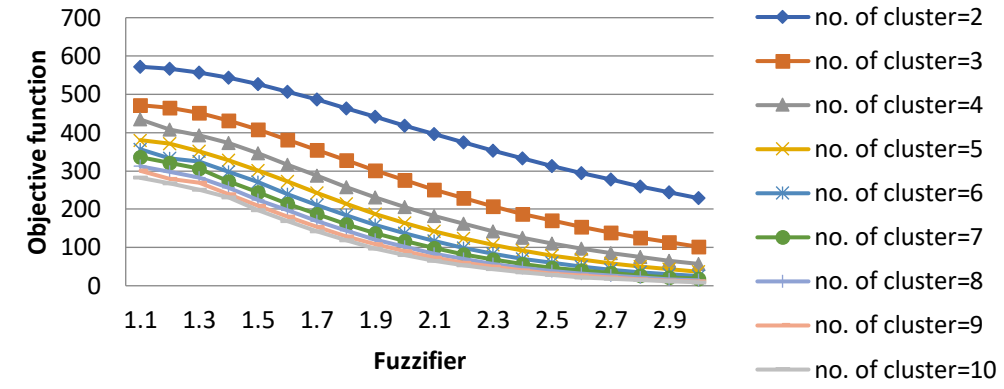
Case 3: Input data matrix with latitude, longitude, elevation, total annual rainfall and standard deviation of total annual rainfall as attributes.

The results are explained in detail in the subsections below:

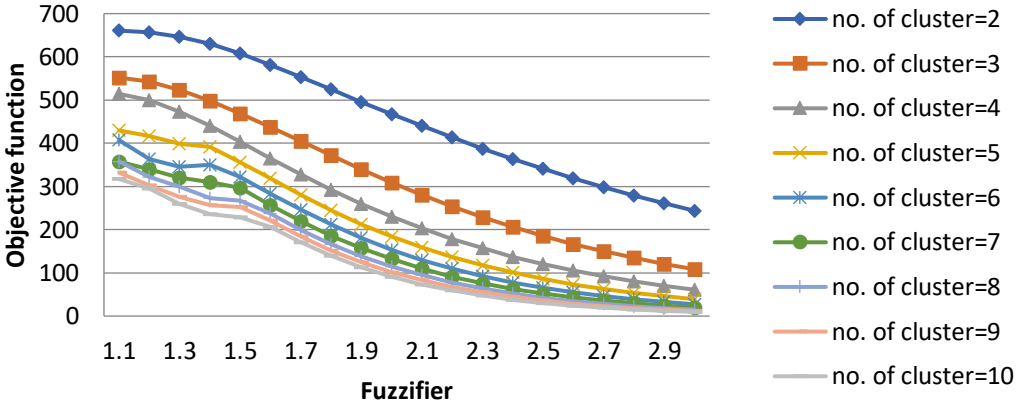
(i) Cluster analysis with Fuzzy C-Means (FCM) algorithm

To acquire reliable results from clustering, most crucial point is to assume the cluster number (c), since the number of regions is unknown beforehand. The best way to do that is to choose a range of values for c, and then to find out the most appropriate one. To achieve that, the cluster number c is changed from 2 to k, k being a quantity lesser than total number of sites, N. The lower bound of c is taken as 2, because the dataset is apparently clustered into more than one group. The interesting point here is to define the upper bound of c, i.e. the parameter k. Varying the k value impacts the reliability of cluster number. From research, it was noticed that increasing the k value generates more consistent cluster number division (Mok et al., 2012). Many works can be found in literature on identification of optimal cluster numbers. In most of the cases, $k \leq N^{1/2}$ is suggested as the upper bound (Xie and Beni, 1991; Pal and Bezdek, 1995; Mok et al., 2012). In a similar way, optimum value of fuzzifier (m) also needs to be found out. Pal and Bezdek (1995) presented that FCM algorithm works well when m varies from 1–2.5. It is hence suggested (Satyanarayana and Srinivas, 2011) to attain a number of sets of clustered regions by selecting a range of values for c and m, and then identify the final clustered regions based on the optimal values of c and m and by means of CVs.

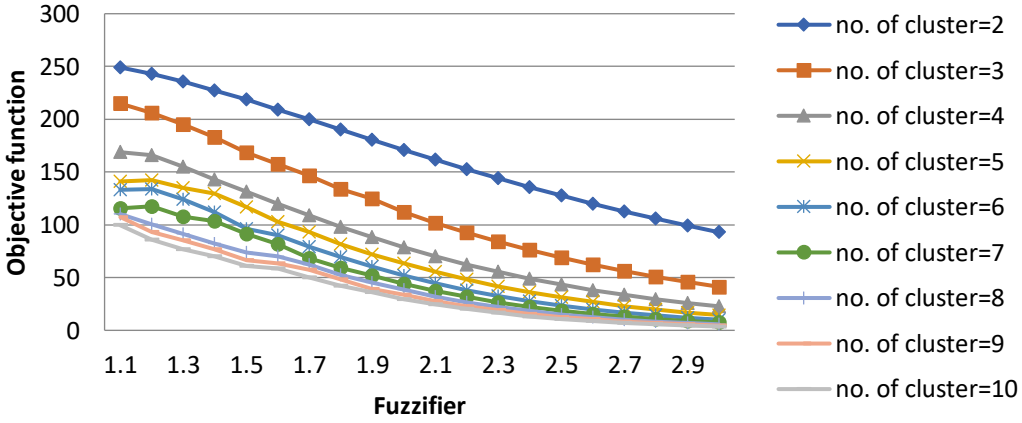
In this study, FCM algorithm was executed for each case with cluster number varying from $c_{min} = 2$ to c_{max} *or* $(k) = (83)^{1/2} \approx 10$, with increment 1. The fuzzifier value (m) is increased from 1.1 to 3.0, with increment 0.1. Change in the values of objective functions with respect to that of fuzzifier m were plotted for cluster number varying from 2 to 10, for all the three cases and are shown in Fig. 3.7(a), 3.7(b) and 3.7(c). It is observed that, as the fuzzifier value increases the optimum value of objective function declines, for a given cluster number. In a similar way, the optimum value decreases with rise in cluster number, for a given value of fuzzifier. The FCM algorithm was found to perform better for m in the range of 1.5–2.5.



(a) Case 1



(b) Case 2



(c) Case 3

Figure 3. 7 Variation in the optimal value of objective function of FCM algorithm with variation of fuzzifier *m* and cluster number *c*, for (a) Case 1: with total monthly rainfall as attributes; (b) Case 2: with standard deviation of total monthly rainfall as attributes; and (c) Case 3: with latitude, longitude, elevation, total annual rainfall and standard deviation of total annual rainfall as attributes

Further, the CVs described in section 3.2.4, were calculated to achieve the optimal value of c and m, for each case. The values of V_{PC}, V_{PE}, V_{FPI}, V_{NCE}, V_{XB}, V_{FS}, and V_K are tabulated in Annexure B. In all the three cases, it is observed that, V_{PC} decreases monotonously whereas V_{PE}, V_{FPI} and V_{NCE} show increase monotonously with increase in the fuzzifier value, m, for a given value of cluster number c. Moreover, they show overall monotonous decreasing (V_{PC}) and increasing (V_{PE}, V_{FPI} and V_{NCE}) tendency with increase in the cluster number c, hence giving always low value of c and m as the optimum set of values for homogeneous clustering. This indicates their ineffectiveness in determining optimum number of homogeneous rainfall regions.

Extended Xie and beni index (V_{XB}) as well as the Kwon index (V_K) were found to be relatively effective in these cases. They didn't show any monotonous increasing or decreasing tendency with change in the values of c and m, and were in line with each other. In case of Fukuyama and Sugeno index (V_{FS}), although it didn't have any monotonic tendency, the values were found to be in a conflict to the values of V_{XB}, and V_K in some cases. Taking into account of the similar studies on homogeneous rainfall region identification, found in literature (Rao and Srinivas, 2006a; Satyanarayana and Srinivas, 2011; Farsadnia et al., 2014), V_{XB} and V_K were opted for clustering of rainfall regions. Henceforth, the clusters found from these two indices were considered for further analysis. The optimal value of clusters c=3 was identified, using these CVs with the corresponding value of m=1.5 for all three cases.

(ii) Homogeneity test and adjustment of heterogeneous clusters

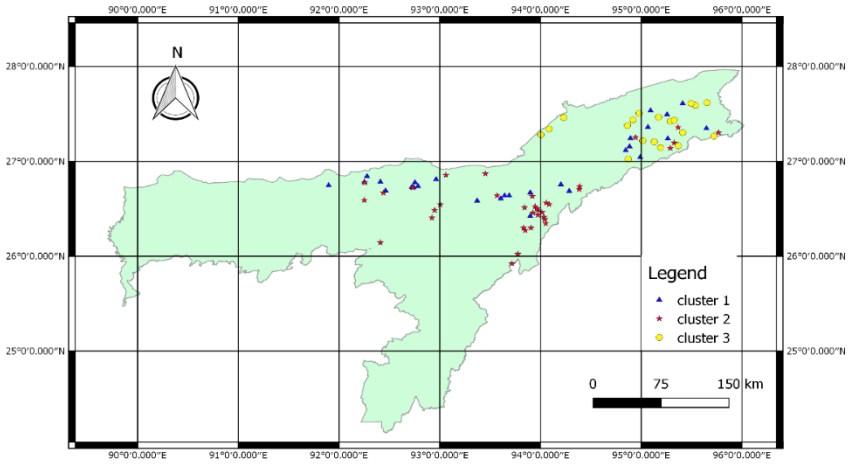
Since clusters resulting from clustering algorithms are not utterly homogeneous, further adjustments are required to come to a definite conclusion. To process with that, at first it is required to identify the homogeneity of the clustered regions obtained from the algorithm, by performing discordance measures and heterogeneity measures, proposed by Hosking and Wallis (1993, 1995). Discordancy tests give the D_i values for the stations in each cluster, which reveals the discordant stations in a particular cluster. H_1, H_2 and H_3 values are calculated for each of the clusters to determine the heterogeneity level within the cluster. After finding these values, some adjustments may be executed to decrease the value of heterogeneity measure, till it becomes acceptably homogeneous.

The H values of homogeneity test for all the three cases are given in Table 3.4. Initially, out of the nine clusters (three clusters for each case), four clusters were acceptably homogeneous, four clusters were possibly heterogeneous and one cluster was definitely heterogeneous. Hence,

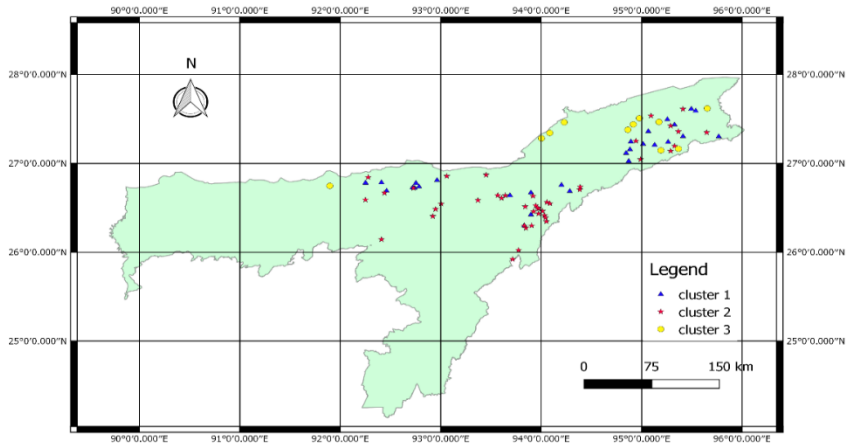
adjustments were done to make the possibly and definitely heterogeneous clusters as homogeneous, by either shifting discordant sites from one cluster to another or by removing those sites if necessary. For case 1, stations Chubwa and Dilli were found to discordant with all the clusters; hence they are removed during adjustment which lead to three acceptably homogeneous clusters with the rest 81 stations. Similarly, station Chubwa was found to be discordant for case 2; hence removed from the clusters and adjustments were done for 82 stations. However, in case 3, no stations were removed. The reason that the two stations did not show any discordancy in case 3 is because the geographic locations of the stations were considered during cluster analysis for case 3. The H values of homogeneity test after adjustment and number of sites in each cluster are shown in Table 3.4. Final clusters formed after adjustment are shown in Fig. 3.8(a), 3.8(b) and 3.8(c).

Table 3. 4 Homogeneity test results

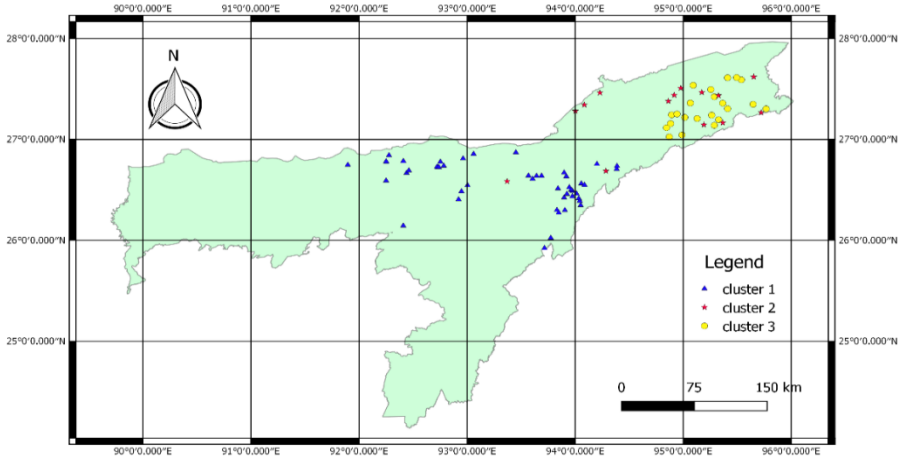
Case	Cluster no.	H test			Homogeneity result	After adjustment			Homogeneity result	Final no of sites
		H1	H2	H3		H1	H2	H3		
1	C1	-0.08	-1.25	-0.99	Acceptably homogeneous	0.68	-0.97	-1.06	Acceptably homogeneous	31
	C2	1.08	-0.69	1.1	Possibly heterogeneous	0.49	-1.04	0.9	Acceptably homogeneous	34
	C3	2.96	1.11	0.46	Definitely heterogeneous	-0.84	0.56	-0.05	Acceptably homogeneous	16
2	C1	-0.98	-0.73	0.04	Acceptably homogeneous	-0.91	-0.62	0.1	Acceptably homogeneous	31
	C2	0.54	-1.25	0.09	Acceptably homogeneous	0.41	-0.71	0.39	Acceptably homogeneous	44
	C3	1.52	0.86	0.69	Possibly heterogeneous	0.41	-1.23	-1.28	Acceptably homogeneous	7
3	C1	1.86	0.91	0.15	Possibly heterogeneous	0.71	0.85	-0.11	Acceptably homogeneous	13
	C2	-0.88	-2.24	-0.25	Acceptably homogeneous	0.73	-1.95	0.16	Acceptably homogeneous	49
	C3	1.62	-0.05	-0.16	Possibly heterogeneous	0.69	-0.86	-0.93	Acceptably homogeneous	21



(a) Case 1



(b) Case 2



(c) Case 3

Figure 3. 8 Clusters formed by the FCM algorithm after adjustment for (a) Case 1: with total monthly rainfall as attributes; (b) Case 2: with standard deviation of total monthly rainfall as attributes; and (c) Case 3: with latitude, longitude, elevation, total annual rainfall and standard deviation of total annual rainfall as attributes

From the analysis it has been found that, in the first two cases clusters are scattered over the whole region (Fig. 3.8(a),(b)). This was due to the consideration of only rainfall data for clustering. Since the average value of rainfall of most of the stations are in the same range, hence the result. In the third case (Fig. 3.8(c)), this scattering of clusters was taken care of by the location parameters, thus signifying the importance of location parameters in homogeneous regionalization. Although two clusters are overlapping with each other in the third case also. Hence, we can draw an overall conclusion from all the three cases that there are two clearly visible clusters, one on the east side and another on the west side of **Cluster 2** [refer Fig. 3.6(a)].

3.4.3. Comparison with similar studies done previously

Although this is the first and foremost study to divide the Brahmaputra valley region of India into hydrologically homogeneous regions with the use of rainfall data collected from the tea gardens by applying fuzzy clustering technique and seven cluster validity indices (CVs), few previous studies can be found on use of various clustering techniques for Indian subcontinent. Hence, comparisons have been made with those closely related studies to validate the performance of the present study. The homogeneous regions developed by Indian Meteorological Department (IMD) displays five large provinces, which are although delineated based on rainfall characteristics but are influenced by contiguity of area and administrative state boundaries. Iyengar and Basak (1994) used principal component analysis (PCA) for regionalization of Indian monsoon rainfall and recommended the PCA approach for further subdivision of the region. Ten homogeneous sequential regions were formed in India from their analysis, in which the stations of upper Brahmaputra valley regions were seemed to form similar kind of clusters as found in the present study, although few stations remained un-clustered. Singh and Singh (1996) have done regionalization of monthly as well as seasonal rainfall for sub-Himalayan areas and Gangetic plains, by using principal component analysis (PCA). They used rainfall data for a period of 114 years (1871-1984) from 90 well distributed stations which resulted into four distinct homogeneous rainfall areas for both monthly and seasonal scales. Srinivasa Raju and Nagesh Kumar (2007) utilized fuzzy cluster analysis (FCA) to classify 159 meteorological stations in India and concluded that FCA method performs well than the Kohonen Artificial Neural Networks (KANN) method in finding meteorologically homogeneous groups. They utilized location parameters (latitude, longitude and elevation) along with other meteorological parameters for clustering and the results exhibited 14 clusters over Indian region, the north-eastern region being in one cluster. Satyanarayana and Srinivas

(2008) have done regional frequency analysis using LSAVs that affects the rainfall in a region instead of observed rainfall data and have used K-means clustering with adjustments and L-statistics for regionalization. 17 homogeneous regions were formed after the analysis, two regions covering the north-eastern states. The upper Brahmaputra valley region came under the same homogeneous cluster, hence producing similar results to those of the present study. Satyanarayana and Srinivas (2011) have done regionalization of rainfall data, based on fuzzy clustering method by utilizing GCM data, location parameters and seasonal rainfall data. The stations of upper Brahmaputra valley regions were seemed to form similar kind of clusters as found in the present study. Stations in middle Assam were found to form one cluster while other stations on the upper Assam formed a different cluster. Saikranthi et al. (2013) used correlation analysis for regionalization based on seasonal and annual rainfall data. They used 51 years (1951–2001) daily rainfall data collected for more than 1000 rain gauges across India for the analysis, which produced 26 homogeneous rainfall zones. However, because of data scarcity north-eastern states were not included in the analysis. Bharath and Srinivas (2015) used wavelet-based global FCM analysis, instead of PCA for determining homogeneous hydrometeorological regions in India. The new approach proposed by them clustered the Indian territory into 29 regions, north-eastern region having 7 clusters. The clusters formed in the upper Brahmaputra valley region were similar to the present study. Kulkarni (2017) has used probability density function to divide the Indian subcontinent into homogeneous clusters, using daily summer monsoon rainfall at 357 square grids of size 10000 sqkm. The study produced five clusters, out of which one cluster covered adjoining regions and all other clusters were scattered indicating irregular behaviour of daily rainfall pattern in India. The study was done by using two time periods 1901–1975 and 1976–2010, and the resulting clusters were found to be extremely different in the two time periods. The clusters formed in the north-eastern region is also different for the two periods which are not entirely in line with the present study. However, they used gridded rainfall data instead of station data, thus the difference. (Mannan et al., 2018) have used climatic variables and self-organizing maps to regionalize India. Artificial neural network is used along with four CVs for clustering and applied on gridded rainfall dataset (0.25° × 0.25°) from IMD for 34 years (1980–2013) as well as climatic variables such as air temperature, surface pressure, geo-potential height, specific humidity, etc. 10 homogeneous regions were formed when only rainfall data was used, whereas incorporation of climatic variables divided the region into 15 regions. The region 2 in their study covered the northeast India with rainfall of 7.2 mm/day.

3.5. SUMMARY

The study was carried out in the north-eastern region of India. The north-eastern region of India is of great concern in the research field of climate change. This is the part of India which receives heavier rainfall than other parts of the subcontinent. High intensity rainfall for short duration and longer dry spell are the major problems due to impact of climate change. An attempt is made to delineate the region into some homogeneous clusters with respect to rainfall, by using fuzzy clustering approach. FCM algorithm was utilized for the same and various combinations were experimented to get an acceptable result for regionalization.

Following are the major outcomes of the analysis:

- Considering the effect of nearest stations' rainfall data in clustering doesn't serve the purpose. In most of the cases, all the stations are found to be in the same cluster. Hence, it contradicts with the whole purpose of regionalization.
- Rainfall data alone is not effective if some stations have very high rainfall value compared to other stations. If the difference in rainfall magnitudes among the stations is very high, it affects the output clustering pattern.
- Similarly, usage of LSAVs (from GCM data) alone is also not effective because of its coarse resolution.
- Inclusion of geographical location parameters (latitude, longitude and elevation) effectively reduces the above limitations, enhancing the chance of getting satisfactory clustering results.
- **Hence, a hybrid approach can be tried with GCM grid data (interpolated to station) along with rainfall data and location parameters (latitude, longitude and elevation).** For the present study area, hybrid approach gave the best result among all other approaches.
- Two major limitations that can be found in small scale clustering are-
    - Use of limited number of rainfall stations for analysis
    - Coarse resolution of GCM grid data.
- Since GCM data is not effective for small region, local rainfall data (with more station data) was used for regionalization, along with location parameters.
- **The best result was exhibited by the approach, in which total annual rainfall, standard deviation of total annual rainfall and location parameters (latitude,**

longitude and elevation) were used as attributes. Hence, this approach can be used for regionalization of small-scale areas.

- As far as the present study area is concerned, the above analysis exhibited three clusters in the northeastern region (hybrid approach used).
    - **Cluster 1:** Region covering the stations Guwahati, Dhubri and Tezpur
    - **Cluster 2**: Region covering the stations Passighat, North-Lakhimpur and Mohanbari
    - **Cluster 3**: Region covering the stations Shillong, Silchar, Imphal and Cherrapunjee.
- After further clustering of **Cluster 2** (small-scale regionalization approach was used), two sub-clusters could be found (Fig. 3.9):
    - **Cluster 2a:** region on the west side of cluster 2
    - **Cluster 2b:** region on the east side of cluster 2.

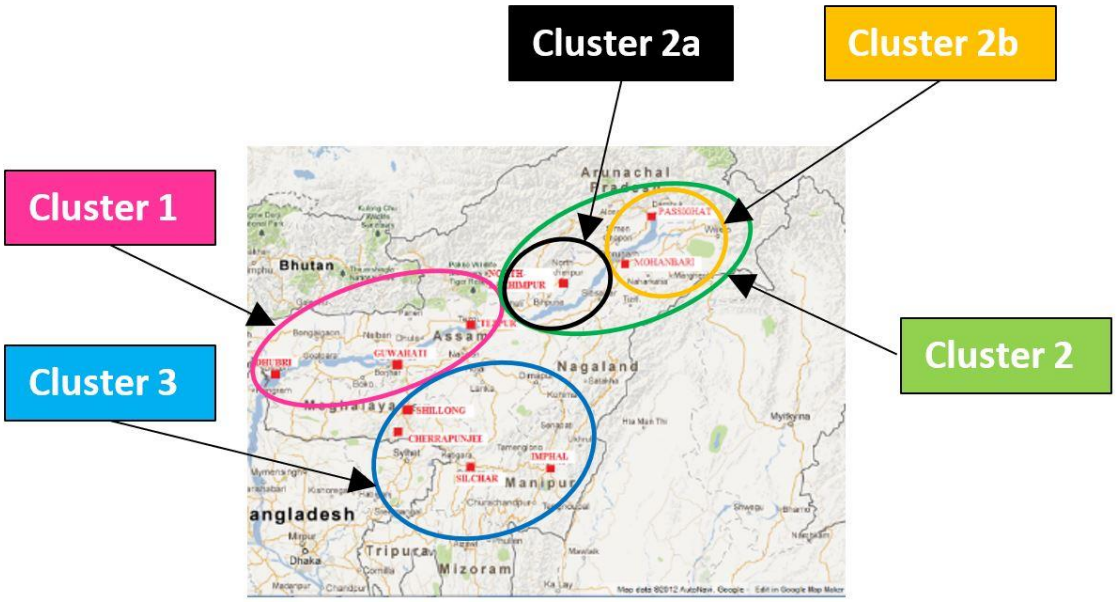


Figure 3. 9 Final clusters obtained in the northeast Indian region

Chapter 4

DEVELOPMENT OF STATISTICAL DOWNSCALING MODELS FOR PREDICTING TEMPERATURE AND RAINFALL FROM GCM OUTPUTS AND TO COMPARE THE RESULTS OF DIFFERENT MODELS WITHIN A CLUSTER

4.1. INTRODUCTION

Usually, GCMs have coarse resolution, covering large geographical areas. However, impact assessment models in many cases require information at finer scales of 50 km or less. To overcome this problem downscaling methods have been developed as these methods can help in obtaining high-resolution climate change information from relatively coarse-resolution global climate models (GCMs). It is assumed that these large-scale climate variables of GCM have significant influence on local scale climate.

Many studies and projects on climate change are already going on. Mehrotra and Mehrotra (1995) have studied the climate change and hydrology with emphasis on the Indian subcontinent. They mentioned that, the major impacts of climate change in India would be on the hydrology, water resources and agriculture of the country. Heyen et al. (1996) have done statistical downscaling of monthly mean air pressure to sea level anomalies. They indicated that increase of air pressure leads to decrease in the mean sea level. Wilby et al. (1998a,1998b) have investigated the empirical relation between climatic variables and local variable and compared different statistical downscaling methods. They found that NCEP reanalysis datasets have dissimilarity with GCM datasets. Eliasson (2000) suggested that the urbanization is always related to climate, i.e., urbanization induces climate change and the climate change causes more problem in urbanization. Gosain et al. (2006) have developed a hydrologic modelling for various river basins in India considering climate change effect. They used HadRM2 daily weather data to determine the spatial-temporal water availability in various river systems. Sharma and Shakya (2006) have analysed the hydrological changes in Bagmati watershed which has hydroelectric plant in upper region of river and agricultural fertile land in lower region. Due to the climate change, precipitation during monsoon decreased and pre and post monsoon precipitation increased, but mean yearly flow in river decreased. Mujumdar (2008) has presented an overview of the current scenario and recent work in India to assess the climate change impact on water resources. Due to climate change, severe water scarcity in one

region and flood hazards in other regions may occur. Kumar and Sarma (2010) conducted a case study in the Sonitpur district of Assam, India, to investigate the extent of climatic variation and scope of utilizing rainwater harvesting (RWH) system for irrigating small agricultural field. Analysis of climatic parameters has revealed the increase in dry spell, increase in temperature range, and decrease in total annual rainfall. Rahman et al. (2010) have studied the effect of climate change on morphological behaviour of the major river systems in Bangladesh (Ganga, Yamuna and Padma), using MIKE 11 and MIKE 21. They predicted the rise in precipitation, temperature and sea level. Deka and Sarma (2011) have carried out a study to quantify the impact of the climate change on the precipitation characteristics of Brahmaputra Basin. The future series generated by temperature model has produced up to 1.8˚C rise in the maximum temperature averaged over month. Sarma et al. (2012) has carried out studies on the impact of climate change on water resources of Brahmaputra basin and adaptation strategies with regard to planning, design and management of water resources systems of the basin. The study indicated significant changes in temperature and rainfall pattern of this basin in future. Sarma et al. (2013) developed an optimization model for optimal Ecological Management Practice (EMP) to minimize impact of climate change and degradation of watershed due to urbanization. The study provides an idea for enhancing sustainability of EMPs by utilizing the carbon sequestration potential of vegetation, which can be adopted by industries to earn carbon credits, thus compensating their greenhouse gas (GHG) emissions.

General Circulation Models (GCMs) are widely used in climate change studies; but GCMs are relatively coarse in resolution. Therefore, some form of downscaling is required to produce scenario of finer spatial resolution than currently delivered by raw GCM output. Number of downscaling techniques and GCM combinations were experimented by various researchers. Wilby et al. (1998b) made a comparison of different Statistical downscaling methods using GCM output. Weather generator methods are better than methods like ANN for climate change porjections. Zorita and von Storch (1999) suggested a simple analog method for statistical downscaling and compared the same with other available complex methods. Misra et al. (2003) found that Regional Spectral Model (RSM) performs better than GCM models. Dibike and Coulibaly (2005) have made comparisons of hydrologic as well as downscaling models under climate change impacts. They predicted future variation in river flow and reservoir inflow using downscaled data from SDSM (regression-based approach) and LARS-WG (Weather Generator). Zhang (2005) developed a simple method for statistical downscaling of GCM output to predict soil erosion and crop production. Mehrotra and Sharma (2005, 2006)

attempted to develop nonparametric stochastic spatial downscaling methods for multisite daily rainfall projections. Ghose and Mujumdar (2006) developed a methodology to forecast future rainfall scenario over Orissa using GCM projections by statistical downscaling. The rainfall projection shows that there is a possibility of increase in hydrologic extremes in Orissa in future. Kang et al. (2007) have done statistical downscaling using six dynamic models and predicted the station scale precipitation in the Philippines and Thailand. They found that in some location, the statistical downscaling is not suitable because precipitation is governed mainly by local complicated terrain other than large-scale process. Serrat-Capdevila et al. (2007) have done statistical downscaling of GCM models to quantify the climate change impacts on hydrology. Anandhi et al. (2008) have presented a methodology for precipitation downscaling to river basins in India, with the use of support vector machine. Hessami et al. (2008) have done statistical downscaling of temperature and precipitation using automated statistical downscaling (ASD). Mujumdar and Ghosh (2008) have described a methodology of statistical downscaling based on Support Vector Machine (SVM) and Relevance Vector Machine (RVM). Tisseuil et al. (2010) compared four statistical downscaling models namely Generalized Linear Model (GLM), Generalized Additive Model (GAM), Artificial Neural Network (ANN) and Aggregated Boosted Trees (ABT). They used these models for predicting stream flow. Among these models, ABT was found to perform better. Kannan and Ghosh (2011) attempted to address the limitation of statistical downscaling in capturing correlation between multiple sites by representing the rainfall pattern in a river basin using rainfall state. Najafi et al. (2011a) have done statistical downscaling of precipitation employing independent component analysis (ICA) for predictor selection that determines spatially independent GCM variables. By doing uncertainty ana lysis of the downscaling methods, they found that model structures and input predictors are important factors in the downscaling process and the main sources of uncertainties. Deka and Sarma (2011) have carried out a study to quantify the impact of the climate change on the precipitation characteristics of Brahmaputra Basin. The future series generated by temperature model has produced up to 1.8 $^{\circ}$ C rise in the maximum temperature averaged over month. In the process of model development, it was experienced that the selection of predictors plays an important role in future series generation through downscaling of GCM. Vinnarasi (2012), in her project work, has focused on simulating the impact of climate change with the use of downscaling methods, on the precipitation characteristics and stream flow behaviour of Dhansiri River. Salvi et al. (2013) have projected daily rainfall of multiple sites in India using statistical downscaling at a resolution of 0.5^0, using classification and Regression Tree (CART) and nonparametric kernel regression technique.

Forsythe et al. (2014) have applied stochastic weather generator to assess climate change impacts in a semi-arid region. They stated that sophisticated downscaling methods were needed to evaluate the climate change impacts and the interannual variabilities for better understanding of the future conditions. Fatichi et al. (2015) have done a high-resolution fully distributed hydrological analysis and future climate projections. They found that there is an elevation dependence of climate change impacts in the catchment under consideration. Shamir et al. (2015) have analysed climate change projections of precipitation using eight dynamically downscaled GCM models. They stated that climate change projections increase the uncertainty and aggravates the complexity of water resources management task.

In the previous chapter it was observed that, regionalization of the north-eastern region of India by using fuzzy clustering approach results into the following homogeneous regions:

- **Cluster 1:** Region covering the stations Guwahati, Dhubri and Tezpur
- **Cluster 2**: Region covering the stations Passighat, North-Lakhimpur and Mohanbari
    - **Cluster 2a:** region on the west side of cluster 2
    - **Cluster 2b:** region on the east side of cluster 2
- **Cluster 3**: Region covering the stations Shillong, Silchar, Imphal and Cherrapunjee.

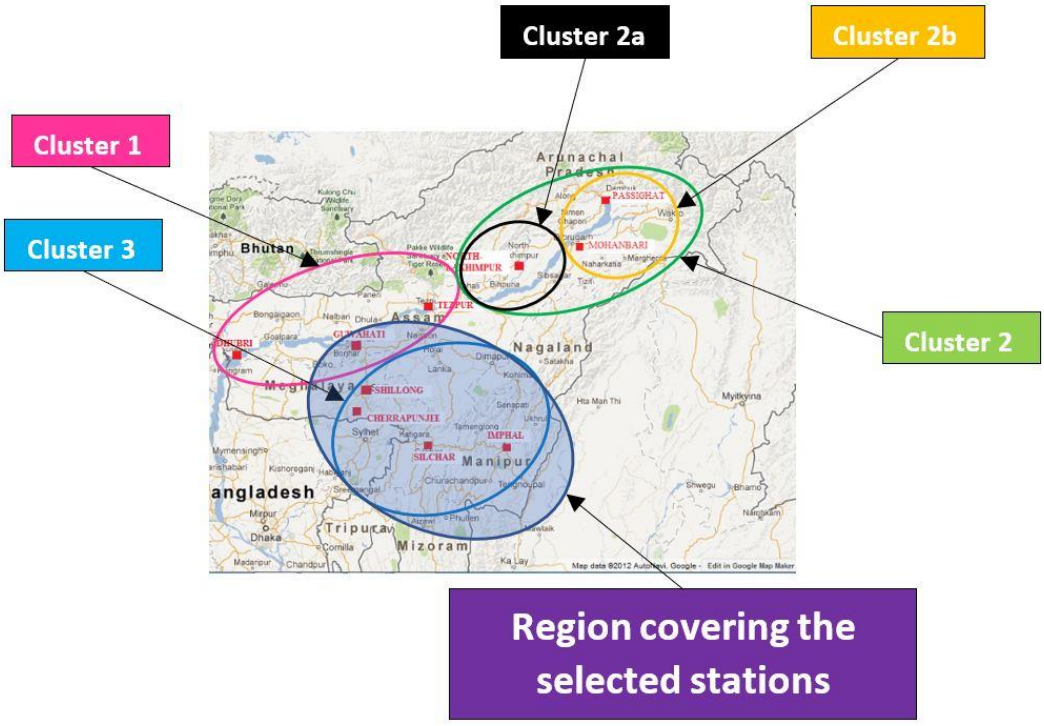


Figure 4. 1 Region covering the 5 selected stations for downscaling

For the purpose of further study, five stations have been selected to understand the variation of rainfall and temperature patterns within a cluster. Since cluster 3 covers a region which has larger variation in orography because of the presence of hilly terrain, hence the four stations present in cluster 3, viz. Cherrapunjee, Imphal, Shillong and Silchar have been selected. Guwahati, containing both hills and plains, is the nearest station to cluster 3 and located in the foothill region of this cluster. Therefore, this station has also been included in the analysis. The selected stations are shown in Fig. 4.1

Three different GCM models have been selected for downscaling. First step of downscaling process is to calibrate the model with the use of predictands (rainfall/temperature) and predictor variables (variables of GCM). In this chapter, downscaling models have been developed via calibration of models. Further, validation of the models has been done and the correlations have been checked. The results of calibration and validation are shown in this chapter and concluding remarks are presented.

4.2. METHODOLOGY

4.2.1. Downscaling using stepwise multiple linear regression

For HadCM3, downscaling of annual daily maximum temperature (T_{max}), annual daily minimum temperature (T_{min}) and rainfall at the five IMD stations has been done, with the use of stepwise Multiple Linear Regression (MLR).

In Multiple Linear Regressions (Yount, 1988) the operation procedure is divided into four basic steps, namely: Specification, Calibration, Validation and Forecast**.** In the specification stage model and predictors are selected. In calibration stage the relation between the output data and input data is obtained. The accuracy of the model is checked in the validation stage by comparing the observed and simulated data. In the forecast stage validated model is utilized to predict the future variations.

The MLR technique is developed based on the least square method. Following assumptions are taken into consideration while using the multiple linear regressions:

- The relation between Y (predictand) and X_1, X_2,…, X_n (predictors) is linear.
- The residuals have a constant variance σ and are normally distributed.
- There is no autocorrelation

- The X variables are fixed.

4.2.1.1. *Specification*

During the specification stage, the input data (both predictand and predictors) are pre-processed before using for calibrating the model. The general steps of data processing are explained below.

- **Interpolation**

The geographical location of a station doesn't always fall on the exact GCM grid point; rather within a grid, multiple stations points may be located. Hence, interpolation of the GCM data is needed to get the required dataset at individual stations. In the present study, GCM grid points are interpolated to the individual stations, with the use of two-dimensional linear interpolations by MATLAB programming.

- **Standardization**

Before calibration, the large-scale climate variables need to be processed. In the process of standardization, the mean (μ) is subtracted from the i^{th} value of predictor/predictand and then it is divided by the standard deviation (σ). The equation can be expressed as,

$$V_{std}(n) = \frac{v_i(n) - \mu(n)}{\sigma(n)} \quad (4.1)$$

Where, V_{std} is the standardized value for i^{th} variable of the n^{th} predictor

v_i is the i^{th} variable of the n^{th} predictor

μ is the mean of all the variables of n^{th} predictor, and

σ is the standard deviation of all the variables of n^{th} predictor.

- **Selection of predictors**

The Predictors are selected by stepwise regression method. This method consists of two key approaches, forward selection and backward elimination. Here, the coefficient of determination (R^2) is obtained between the observed data and a particular predictor. The coefficient of determination lies between 0 and 1. Larger the value, stronger is the correlation. The process continues till it reaches the best t-test, f-test and adjusted R^2 value.

- **Correlation between the predictor(s) and predictand**

It is the statistical relation between the arrays of two variables. It shows how they are correlated with each other. Pearson Correlation (PC) is mostly used to determine the coefficient of correlation. Its value lies between -1 to +1. Larger the absolute value of the correlation, stronger will be the relationship between the two variables. A positive value of PC means the two variables are directly correlated (or directly proportional); whereas negative PC value means they are inversely correlated (or inversely proportional).

- **Cross-correlation among the predictor variables**

While using stepwise Multiple Linear Regression (MLR), it is very important to ascertain whether the independent variables used are inter-dependent or not by doing cross correlation analysis among the independent variables used in the linear regression. Variables which are cross-correlated, if chosen for regression, may produce poorly performing models with high coefficient of determination value. Hence, if the cross-correlation value is more than a certain threshold value, those variables can be eliminated to remove such kind of multi-co-linearity.

4.2.1.2. Model calibration, validation and future projection

There are several ways for performing calibration. Some of these are:

- Multiple linear regressions without additive constant.

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in} \quad (4.2)$$

- Multiple linear regressions with additive constant.

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in} + r_i \quad (4.3)$$

- Multiple linear regression with a multiplying constant

$$y_i = (\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in}) m \quad (4.4)$$

Here, y_i = precipitation, β= coefficient, x_i=predictor, r_i=residual, m_i=multiplying constant.

A relationship (in the form of an equation) is developed between the predictand and the predictors by using the regression method during the calibration stage, which is then used for validating the model with a known set of data. If the coefficient of determination (R^2) between

the known (observed) data and the simulated data is more than an acceptable value (usually 0.8), then the developed model is said to be well-correlated with the original dataset and can be considered as suitable for future projection. Based on the model developed during calibration and validation, future scenarios are generated.

4.2.2. Downscaling using statistical downscaling model (SDSM) software

For models ESM2G and ESM2M, downscaling of maximum temperature, minimum temperature and rainfall at the five IMD stations has been done, with the use of Statistical Downscaling Model (SDSM) software. Future forecast of temperature and rainfall have been done by downscaling this climate model data till the end of 21st century.

SDSM or Statistical downscaling model (Wilby et al., (2002) is a combination of multiple linear regression and stochastic weather generator. It produces regression parameters by establishing statistical relationships between local-scale and large-scale variables. The stochastic weather generator uses these relations, along with GCM variables, to calibrate the model and simulates daily time series data. SDSM uses two types of optimization methods, ordinary least squares (OLS) and dual simplex (DS). Both methods produce similar results, although OLS works faster than DS (Huang et al., 2011).. Three sub-models are there in SDSM: annual (SDSM-A), monthly (SDSM-M) and seasonal (SDSM-S). The regression parameters generated by SDSM-A are similar for 12 months, whereas SDSM-M generates 12 different regression equations for each of the 12 months. Similarly, SDSM-S generates 4 regression equations for each of the 4 seasons. SDSM accounts for both conditional and unconditional processes. The unconditional process is used for independent variables (temperature, humidity, etc.) and the conditional process is used for variables such as precipitation and evaporation (Wilby, Dawson and Barrow, 2002). Further details on the SDSM model could be found in Wilby et al. (2002).

It is observed that rainfall data is not distributed normally, but in the case of temperature, the data is distributed normally. SDSM can transform the data to make it normal before using the data in regression equations. In the present study, the fourth root model transformation is used for rainfall to render it normal before using it in a regression equation.

4.2.2.1. *Stepwise explanation of the downscaling process using SDSM*

- **Screening of predictors**

Screening of predictors is the first and the most influencing step of the downscaling process, as it gives the insight to the decision on the appropriate predictor selection. Most of the time, we tend to use the predictors with the greatest correlations with the predictands; but we somehow ignore the inter-correlation among the predictors which affects the regression process. Hence, in this study, a stepwise screening of predictors is used, as proposed by Mahmood and Babel (2013). The procedure is explained below:

1. First, the correlation matrix between observed data (predictand) and all the predictors (GCM data) is made and arranged in descending order of correlation coefficient. First predictor with highest correlation is considered as the super-predictor (SP) and correlation coefficient is super correlation coefficient. To find out the next best predictor, the top 12 predictors with higher correlations are considered.
2. The absolute correlation coefficient between the predictor and predictand (R_1), the absolute correlation coefficient between individual predictors (R_2), absolute partial correlation (Pr), and P-value are obtained by regressing (11 in this case) the remaining highly correlated predictors individually in the presence of SP.
3. Then the predictors which have a P-value greater than 0.05 are removed to render the results statistically significant and the predictors which are highly correlated (0.5 in this study) are taken out in order to remove any multi-co-linearity. The correlation coefficient up to 0.7 between two predictors is considered acceptable.
4. Next the percentage reduction in an absolute partial correlation (PRP) is calculated for each predictor using the following equation.

$$PRP = \frac{Pr - R_1}{R_1} \tag{4.5}$$

5. The predictor which has a min PRP is selected as the second most suitable predictor and is termed as second super-predictor.
6. The third, fourth, and following predictors can be obtained by repeating steps 2 to 5.

An example of the process of screening of variable is shown in Table 4.1.

Table 4. 1 Example of the process of screening of predictors

Rank	ESM2G predictors	R1 (<0.5)		R2 (<0.7)		Pr		P-value	PRP(%)
		Correlation	Correlation (Abs)	Correlation	Correlation (Abs)	Correlation	Correlation (Abs)	<0.05	(Pr-R1)/R1
1	huss	0.211	0.211						
2	rlds	0.204	0.204	0.859	0.859	0.046	0.046	0.002	77.451
3	psl	-0.179	0.179	-0.354	0.354	-0.114	0.114	0	36.313
4	clt	0.174	0.174	0.75	0.75	0.025	0.025	0.112	85.632
5	tasmin	0.165	0.165	0.539	0.539	0.062	0.062	0	62.424
6	rhsmin	0.162	0.162	0.865	0.865	-0.041	0.041	0.008	74.691
7	rhs	0.157	0.157	0.877	0.877	-0.058	0.058	0.0001	63.057
8	prc	0.151	0.151	0.665	0.665	0.015	0.015	0.307	90.066
9	rlut	-0.145	0.145	-0.678	0.678	-0.004	0.004	0.545	97.241
10	rhsmax	0.141	0.141	0.847	0.847	-0.071	0.071	0	49.645
11	hfls	0.136	0.136	0.705	0.705	-0.018	0.018	0.244	86.765
12	pr	0.118	0.118	0.55	0.55	0.002	0.002	0.557	98.305

- **Calibration and Validation**

Based on the availability, observed daily dataset 1971–2000 is used for the calibration of max temp, min temp and rainfall. In this study, SDSM is developed using annual, monthly and seasonal sub-models. The developed annual, monthly and seasonal sub-models of SDSM are denoted by SDSM-A, SDSM-M and SDSM-S, respectively. While calibrating, the same predictors are used to calibrate all the sub-models. These models are developed individually for each of the predictands (max temp, min temp and rainfall) at each site. The unconditional sub-model is used for max temp and min temp without any transformation and conditional sub-model is used for rainfall with fourth root transformation. The models are developed for both ESM2G and ESM2M. Hence a total of 3 (sub-models) ×2 (ESMs) = 6 models are developed at each site for each of the predictands (max temp, min temp and rainfall).

Optimization of the best fit is done by Ordinary Least Square (OLS) method. With the developed models, max temp, min temp and rainfall are simulated for 1971–2000 for validation using the ESM-2G and ESM-2M predictors. A total of 20 ensembles are simulated using the annual, monthly and seasonal SDSMs and the mean of these ensembles is used for future projection. The outputs of SDSM are compared with observed data by calculating the coefficient of determination (R^2) and Pearson Correlation (PC).

- **Future projection**

Based on the 6 models developed (after validation), future scenarios are generated for RCP2.6 and RCP8.5 for all three parameters. A total of 6 (calibrated models) × 2 RCPs = 12 future series are generated for rainfall at each of the 5 IMD stations. The development of the future scenarios of the predictands for a particular site is shown in Fig. 4.2 in a hierarchical form.

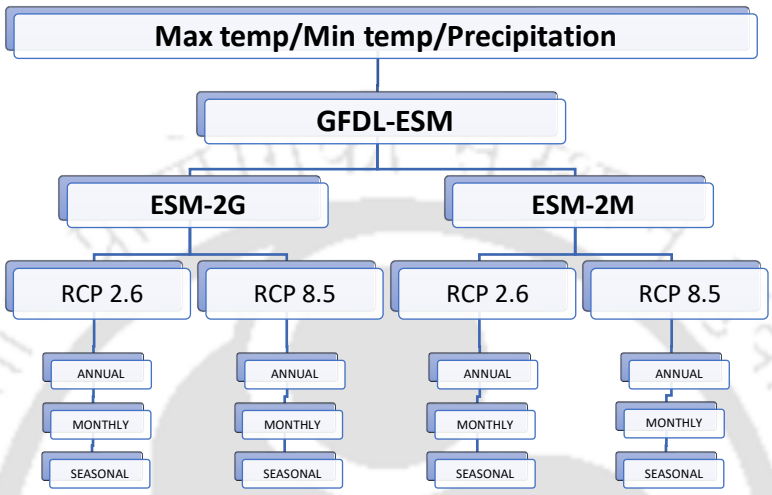


Figure 4. 2 Hierarchical diagram of the future scenario generation

4.2.3. Bias correction

Major drawback of using these climate models is the biases involved in them. Particularly in case of rainfall, the magnitude of these biases fluctuates to a large extent as precipitation is an unevenly distributed parameter in temporal scale. Hence to reduce these biases, corrections need to be done before utilizing the predicted temperature and rainfall for impact assessment studies. There are several bias correction methods that have been suggested for correcting precipitation simulations (Watanabe et al., 2012; Mahmood and Babel, 2013). Here, following equations have been used for bias correction (Mahmood and Babel, 2013).

$$T_{bc} = T_{ds} - (\overline{T_{cont}} - \overline{T_{obs}}) \tag{4.6}$$

$$P_{bc} = P_{ds} \times (\frac{\overline{P_{obs}}}{\overline{P_{cont}}}) \tag{4.7}$$

where, T_{bc} and P_{bc} are bias corrected temperature and precipitation series, T_{ds} and P_{ds} are downscaled temperature and precipitation series from MLR/SDSM, $\overline{T_{obs}}$ and $\overline{P_{obs}}$ are mean of

the observed temperature and precipitation series, and $\overline{T_{cont}}$ and $\overline{P_{cont}}$ are mean of downscaled temperature and precipitation data series for the calibration period, respectively.

4.2.4. Trend analysis

To understand the trend of the parameters (T_{max}, T_{min}, and rainfall) over the observed and projected period, trend analysis needs to be done. Various statistical test methods are available that can detect the trend in a hydrological and hydrometeorological data series. These methods are categorized as parametric and non-parametric tests. Parametric tests are used for independent dataset that has a normal distribution. However, for hydrological dataset this is hardly true. Hydrological parameters generally do not show normal distribution pattern, especially for parameters like rainfall which may have a number of outliers. Non-parametric tests represent monotonic linear dependence (Rossi et al., 1992; Davis, 2002) and are better at tolerating the influence of the presence of outliers (Lanzante, 1996). Hence, these types of tests are found to be suitable for trend analysis of hydrological time series.

Mann-Kendall (MK) trend test (Mann, 1945; Kendall, 1975) is the most widely used non-parametric test for trend analysis of hydrological time series (Basistha et al., 2009; Longobardi and Villani, 2009; Oguntunde et al., 2012; Wang et al., 2012; Yang et al., 2012; Yanming et al., 2012; Anghileri et al., 2014). MK trend test is not influenced by the actual distribution of the data and hence is less sensitive to outliers. The test either accepts or rejects the null hypothesis. H = 1 implies rejection of the null hypothesis, indicating presence of a trend (increasing or decreasing). H = 0 implies failure to reject the null hypothesis, indicating absence of a trend. The statistics S of MK test, variance of S i.e. V(S) and the standardized test statistics Z can be calculated as follows:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_j - x_i) \tag{4.8}$$

where,

$$sgn(x_j - x_i) = \begin{cases} 1, & x_i < x_j \\ 0, & x_i = x_j \\ -1, & x_i > x_j \end{cases} \tag{4.9}$$

$$V(S) = \frac{1}{18} \left[n\,(n\,-\,1)\,(2n\,+\,5)\,-\,\sum_{p=1}^{q} t_p\,(t_p\,-\,1)\,(2t_p\,+\,5) \right] \tag{4.10}$$

$$Z = \begin{cases} \frac{S-1}{\sqrt{VAR(S)}}, & if\ S > 0 \\ 0, & if\ S = 0 \\ \frac{S+1}{\sqrt{VAR(S)}}, & if\ S < 0 \end{cases} \quad (4.11)$$

Here x_i and x_j are data points (in chronological order) at time i and j respectively, of the time series x_1, x_2,..., x_n; n is the length of the time series, q is the number of tied groups (set of data with same value), t_p is the number of data points in the p^{th} group.

A positive value of S indicates an increasing trend, while negative value indicates a decreasing trend. Similarly, a positive Z value indicates an increasing trend, whereas negative Z value implies a decreasing trend. If $|Z| > Z_{1-\alpha/2}$, null hypothesis is rejected indicating significant trend in the time series. The critical value of $Z_{1-\alpha/2}$ for a significance level (p) of 0.05 is 1.96. It means that, Z value greater than +1.96 indicates significant increasing trend and lower than −1.96 implies significant decreasing trend.

However, the original MK test has some limitations. It requires the data series to be serially independent and no autocorrelation should be present. In presence of positive or negative autocorrelation, MK test leads to erroneous results (Hamed and Ramachandra Rao, 1998; Yue et al., 2002; Chatterjee et al., 2016). While positive autocorrelation detects significant trend even if there may be actually no trend, negative autocorrelation may neglect the significance and hence may fail to detect any trend (Cox and Stuart, 1955). To address this issue, pre-whitening of the data series needs to be done to remove the serial correlation, before applying the MK test (Hamed and Ramachandra Rao, 1998; Von Storch and Navarra, 2010). Hamed and Ramachandra Rao (1998) proposed a modified version of MK test, which is applicable for autocorrelated data series. This modified method incorporates a revised equation for V(S), thus reducing the effect of autocorrelation on variance of S statistics, and can be applied for trend analysis of hydrological data series. The revised V(S) value can be calculated as follows:

$$V^*(S) = V(S)\frac{n}{n_S^*} \quad (4.12)$$

where, $\frac{n}{n_S^*}$ represents correction due to autocorrelation in the time series. It can be calculated empirically as,

$$\frac{n}{n_S^*} = 1 + \frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n-1}(n-i)(n-i-1)(n-i-2)\rho_S(i) \quad (4.13)$$

Here, n is the number of data points (actual sample size), n* is the effective sample size and $\rho_S(i)$ is the autocorrelation function of the ranks of data points that can be calculated by taking inverse of the following equation:

$$\rho(i) = 2\sin\left(\frac{\pi}{6}\rho_S(i)\right) \tag{4.14}$$

4.3. DATA COLLECTION

Two types of data have been used in this study, namely

- Observed rainfall data of 5 selected station, collected from IMD
- Global Climate model (GCM) data collected from IPCC AR4 and AR5
    - AR4: HadCM3, scenario A2 (considered as equivalent to RCP8.5)
    - AR5: GFDL-ESM2G and GFDL-ESM2G (RCP2.6 and RCP8.5)

4.3.1. Observed rainfall data

Observed daily rainfall data have been collected from Indian Meteorological Department (IMD) under an MOU between IIT Guwahati and IMD. The time period of data collection is from 01-01-1969 to 31-01-2012. Details are given in section 3.3.1 of chapter 3.

The 5 IMD stations selected, along with their latitude-longitude-elevation, can be found in Table 3.1 in the previous chapter. The locations of the stations are shown in the Fig. 3.2. The study area is situated near the Meghalaya plateau and due to the presence of hilly terrain this region has larger variation in orography. Therefore, it is assumed that chosen stations will be suitable to study the effect of orography in downscaling, in the next chapter.

4.3.2. Global Climate Model (GCM) data

The GCM data are collected from Fourth Assessment Report (AR4) and Fifth Assessment Report (AR5) of Intergovernmental Panel on Climate Change (IPCC). Three GCM models are used in the study: HadCM3, ESM2G and ESM2M. The models are explained in detail in section 3.3.2 of chapter 3.

Here it is required to be mentioned that, since HadCM3 and the two ESM models were already used for clustering, hence same models are being used for downscaling. The reason of using

similar models for homogeneous clustering and for developing downscaling model is to understand the consistency of the GCM variables in both cases and to check whether the variables are able to capture the orographic effect within the cluster or not.

4.4. RESULTS AND DISCUSSION

4.4.1. Analysis of historical trend of rainfall and temperature data

For analysing the existing trends of maximum temperature (t_{max}) and minimum temperature (t_{min}) and rainfall, the IMD dataset are used which ranges from 1971 to 2000. These datasets are examined on both annual and monthly basis to determine the average change in rainfall and temperature in a year and in a month, respectively, over the 30 years of historical period of record.

a) Cherrapunjee

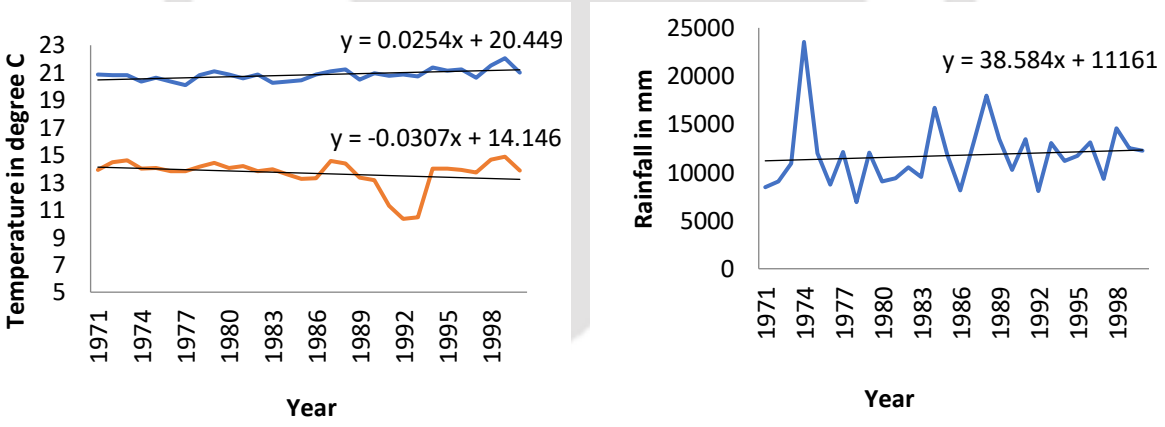


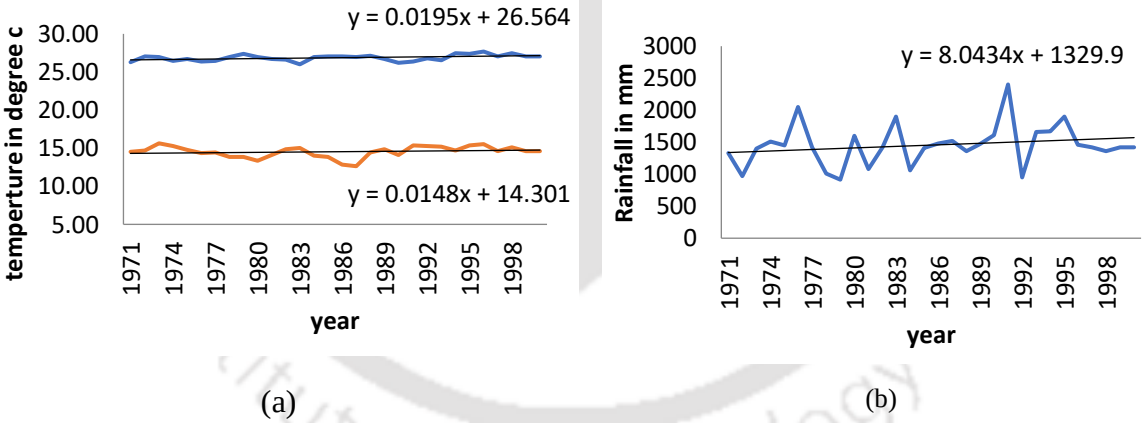
Figure 4. 3 (a) Annual mean t_{max} and t_{min}, and (b) total annual rainfall of Cherrapunjee

The annual mean of t_{max} shows an increasing trend from 1971 to 2000 (Fig. 4.3(a)), with an approximate rate of 0.03 °C per year. In contrast, annual mean of t_{min} shows a decreasing trend with approximate rate of 0.03 °C per year. The 30-year annual mean t_{max} is 20.8 °C and t_{min} is 13.7 °C, which give 30-year annual mean temperature of Cherrapunjee as 17.3 °C. The monthly trend analysis of mean t_{max} shows increasing trend from 1971 to 2000 for all the months except March month, with higher increasing rate in autumn-winter (Oct-Nov-Dec) and lower increasing rate in spring-summer (May-Jun-Jul-Aug). Monthly mean of t_{min} shows decreasing trend for all 12 months with higher rate in Oct-Nov and Jan-Feb, and lower rate in other months.

It is observed that the mean annual t_{min} is significantly low in early 90's (1991 to 1993) which is also found in case of monthly analysis of mean t_{min}. The highest t_{max} is 22.1 °C (year 1999) and lowest t_{min} is 10.3°C (year 1992).

The total annual rainfall analysis shows an increasing trend over the 30 years period (Fig. 4.3(b)), with a rate of increase 38.58 mm per year. 30-year average of annual rainfall is 11759 mm. However, the highest annual rainfall occurred in the year 1974, which is 23539.7 mm. The monthly trend analysis of total monthly rainfall shows an increasing trend in the initial part of the year (from Jan), the rate being highest during May-June, then from July starts to show decreasing trend till Dec (April and Aug being exception). Total annual dry day analysis does not show much variation for the 30 years period, with an average of 187 days per year. Total monthly dry day analysis shows increasing trend for April month and decreasing rate in Jun-Jul-Aug, for other months not much variation.

b) Imphal



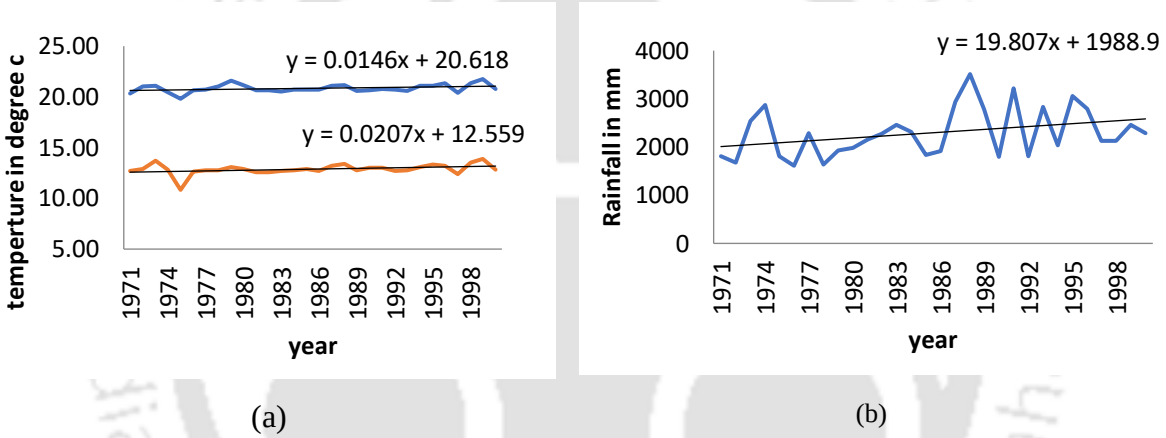
(a) (b)

Figure 4. 4 (a) Annual mean t_{max} and t_{min}, and (b) total annual rainfall of Imphal

Both annual mean of t_{max} and t_{min} show increasing trends from 1971 to 2000 (Fig. 4.4(a)), with approximate rates of 0.02 °C per year and 0.01 °C per year, respectively. The 30-year annual mean t_{max} is 26.9 °C and t_{min} is 14.5 °C, which give 30-year annual mean temperature of Imphal as 20.7 °C. The monthly trend analysis of mean t_{max} shows increasing trend from 1971 to 2000 for all the months except March month, with higher increasing rate in autumn-winter (Sept-Oct-Nov). Monthly mean of t_{min} shows increasing trend except for Nov month, with higher rate in Feb-Mar. The highest t_{max} is 27.7 °C (year 1996) and lowest t_{min} is 12.6 °C (year 1987).

The total annual rainfall analysis shows an increasing trend over the 30 years period (Fig. 4.4(b)), with a rate of increase 8.04 mm per year. 30-year average of annual rainfall is 1454.6 mm. However, the highest annual rainfall occurred in the year 1991, which is 2402.1 mm. The monthly trend analysis of total monthly rainfall shows an increasing trend in most of the months, except Mar and Nov, the rate being highest during May to Sept. Total annual dry day analysis shows a slight decreasing trend for the 30 years period, with an average of 208 days per year. Total monthly dry day analysis shows higher rate of decrease in monsoon (May-Jun-Jul-Aug-Sept).

c) Shillong



(a) (b)

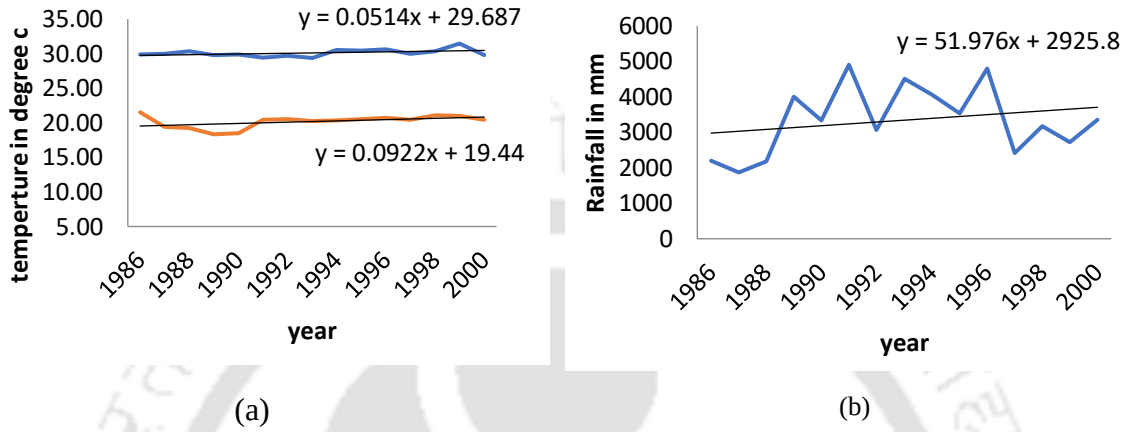
Figure 4. 5 (a) Annual mean t_{max} and t_{min}, and (b) total annual rainfall of Shillong

Both annual mean of t_{max} and t_{min} show increasing trends from 1971 to 2000 (Fig. 4.5(a)), with approximate rates of 0.01 °C per year and 0.02 °C per year, respectively. The 30-year annual mean t_{max} is 20.8 °C and t_{min} is 12.9 °C, which give 30-year annual mean temperature of Shillong as 16.9 °C. The monthly trend analysis of mean t_{max} shows increasing trend from 1971 to 2000 for all the months except March month, with higher increasing rate in June and lower increasing rate in Jan. Monthly mean of t_{min} shows decreasing trend for Dec-Jan and increasing trend for other months with higher rate during monsoon (May-Jun-Jul-Aug-Sept). The highest t_{max} is 21.8 °C (year 1999) and lowest t_{min} is 10.8 °C (year 1975).

The total annual rainfall analysis shows an increasing trend over the 30 years period (Fig. 4.5(b)), with a rate of increase 19.8 mm per year. 30-year average of annual rainfall is 2296 mm. However, the highest annual rainfall occurred in the year 1988, which is 3512.5 mm. The monthly trend analysis of total monthly rainfall shows higher increasing trends in the pre-monsoon season (Mar-Apr-May), and post-monsoon season (Aug-Sept-Oct). Total annual dry

day analysis shows a slight decreasing trend over the 30 years period, with an average of 184 days per year. Total monthly dry day analysis shows higher decreasing rate during Aug-Sept-Oct.

d) Silchar



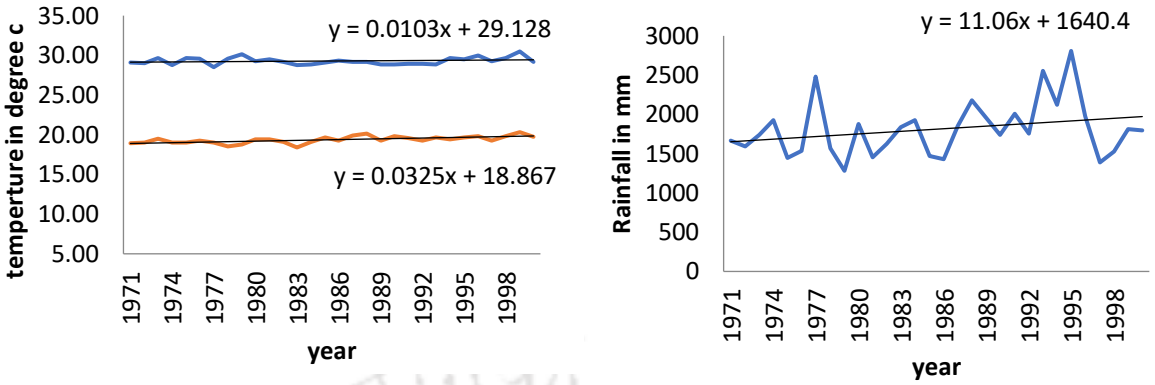
(a) (b)

Figure 4. 6 (a) Annual mean t_{max} and t_{min}, and (b) total annual rainfall of Silchar

Both annual mean of t_{max} and t_{min} show increasing trends from 1986 to 2000 (Fig. 4.6(a)), with approximate rates of 0.05 °C per year and 0.09 °C per year, respectively. The 15-year annual mean t_{max} is 29.5 °C and t_{min} is 20.2 °C, which give 15-year annual mean temperature of Silchar as 24.8 °C. The monthly trend analysis of mean t_{max} shows increasing trend from 1986 to 2000 for almost all the months, with higher increasing rate in April and Oct-Nov-Dec. Monthly mean of t_{min} shows increasing trend for all months with higher rate during pre-monsoon (Mar-Apr) and post-monsoon season (Oct-Nov). The highest t_{max} is 31.5 °C (year 1999) and lowest t_{min} is 18.3 °C (year 1989).

The total annual rainfall analysis shows an increasing trend over the 15 years period (Fig. 4.6(b)), with a rate of increase 51.97 mm per year. 15-year average of annual rainfall is 3302.1 mm. However, the highest annual rainfall occurred in the year 1991, which is 4903.7 mm. The monthly trend analysis of total monthly rainfall shows higher increasing trends in Mar and in monsoon season (May to Aug) and decreasing trend in Apr, Oct and Nov. Total annual dry day analysis shows a slight decreasing trend over the 15 years period, with an average of 185 days per year. Total monthly dry day analysis shows higher decreasing rate during Mar and in monsoon (May to Aug) and increasing rate during Oct-Nov.

e) Guwahati



(a) (b)

Figure 4. 7 (a) Annual mean t_{max} and t_{min}, and (b) total annual rainfall of Guwahati

Both annual mean of t_{max} and t_{min} show increasing trends from 1971 to 2000 (Fig. 4.7(a)), with approximate rates of 0.01 °C per year and 0.03 °C per year, respectively. The 30-year annual mean t_{max} is 29.3 °C and t_{min} is 19.4 °C, which give 30-year annual mean temperature of Guwahati as 24.3 °C. The monthly trend analysis of mean t_{max} shows increasing trend from 1971 to 2000 for all the months except March month, with higher increasing rate in Nov-Dec and May month and lower increasing rate in Feb month. Monthly mean of t_{min} shows increasing trend for all 12 months with higher rate in Nov-Dec and May, and lower rate in other months. The highest t_{max} is 30.5 °C (year 1999) and lowest t_{min} is 18.4 °C (year 1983).

The total annual rainfall analysis shows an increasing trend over the 30 years period (Fig. 4.7(b)), with a rate of increase 11.06 mm per year. 30-year average of annual rainfall is 1811.8 mm. However, the highest annual rainfall occurred in the year 1995, which is 2808.1 mm. The monthly trend analysis of total monthly rainfall shows an increasing trend in the pre-monsoon season (May) and later monsoon seasons (Aug-Sept-Oct). Total annual dry day analysis does not show much variation for the 30 years period, with an average of 223 days per year. Total monthly dry day analysis shows a decreasing trend in monsoon (May-Jun-Jul-Aug-Sept), for other months not much variation.

4.4.1.1. Salient points on meteorological characteristics of the stations

Among the stations, rainfall of Cherrapunjee is the highest (11759 mm), rest of the stations have almost similar amount of annual rainfall, Imphal have the lowest value (1454.6 mm). Silchar is the hottest with high annual mean t_{max} and t_{min} values. Shillong and Cherrapunjee are the coolest stations with low t_{max} and t_{min} values. Year 1999 is the hottest year within the 30-

year historical period for most of the stations. The annual trends of maximum temperature, minimum temperature and rainfall for all the stations are summarized in Fig. 4.8, 4.9 and 4.10, respectively. The meteorological parameter values are given in Table 4.2.

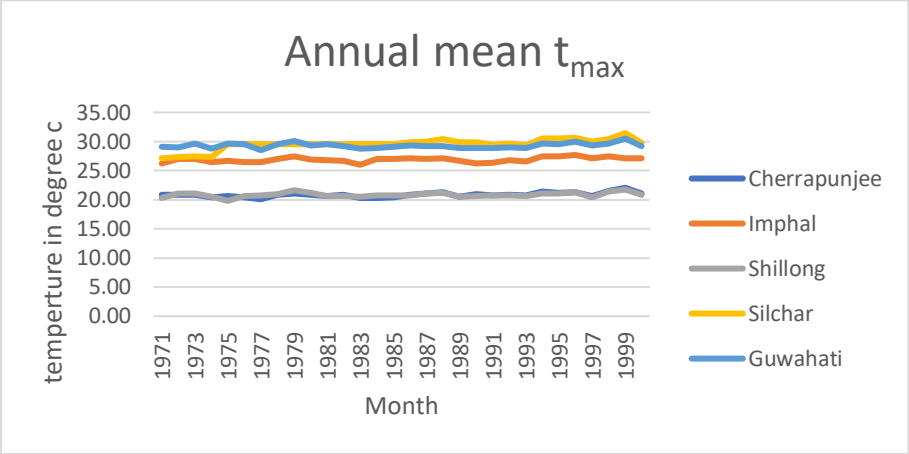


Figure 4. 8 Historical trend of annual mean maximum temperature (t_{max})

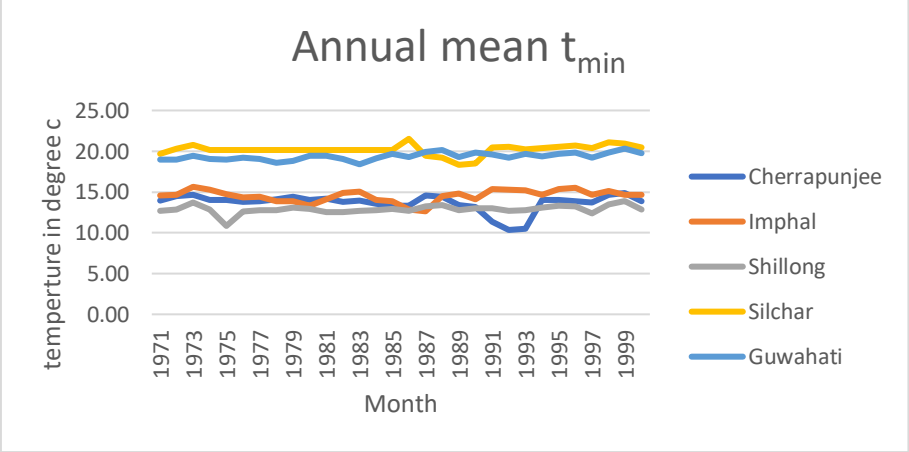


Figure 4. 9 Historical trend of annual mean minimum temperature (t_{min})

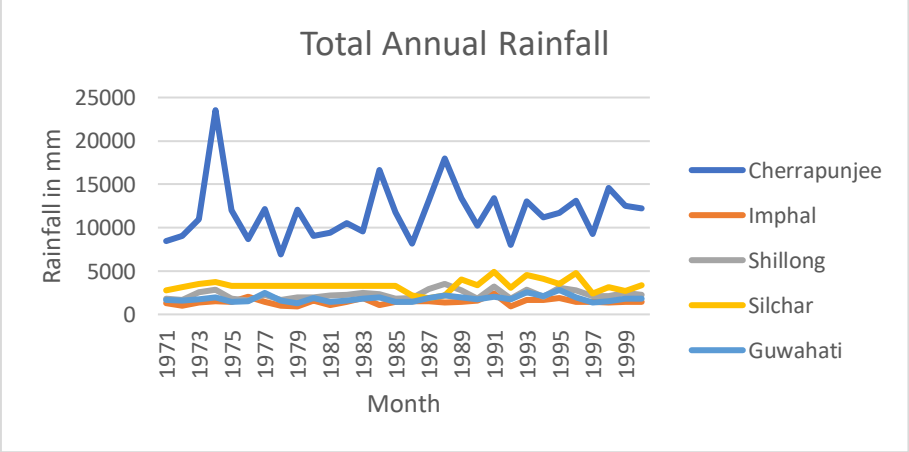


Figure 4. 10 Historical trend of total annual rainfall

Table 4. 2 Summary of the meteorological parameters of the stations

Meteorological Parameters		Cherrapunjee		Imphal		Shillong		Silchar		Guwahati	
			year		year		year		year		year
T_{max}(in degree C)	average	20.8		26.9		20.8		29.5		29.3	
	max	22.1	1999	27.7	1996	21.8	1999	31.5	1999	30.5	1999
T_{min} (in degree C)	average	13.7		14.5		12.9		20.2		19.4	
	min	10.3	1992	12.6	1987	10.8	1975	18.3	1989	18.4	1983
Annual mean temp (in degree C)		17.3		20.7		16.9		24.8		24.3	
Rainfall in mm	average	11759		1454.6		2296		3302.1		1811.8	
	max	23539.7	1974	2402.1	1991	3512.5	1988	4903.7	1991	2808.1	1995

4.4.1.2. Mann-Kendall trend test

The trend analysis, explained above, gives an insight to the increasing and decreasing trend of a parameter. However, to understand the significance of the trend, Mann-Kendall (MK) trend test is performed. The results of the MK test are given in Table 4.3. Cherrapunjee, Imphal, Shillong and Silchar has shown significant increasing trend for maximum temperature during the observed period of record, whereas Guwahati has shown non-significant increasing trend. For minimum temperature, Cherrapunjee has shown non-significant decreasing trend; whereas Imphal, Silchar have shown non-significant increasing trend and Shillong, Guwahati have shown significant increasing trend. For rainfall, all the stations have shown increasing trend, but only the trend of Shillong has indicated significant increase during the historical period of record.

Table 4. 3 Trend analysis of annual average temperature (daily maximum and minimum) and rainfall for past (1971–2000) climate using Mann-Kendall trend test

Meteorological Parameters		Cherrapunjee	Imphal	Shillong	Silchar	Guwahati
T_{max}	trend	++	++	++	++	+
T_{min}	trend	-	+	++	+	++
Rainfall	trend	+	+	++	+	+

+ represents non- significant increasing trend (P > 0.05)
- represents non- significant decreasing trend (P > 0.05)
++ represents significant increasing trend (P < 0.05; MK Z-value > 1.96).
-- represents significant decreasing trend (P < 0.05; MK Z-value < -1.96).

4.4.2. Development of the model for temperature and rainfall data

For developing the downscaling model, three GCM models are used: HadCM3, ESM2G and ESM2M. Historical datasets of the three GCM models are different from each other. HadCM3 is developed under IPCC's Assessment Report (AR4) for which data has been downloaded from 2000 to 2100. ESM2G and ESM2M have simulated datasets for historical data (up to 2005) and future scenario datasets for different RCPs (2006-2100) under CMIP5 protocol. IMD daily data are available from 1971-2000 and 2006-2011. There are no recorded data for the period 2001-2005. Since the datasets are different, hence the base periods chosen for the three models are also different.

4.4.2.1. Development of the model HadCM3 (A2 scenario)

One of the most critical parts of a hydrological study is the scarcity of observed data. In most cases, 30 years of data is considered to be enough for calibrating a meteorological trend. Some researchers have also used short period of data (10 years). But there is no proper recommendations or guidelines that suggests what is the best data-length to get a satisfactory result for any meteorological study. In north-eastern states of India, data scarcity is a very common case both in temporal and spatial scale. The orographic variations in the region is very high. Compared to that number of meteorological stations available is relatively small. Moreover, recorded datasets available from some of the stations are of short length, which might be because of less accessibility to the remote locations of those stations. When the observed data collected is less, it is difficult to choose which set of data should be used for calibration and which set should be used for validation. The future dataset obtained from downscaling will be acceptable, if the calibrated model is precisely validated before it is applied for prediction. Previous studies show that data scarcity degrades the quality of downscaling results by overestimating or underestimating the downscaled parameters. Wagner et al. (2015) stated that limited baseline period had led to underestimation of temperatures in Western Ghats, India. Literatures can be found where researchers have faced this issue and hence tried new approaches to make the downscaling results better (Wagner et al., 2015; Huan-Yu LIN et al., 2018; Zuo et al., 2019).

In case of HadCM3, six years daily data of IMD (2006-2011) are available which is a very short dataset. Therefore, we need a new method for finding out the best fit dataset for calibrating the model. In this study, different combinations of calibration-validation sets are compared to find out their effectiveness and related challenges. Based on the results, an attempt

is made to find out the best combination of data for calibration as well as validation that will precisely predict the future pattern of rainfall in data scares areas. Available dataset is converted to monthly data and used as base period for calibration and validation. The approach is explained below and the calibration-validation results are presented.

- **Screening of predictors**

Screening of predictors has been done as per the method explained for MLR in the methodology section (section 4.2.1). The datasets have been interpolated and standardized before use. There are 25 predictors available for HadCM3 model. It is not very wise to use all the available predictors in the MLR for calibration, which will make the process tedious. Hence an initial screening has been done by observing the correlation of these predictors with the observed data. Those predictors having correlations well enough to be used for calibration are considered. For this purpose, Pearson correlations (PC) between the predictand variables and predictors are calculated for each of the 5 stations. The lists of selected predictors for maximum temperature, minimum temperature and rainfall are given in Annexure C. Furthermore, cross-correlation analysis has also been performed to remove the predictors having correlation value more than 0.85 (threshold). Cross-correlation tables are provided in Annexure D. The correlations were calculated for each station separately, which is explained in the next paragraph.

- **Calibration and validation**

The base period considered for calibration and validation is very short, since only six years data are available (2006-2011). In the present study, different combinations are tried to find out the best combination of data for calibration. Six combinations are considered here as tabulated below in Table 4.4.

Table 4. 4 Different combinations of base period for calibration

	combination 1		**combination 2**		**combination 3**		**combination 4**		**combination 5**		**combination 6**	
	calib	**valid**	**calib**	**valid**	**calib**	**valid**	**calib**	**valid**	**calib**	**valid**	**calib**	**valid**
Years	2006-2007-2008	2009-2010-2011	2009-2010-2011	2006-2007-2008	2006-2008-2010	2007-2009-2011	2007-2009-2011	2006-2008-2010	2006-2007-2008-2009	2010-2011	2006-2007-2008-2009-2010	2011

Note: calib: calibration period; valid: validation period

Stepwise MLR has been performed for all the six combinations separately to find out the final set of predictors, which are different for different combinations. Selected predictors and their R^2 values (Co-efficient of determination) are provided in Annexure E with figures. From the R^2 values, it has been observed that combination 6 (5 years for calibration and 1 year for validation) is giving the best results for all stations. However, it may be presumed by some researcher that this has happened due to the consideration of short period of validation data. Hence, to find out the correlation of the calibration-validation datasets, statistical parameters are evaluated for all the six combinations, at all the stations. The parameters considered are, Pearson correlation (PC), Coefficient of Determination (R^2), Root Mean Square Error (RMSE), Relative error in mean (RE_μ), Relative error in standard deviation (RE_σ) and Percentage change in the maximum value of the series. These Statistical parameters results are tabulated in Annexure F.

From the statistical parameter analysis, it is observed that in most of the cases the similarities between the two datasets (calibration-validation) are lowest for Combination 6, yet this combination is giving better results than the other combinations. This can be attributed to the reason that the calibrated model becomes more reliable due to consideration of more data points from the observed series. However, it should be noted that, the correlation coefficients (Person correlation and Coefficient of Determination, R^2) should not be very low or exceptionally low. Furthermore, the other error parameters namely, Root Mean Square Error (RMSE), Relative error in mean (RE_μ), Relative error in standard deviation (RE_σ) and Percentage change in the maximum value of the series, should be within considerable range. Further observation of the statistical results shows that, the best combination is the one with lower errors in RMSE, RE_μ, RE_σ and percentage change in maximum. Hence, we may say that lower the errors, better is the combination and better is the model calibrated.

Hence, combination 6 is used for simulation of future series. Coefficient of determination (R^2) values for combination 6 are shown in Fig. 4.11.

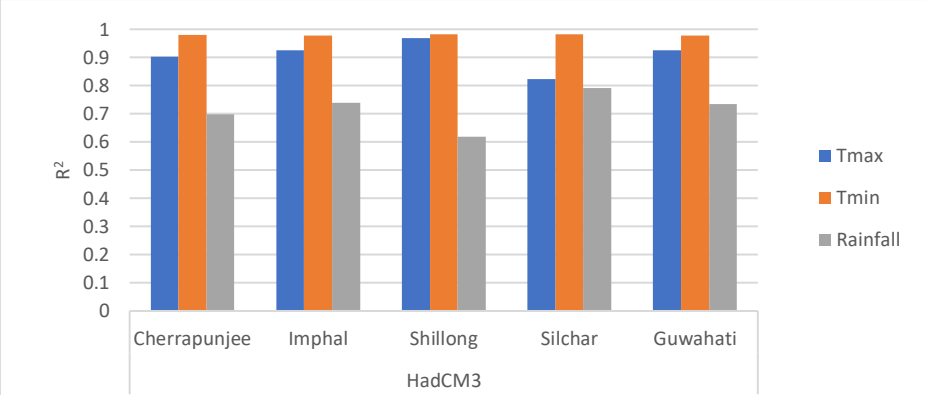


Figure 4. 11 Coefficient of determination (R^2) between observed and simulated daily maximum temperature (T_{max}) and daily minimum temperature (T_{min}) and rainfall for HadCM3

4.4.2.2. Development of the models ESM2G and ESM2M

- **Screening of predictors**

In this study, stepwise screening of predictors (Mahmood and Babel, 2013) has been used as explained in the methodology section (section 4.2.2). Based on the availability, observed daily dataset 1971–2000 is used for the calibration of maximum temperature, minimum temperature and rainfall for all stations (1986-2000 for Silchar station). 4 predictors are selected out of the 23 predictors for each predictand, both for ESM2G and ESM2M. The selected predictors for maximum temperature, minimum temperature and rainfall are shown in Tables 4.5, 4.6 and 4.7, respectively. The predictors are arranged in descending order of their suitability. (SP1, SP2, SP3 & SP4).

Table 4. 5 Selected predictors for maximum temperature

Cherrapunjee		**Imphal**		**Shillong**		**Silchar**		**Guwahati**	
ESM-2G	**ESM-2M**	**ESM-2G**	**ESM-2M**	**ESM-2G**	**ESM-2M**	**ESM-2G**	**ESM-2M**	**ESM-2G**	**ESM-2M**
rlds	tasmin	tasmin	tasmin	tasmin	tasmin	rlds	tasmin	tasmin	tasmin
tasmax	rhs	hfls	hfls	hfls	hfls	hfls	rhs	hfls	hfls
rhsmin	tasmax	prc	prc	clt	clt	tasmax	clt	prc	clt
prc	prc	rhsmin	tasmax	pr	pr	prc	hfls	rhsmin	pr

Table 4. 6 Selected predictors for minimum temperature

Minimum temperature									
Cherrapunjee		**Imphal**		**Shillong**		**Silchar**		**Guwahati**	
ESM-2G	**ESM-2M**	**ESM-2G**	**ESM-2M**	**ESM-2G**	**ESM-2M**	**ESM-2G**	**ESM-2M**	**ESM-2G**	**ESM-2M**
rlds	rlds	rlds	rlds	rlds	tasmin	rlds	rlds	rlds	tasmin
hfls	hfls	hfls	tasmax	tasmax	rhsmin	hfls	clt	tasmax	rhs

rhs	tasmax	rhs	rhsmin	rhsmin	hfls	rlus	hfls	rhs	hfls
tas	rhsmin	rlus	hfls	hfls	tasmax	rhsmin	rlus	hfls	clt

Table 4. 7 Selected predictors for rainfall

Rainfall									
Cherrapunjee		**Imphal**		**Shillong**		**Silchar**		**Guwahati**	
ESM-2G	**ESM-2M**	**ESM-2G**	**ESM-2M**	**ESM-2G**	**ESM-2M**	**ESM-2G**	**ESM-2M**	**ESM-2G**	**ESM-2M**
huss	huss	clt	rhsmin	huss	rlds	tasmin	tasmin	rlds	tasmin
rhsmax	rhsmax	huss	sfcwindmax	rlds	sfcwindmax	huss	hfls	rlus	hfls
clt	prc	hfls	rlds	sfcwindmax	hfls	prc	vas	tas	prc
hfls	rlut	rlds	tasmax	prc	rlut	sfcwind	rlut	rlut	rhsmin

It was observed that, although the same GCM variables were used for homogeneous clustering as well for downscaling, the correlation coefficients do not show any consistency within a cluster. The selected predictors are different for different stations, which eventually gives different downscaling models for different stations within a cluster, although the areal distances between those stations were not very large. This indicates that orography plays a very important role in variations of hydrometeorological parameters at a local scale.

- **Calibration and Validation**

Annual (SDSM-A), monthly (SDSM-M) and seasonal (SDSM-S) sub-models of SDSM are developed individually for each of the predictands) at each of the 5 IMD sites. Calibration is done for the period 1971-2000. Each of the calibrated models is then validated with simulated data.

Major drawback of using these climate models is the biases involved in them. Particularly in case of rainfall, the magnitude of these biases fluctuates to a large extent as rainfall is an unevenly distributed parameter in temporal scale. Hence to reduce these biases, corrections need to be done before utilizing the predicted temperature and rainfall for impact assessment studies. There are several bias correction methods that have been suggested for correcting rainfall simulations (Watanabe et al., 2012; Mahmood and Babel, 2013). The models were validated with observed data by calculating the coefficient of determination (R^2). R^2 values are shown in Fig 4.12.

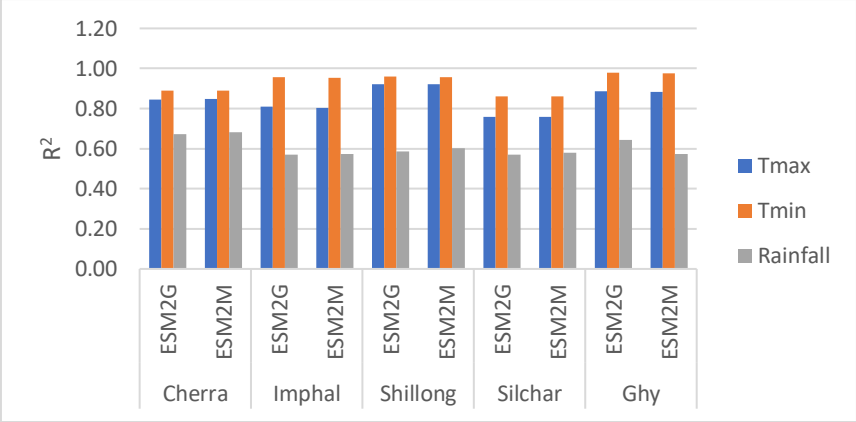


Figure 4. 12 Coefficient of determination (R^2) between observed and simulated daily maximum temperature (T_{max}) and daily minimum temperature (T_{min}) and rainfall for ESM2G and ESM2M

A value greater than 0.8 for R^2 is considered as a good correlation between observed and simulated data. From both Fig. 4.11 and Fig. 4.12, it was realized that the models developed for daily maximum and minimum temperatures have good correlation (R^2 value) between the observed and simulated data, indicating well developed downscaling model. This implies that, local scale orographic arrangement has less effect on these independent hydrometeorological parameters, while developing downscaling model for future projection. However, in case of rainfall, having a good correlation (more than 0.8 is very difficult, as daily precipitation is not normally distributed and is highly unpredictable. R^2 value more than 0.6 gives acceptable simulation results in case of rainfall, although there are still a lot of uncertainties associated with the simulated results. Hence, there is need to develop better correlated models with the use of local orographic variances which might be affecting the outputs of downscaling models.

4.5. SUMMARY

Following are the major outcomes of the analysis:

- A value greater than 0.8 for R^2 is considered a good correlation between observed and simulated data. However, in case of rainfall, to have a good correlation (more than 0.8) using daily rainfall data is very difficult, as daily rainfall is not normally distributed and is highly unpredictable. Although many of the literatures have considered that, R^2 value more than 0.6 gives acceptable simulation results in case of rainfall, **there is need to develop better correlated models as it induces more uncertainty to the projected series.**

- It was observed that, although the same GCM variables were used for homogeneous clustering as well for downscaling, the correlation coefficients do not show any consistency within a cluster. The regression models are different for different stations within a cluster, although the areal distances between those stations were not very large. **This indicates that orography plays a very important role in rainfall variations at a local scale.**
- The R^2 values for daily maximum and minimum temperature are showing good results. Hence, **use of local scale orographic arrangement may not be considered essential in developing downscaling model for temperature.**

Chapter 5

ASSESSMENT OF OROGRAPHIC EFFECT ON MODEL SIMULATIONS BY INCORPORATING OROGRAPHIC FACTOR

5.1. INTRODUCTION

It is a well-known fact that orographic arrangement of a region affects the meteorological behaviour of that particular region. Specially, when the study area is small, the local effects of orography should be incorporated for any hydrological study. GCM models are of coarse resolution, hence most of the time they fail to explain the local scale impacts of meteorological parameters and the orographic arrangement. Some studies are there which incorporated the topography of the study area for analyses, but no detailed study has been done on climate change considering the effects of the orographic arrangements. Hence there is a need for developing methodologies for climate change impact studies which will take into account for the overall orographic arrangement, including locational parameters (latitude, longitude and elevation) and the complexity of the terrain.

The objective of this chapter is to introduce a new factor based on orographic arrangement in the study area, which can be added in the models to improve their efficiency in calibrating and simulating future rainfall series. At first, to understand the complexity of the terrain, topographical characteristics of the region have been studied by performing terrain analysis using QGIS software. Furthermore, the wind data has also been analysed to understand the wind flow pattern in the region. The orographic factor has been derived based on the location and height of hills near the stations, distance of a station from hill, elevation of the stations, wind flow direction and seasonality factor. The variable has been included in the downscaling models and its efficiency has been assessed. Finally, the results and concluding remarks are presented.

5.2. METHODOLOGY

5.2.1. QGIS 2.18.2

In this study, the software QGIS (version 2.18.2) is used for studies related to topographic characteristics of the region. QGIS is an open source GIS software that is freely available with

a user-friendly graphical user interface (GUI). QGIS is licensed under the GNU General Public License (GPL). This software is an official project of the Open Source Geospatial (OSGeo) Foundation that started in May-Jun' 2002. Traditional GIS softwares are expensive and complex, whereas QGIS is free and feasible for anyone with basic access to a computer. QGIS currently runs on most Unix platforms, Windows, and macOS and is developed using the C++ and Qt toolkit, which makes it so comprehensible.

QGIS 2.18.2 holds new features and extends the programmatic interface over previous versions. This version contains a lot of bug fixes and several new features and improvements over the previous version QGIS 2.14 in symbology, labelling, rendering, data management, forms and widgets, map composer, processing, data providers, QGIS server, plugins, programmability, new pyqgis classes, new expression functions, etc.

5.2.2. Terrain analysis using QGIS 2.18.2

To study the topographical characteristics, terrain analysis of the region can be done with the help of QGIS software. QGIS allows us to do analysis of some features like **hillshade, contour, slope, aspect** and **Terrain Ruggedness Index (TRI).**

The DEM that we use for topographical study gives an abstract detail of the elevation of the terrain. Though it contains all the information about the terrain, because of the 2D appearance it looks like a basic map. To get a better appearance at the terrain, hillshade analysis can be done, which is a raster that creates a 3D-looking image by mapping the terrain with the use of light and shadow effect. Another feature that helps in knowing the variations in elevation is contour. More the number of contours, more is the variations in elevation for a particular contour interval and vice versa.

Another helpful thing in topographical study is to identify how steep the terrain is. Slope analysis and aspect analysis help in this regard. The slope is defined as the change in elevation with respect to the horizontal distance between two points. It represents the steepness or inclination of the ground. Buytaert et al. (2006) and Kumari et al. (2017) have used slope as one of the topographical variables for spatial interpolation of rainfall in mountainous regions. Usually, slope can be measured as either a degree value from one grid cell in the raster to the next (from 0 to 90 degree) or as a percent ratio (change in elevation ÷ horizontal distance = slope %). Various guidelines could be found that categorize the slope into different steepness

ranges. The one that is found to be useful was initially defined by the Barcelona Field Studies Centre (BFSC), as Slope Steepness Index (SPI). The categories were slightly modified by combining some ranges for ease of use and better visualization of the map, in this particular study only. The modified SPI ranges are given in Table 5.1.

Table 5. 1 Slope Steepness Index (SPI)

Slope (%)	Approx. Degree	SPI (by BFSC)	Modified SPI
0.0–0.5	0.0	Level	Level slope (0-5%)
0.5–2.0	0.3–1.1	Nearly level	Moderate slope (5%-15%)
2.0–5.0	1.1–3.0	Very gentle slope	
5.0–9.0	3.0–5.0	Gentle slope	
9.0–15.0	5.0–8.5	Moderate slope	
15.0–30.0	8.5–16.5	Strong slope	Strong slope (15%-30%)
30.0–45.0	16.5–24.0	Very strong slope	Very strong slope (30%-45%)
45.0–70.0	24.0–35.0	Extreme slope	Extreme slope (45%-100% or more)
70.0–100.0	35.0–45.0	Steep slope	
> 100.0	> 45.0	Very steep slope	

The aspect of a terrain indicates the direction it's facing in. It can be considered as the direction of the slope. Aspect detects the downslope direction of the maximum rate of change in value from each cell to its neighbors. Typically, aspect is described by dividing the angle of slopes in different directions into eight quadrants: North, North-East, East, South-East, South, South-West, West and North-West. An aspect value of 0 indicates that the slope is North-facing. Similarly, 90 is East-facing, 180 is South-facing and 270 is West-facing and so on.

Terrain heterogeneity is another useful as well as important variable in topographical study. Slope and aspect analysis help in understanding the steepness and the direction of slope, but they do not provide any insight on the terrain roughness. The Terrain Ruggedness Index (TRI) is a topographic index developed by Riley et al. (1999) that gives a quantitative measurement of terrain heterogeneity. This variable helps in understanding whether slope at a location is surrounded by neighbors that are of similar slopes or not, thus giving some idea about the

ruggedness of the region. The TRI is calculated for each cell by considering the elevation of eight neighbor cells in 3×3-pixel grid format using the following formula,

$$TRI = \frac{1}{8}\sqrt{\sum(x_{ij} - x_{00})^2} \quad (5.1)$$

Where, x_{ij} is elevation of each neighbor cell and x_{00} is the elevation of the center cell (grid cell for which TRI is calculated).

The continuous range of TRI values are classified into seven classes of unequal ranges for better understanding (Riley et al., 1999) as given below:

- 0-80 = level terrain surface;
- 81-116 = nearly level surface;
- 117-161 = slightly rugged surface;
- 162-239 = intermediately rugged surface;
- 240-497 = moderately rugged;
- 498-958 = highly rugged; and
- 959-4367 = extremely rugged surface.

The TRI model developed by them is appropriate to use in large area analyses. Although they used 1 km^2 grid cell format for developing the model, it was mentioned that this TRI variable can be used for smaller areas also, provided that higher quality data or DEMs are available for the region.

5.3. DATA COLLECTION

To study the topographical characteristics of the region, following datasets have been used in this study, namely

- SRTM DEM data of 30-meter resolution
- Rainfall data collected from various sources
- Wind data collected from IMD for the 10 stations
- Global Climate model (GCM) data collected from IPCC AR5
    - AR5: GFDL-ESM2G and GFDL-ESM2G (RCP2.6 and RCP8.5)

5.3.1. SRTM DEM data of 30-meter resolution

To study the topographical characteristics of the region, 30-meter Shuttle Radar Topography Mission (SRTM) data covering the whole northeast Indian region are obtained from the NASA servers. Since the study is to be done in a basin or a large watershed level for orographic study, hence 30-meter resolution data are considered to be fair enough. Digital Elevation Model (DEM) files come in tiles of 3601x3601 pixel sizes (1-arcsecond) in a geographic coordinate system projection (EPSG:4326). A total of 54 DEM tiles are obtained covering the region (latitude: 23^0 to 29^0, longitude: 88^0 to 97^0). These DEM tiles are merged and further study is done in the software QGIS 2.18.2. The 30-meter DEM covering the northeast Indian region is shown in Fig. 5.1. The 30-meter DEM is reconstructed for various resolutions (1 km, 2 km, 5 km, 10 km and 20 km) for topographical study, which are given in Annexure G.

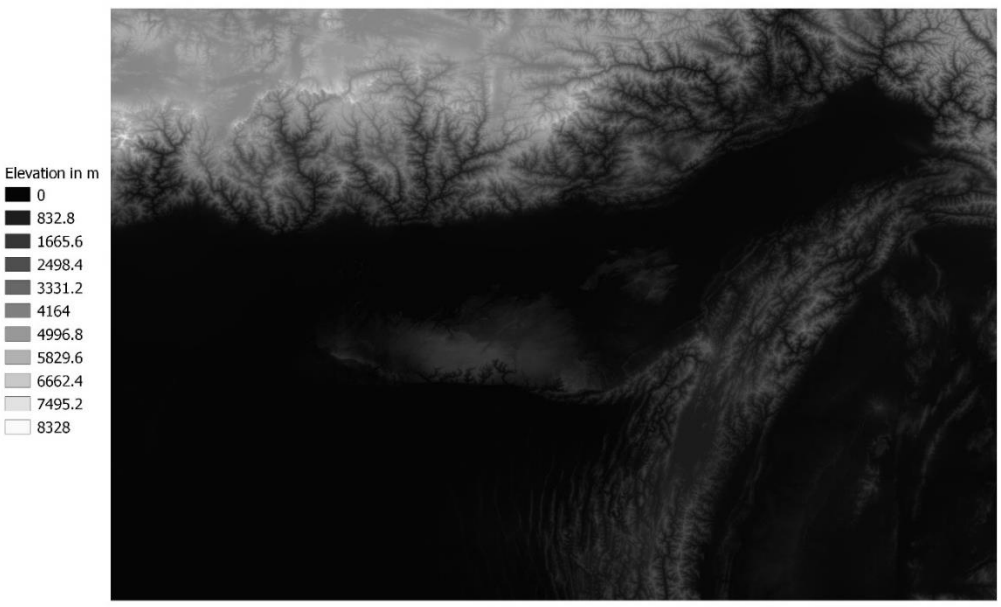


Figure 5. 1 30-meter DEM covering the northeast Indian region

5.3.2. Rainfall data collected from various sources

To find out the correlation of the topographic variables with rainfall, we need the rainfall data of the region at a finer scale. The 10 IMD stations available in the region are situated far apart from each other, which is why they cannot give a good raster map of the region. Also, the gridded IMD data can only generate $0.5^0 \times 0.5^0$ resolution raster map. Hence there is a need of more station points in the region. To fulfill the requirement of these rainfall data, following data are collected from various sources:

i. ***Rainfall data from tea gardens***: These data are collected from various tea gardens of Assam. A total of 105 station rainfall data are found, which are scattered all over Assam. These datasets have various lengths of rainfall series (from 1-year daily data to as long as 84 years).

ii. ***District wise rainfall data***: These data are collected from the site https://knoema.com/, which provides district wise rainfall data all over India. Data are available for 2004-2013, contributed by various sources. The datasets are available in both monthly and annual series.

iii. ***IMD data***: Annual and monthly datasets for 2004-2010 are also available at IMD, for most of the districts of India.

All these station points are shown in Fig. 5.2.

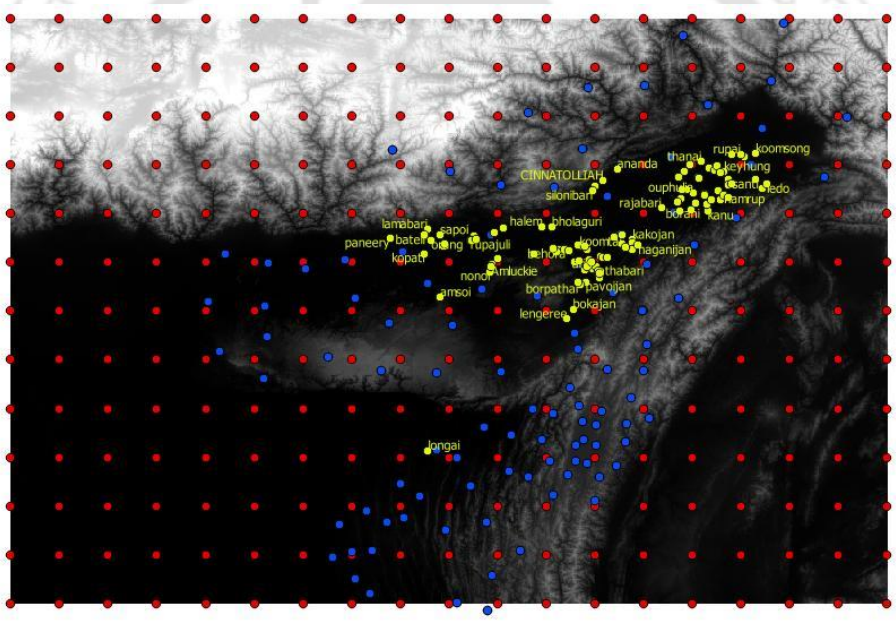


Figure 5. 2 All rainfall station points covering the northeast Indian region (Tea Garden stations: yellow; District wise rainfall data centroids: blue; IMD gridded data: red)

5.3.3. Wind data collected from IMD for the 10 stations

To understand the behaviour of rainfall, orographic arrangements like location of hills, direction of wind and its effect on rainfall are needed to be studied thoroughly. In the present study, IMD wind data are collected to study the behaviour of wind in the Northeast region. The dataset is available from 1969-2011 at the 10 IMD stations, with no data from 2001-2005. The primary and secondary directions of wind in the region at the 10 IMD stations are studied from the available data.

5.3.4. Global Climate model (GCM) data collected from IPCC AR5

The GCM data are taken from Fifth Assessment Report (AR5) of Intergovernmental Panel on Climate Change (IPCC). Two GCM models are used in the study: ESM2G and ESM2M, developed by the Geophysical Fluid Dynamics Laboratory (GFDL), USA. These ESMs are based on a coupled atmospheric and oceanic circulation model, with representations of sea ice, land and iceberg dynamics and they incorporate interactive biogeochemistry, including the carbon cycle. The models are explained in detail in section 3.3.2.2 of chapter 3.

5.4. RESULTS AND DISCUSSION

5.4.1. Study of the topographical characteristics of the northeast Indian region and its relationship with rainfall patterns

5.4.1.1. Study of the rainfall pattern

The new rainfall datasets, along with the previous data (10 IMD station data, IMD grid point data) are clubbed together to find out the rainfall raster maps of the northeast Indian region. The rainfall raster maps for annual average rainfall are generated for 30m, 1 km, 2 km, 5 km, 10 km and 20 km resolution. 30-meter raster map of rainfall is shown in Fig. 5.3 and rest are attached in Annexure G.

The variation of annual rainfall patterns of different stations is clearly visible in the raster map. The region covering the stations Passighat, Mohanbari, North-Lakhimpur has higher rainfall compared to middle region with stations Tezpur, Imphal and Guwahati. The region covering Shillong, Silchar and Cherrapunjee has higher rainfall than any other part of the northeastern region. However, it was realized that the raster generated could not capture the exceptionally high rainfall value of Cherrapunjee station, although it was found to be the highest among all stations. This may be attributed to the influence of gridded rainfall data, which is basically a interpolated dataset. If the rainfall of the five stations under consideration (Cherrapunjee, Imphal, Shillong, Silchar and Guwahati) are compared, the variation of annual rainfall among the station can be clearly seen, although they were grouped in the same homogenous cluster during fuzzy cluster analysis. It indicates that the local orographic arrangement plays an important role in this peculiar region.

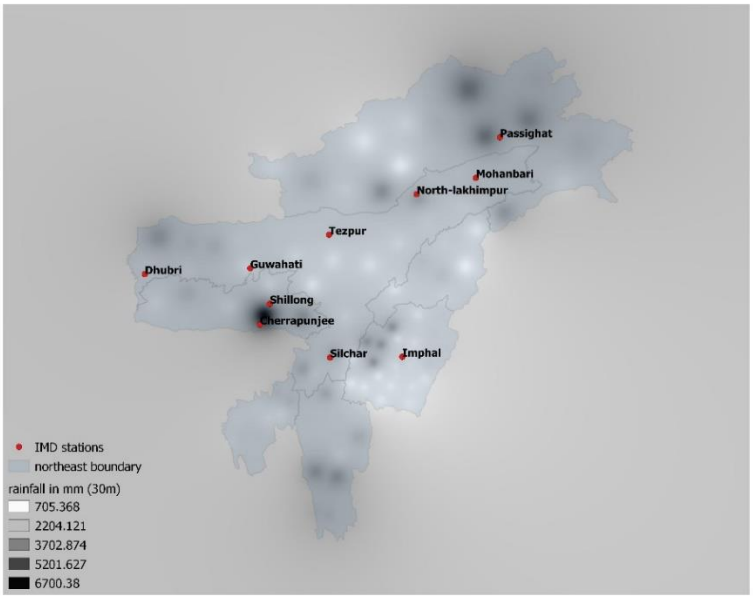


Figure 5. 3 30-meter rainfall raster map of the northeast Indian region for the combined rainfall dataset

5.4.1.2. Terrain analysis of the region

For the present study, **hillshade, contour, slope, aspect** and **Terrain Ruggedness Index (TRI)** were considered to incorporate in the topographical characteristics study for the NE Indian region.

The terrain analyses have been done for the northeastern region with the help of DEM of various resolutions, as mentioned in section 5.3.1. The 30-meter raster map of hillshade analysis is shown in Fig. 5.4, from which the location and extent of hilly region can be easily recognized. Other raster maps, with resolution ranging from 1km to 20km, are given in Annexure G. Stations Mohanbari, North-Lakhimpur, Tezpur, Guwahati and Dhubri are situated in plain area, near the banks of Brahmaputra river. Passighat station is located at the foothills of Eastern Himalaya-Patkai range. Stations Cherrapunjee, Imphal, Shillong and Silchar are located in hilly area and the variation in elevations of the hills can be realized from the map. The variations in elevation can also be explained with the help of contour map, shown in Fig. 5.5. The contour map was generated for 500-meter contour interval. It can be seen that for the five stations under consideration, number of contour lines are more since the variations in elevation are more. Compared to that, other six stations are situated in plain areas with less or no contour lines.

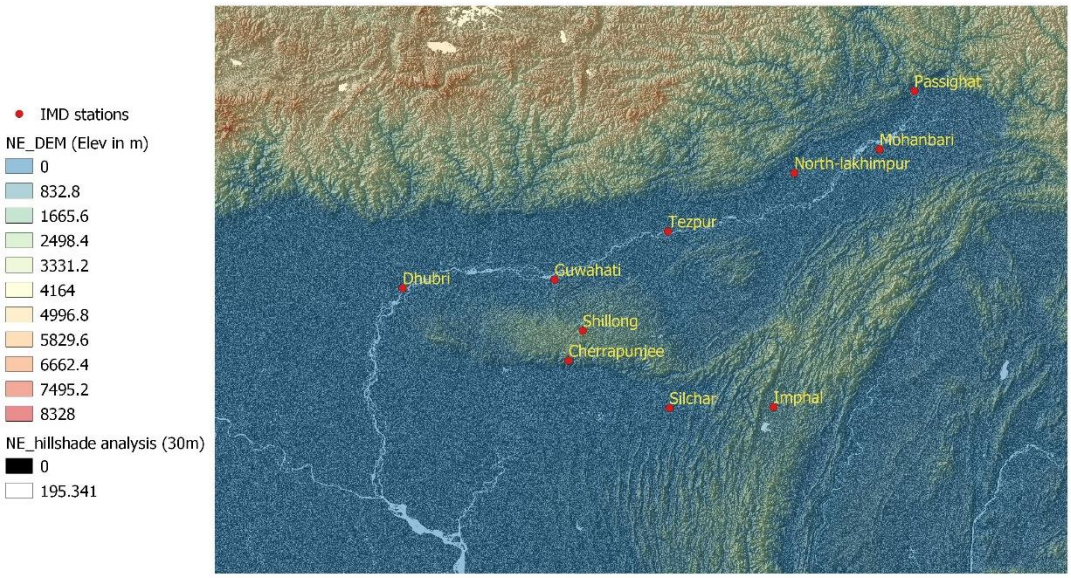


Figure 5. 4 30-meter raster map of hillshade analysis of the northeast Indian region

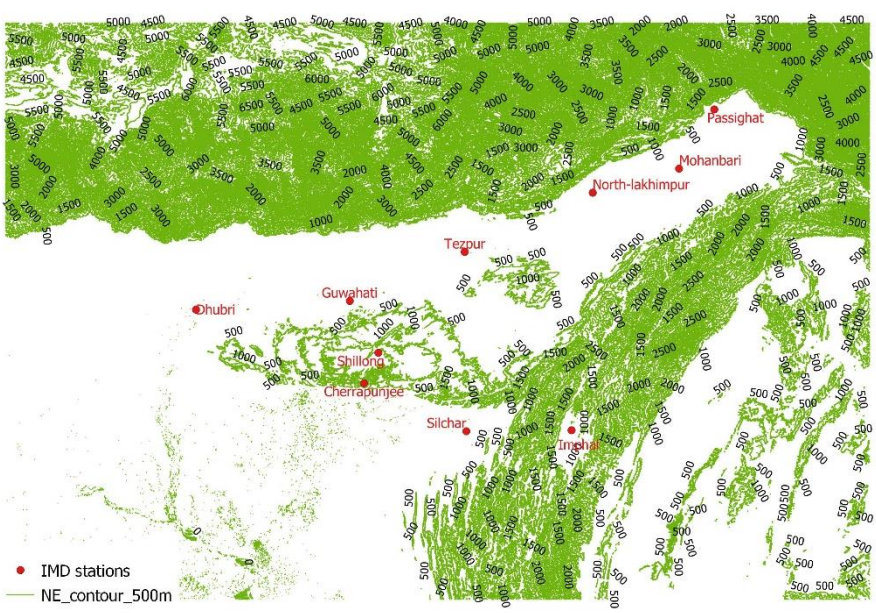


Figure 5. 5 Contour map of the northeast Indian region with 500 m contour interval

To determine the steepness of the study area, slope analysis has been done. 30-meter raster map of slope analysis is shown in Fig. 5.6. Stations Passighat, Mohanbari, North-Lakhimpur, Tezpur and Dhubri are situated in level ground with less than 5% slope steepness index (SPI). Whereas, stations Cherrapunjee, Imphal, Shillong, Silchar and Guwahati are located in areas with high SPI values. Variation in slope steepness in the region covering these five stations can be described from the figure. The adjacent areas of stations Guwahati and Silchar have

moderate to strong slopes. Shillong and Cherrapunjee are situated in an area with very strong to extreme steep slopes. Although Imphal is at a relatively level ground, the station is surrounded by extreme steep slopes with very high SPI values.

The direction of steepness of the hills are determined from aspect analysis, 30-meter raster map of the same is shown in Fig. 5.7. Raster maps of slope and aspect analysis with different resolutions are given in Annexure G.

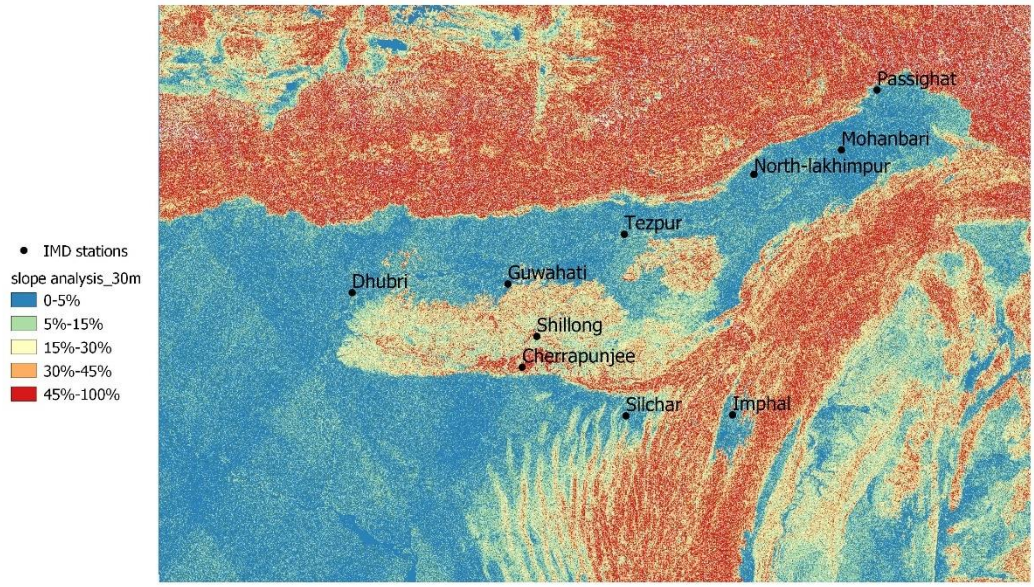


Figure 5. 6 30-meter raster map of slope analysis of the northeast Indian region

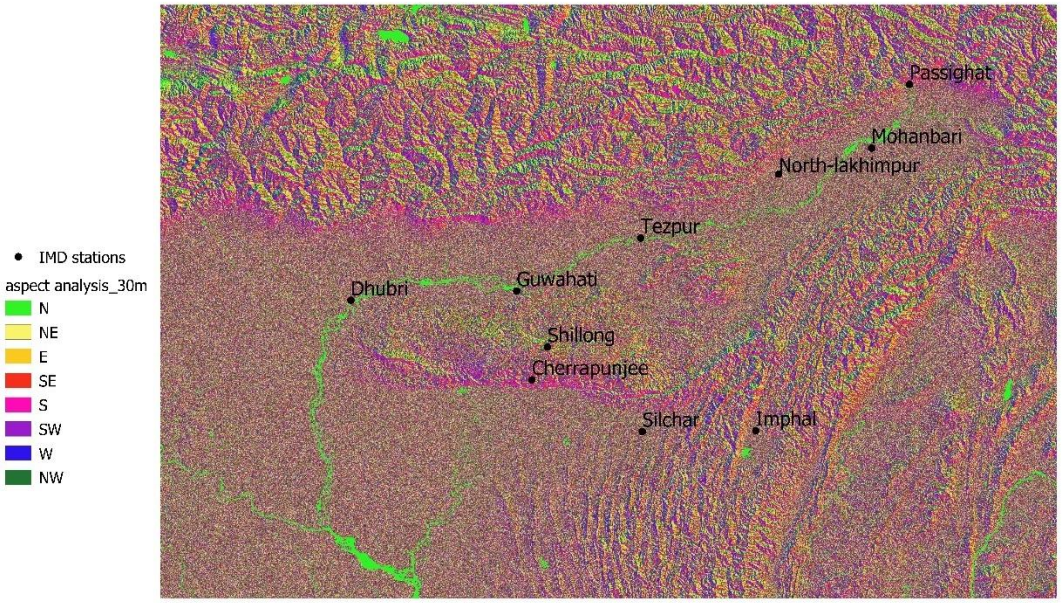


Figure 5. 7 30-meter raster map of aspect analysis of the northeast Indian region

TRI analysis has been done to understand about the ruggedness of the region, i.e. whether slope at a location is surrounded by neighbors that are of similar slopes or not. Riley et al. (1999) stated that the TRI model developed by them is appropriate to use in large area analyses and used 1 km^2 grid cell format for developing the model. Fig. 5.8 shows the 1-km raster map of TRI analysis for the northeastern Indian region. It is evident from the figure that the region covering the stations Cherrapunjee, Imphal, Sillong, Silchar and Guwahati has higher ruggedness, compared to other five stations. The ruggedness index varies from slightly rugged to highly rugged for the region, indicating large variation in slopes. Raster maps of TRI analysis with different resolutions are given in Annexure G.

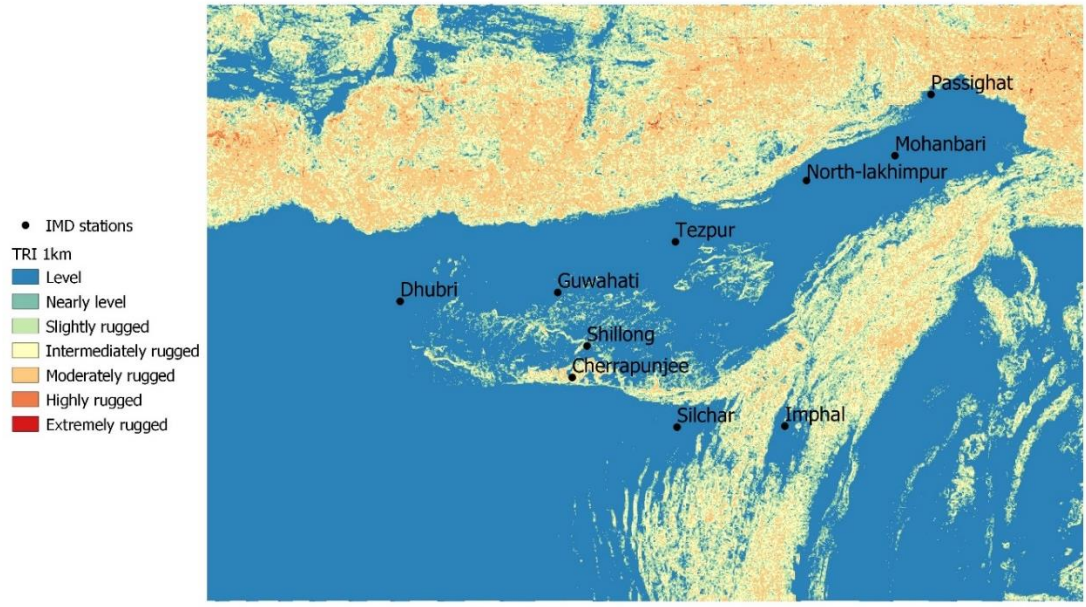


Figure 5. 8 Raster map (1km) of TRI analysis of the northeast Indian region

5.4.2. Analysis of wind data

To develop the orographic factor, daily wind data has been collected from IMD for 1969-2011 and the variation in the wind flow pattern has been studied rigorously on the 10 IMD stations. The primary and secondary directions of wind at those stations were assessed, both for annual and monthly data by taking mode of the individual datasets. However, only results of the 5 stations under consideration are presented in this section. Other stations results are not given for brevity.

For better understanding the wind directions are shown in Fig. 5.9. It was found that, the moisture bearing wind mainly comes from the south west direction. The primary and secondary

directions for annual and monthly data are given below. Here, **N** means wind flowing toward the north direction, **NE** means towards north-east direction, and so on.

1. **Cherrapunjee**

Annual	primary direction	= **NE** direction for all years
	secondary direction	= **NE-E** direction for all years
Monthly	primary direction	= **SW** direction from Feb to Sept
	secondary direction	= **W** direction Feb to May;
		= **E** direction June to Jan

2. **Imphal**

Annual	primary direction	= **S** direction for all years
		= **NW** direction from 1969-89;
		= **W** direction from 2001-12
	secondary direction	= **SE** direction for all years
		= **W** direction from 1969-89;
		= **NW** direction from 2001-12
Monthly	primary direction	= **S** direction for all months
		= **NW** direction for all months
	secondary direction	= **SE** direction for all months
		= W direction for all months

3. **Shillong**

Annual	primary direction	= **NE** direction for all years
	secondary direction	= **W** direction for all years
Monthly	primary direction	= **SW** direction from Feb to Sept;
		= **E** direction from Oct to Jan
	secondary direction	= **W** direction Feb to Sept;
		= **SE** direction Oct to Jan

4. **Silchar**

Annual	primary direction	= **E** direction for all years
	secondary direction	= **NE** direction for all years
		= **NW** direction for all years
Monthly	primary direction	= **E** direction for all months;

= **W** direction for all months

secondary direction = **NE** direction for all months;

= **NW** direction for all months

5. Guwahati

Annual	primary direction	= **NE** direction for all years
	secondary direction	= **E** direction from 1972-93;
		= **N** direction from 1994-12;
		= **W** direction for all years
Monthly	primary direction	= **NE** direction for all months
	secondary direction	= **E** direction for all months
		= **W** direction from Feb to Sept;
		= **E** direction from Oct to Jan

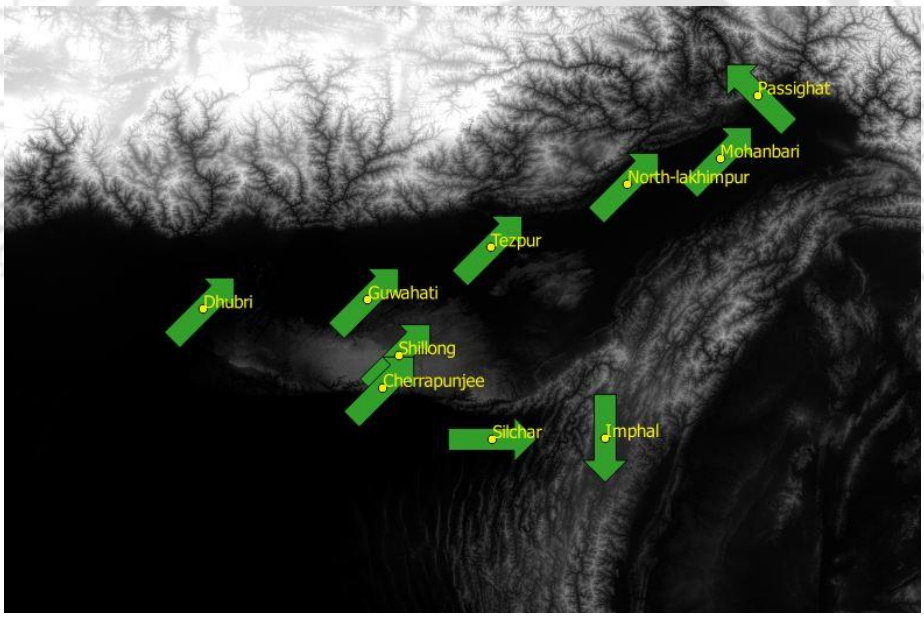


Figure 5. 9 Primary wind directions at the 10 IMD stations in the northeast

5.4.3. Development of a new downscaling approach with inclusion of the orographic factor (O_f)

Due to variations in wind direction within a year, the highest elevation of the hill obstructing the wind flow as well as distance of the station from that hill also varies with time. Based on this, the orographic factor (O_f) is developed.

Inclusion of the orographic factor (O_f) in the model calibration can be expressed by the following relation

Rainfall = f {GCM predictors, Orographic factor (O_f)}

$$O_f = f(S, E_s, E_h, d_w or d_l) \quad (5.2)$$

where, S = Seasonality factor; E_s = Elevation of the station (in m); E_h = Elevation of the hill (in m); d_w = distance of the station from hill in km (if on windward side); d_l = distance of the station from hill in km (if on leeward side).

Here, S, E_h and d_w (or d_l) are time dependant variables, whereas E_s is a constant value. From general observation as well as correlating the above parameters with rainfall pattern, following equations were formed:

For stations on the windward side, $O_f = \frac{SE_sE_h}{d_w}$ (5.3)

For stations on the leeward side, $O_f = \frac{Sd_lE_s}{E_h}$ (5.4)

The orographic factor was evaluated for all the five stations and its correlations were found with monthly rainfall of the respective stations. Table 5.2 presents the Pearson correlation (PC) values of the stations. The location of the stations based on annual primary wind direction are also given in the table.

Table 5. 2 Correlation (PC) between monthly rainfall and orographic factor and location of a station based on primary wind direction

STATION NAME	PC	LOCATION
Cherrapunjee	0.67	windward
Imphal	0.79	leeward
Shillong	0.74	leeward
Silchar	0.86	windward
Guwahati	0.76	leeward

Same procedure was followed for screening of predictors (Mahmood and Babel, 2013). Observed daily dataset 1971–2000 is used for the calibration of annual rainfall for all stations (1986-2000 for Silchar station). In this study, SDSM is developed using annual model only, as it was found to be showing better relation in the previous analysis. In this case, orographic

factor (O_f) is included in the regression process along with other selected predictors. Bias correction was done and the models were validated with observed data by calculating the coefficient of determination (R^2). Fig. 5.10 shows the R^2 values between the observed rainfall and bias corrected simulated rainfall for GCM model and the new model (GCM+O_f).

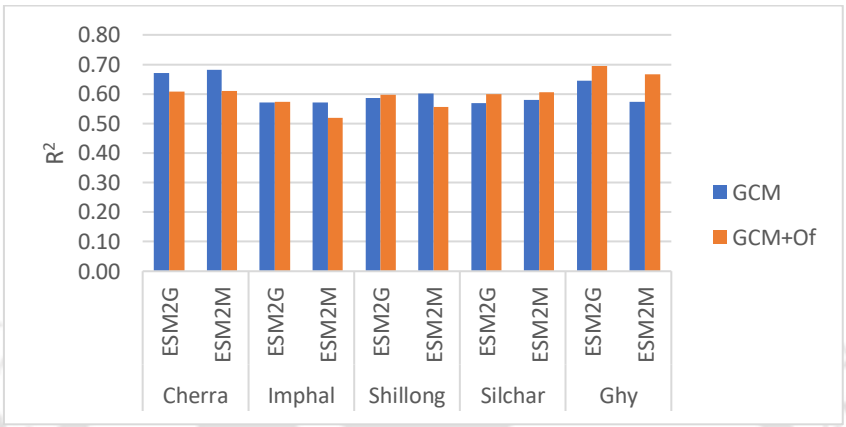


Figure 5. 10 Coefficient of determination (R^2) between the observed and simulated annual rainfall for GCM model and the new model (GCM+O_f)

From the above figure it was found that, the orographic factor could not improve the R^2 values for stations Cherrapunjee, Shillong and Imphal. However, for stations Guwahati and Silchar, the orographic factor has exhibited increase in the R^2 value to a decent extent. Hence, for a station that is very near to a hill or located above the hill, not much difference is found in conventional GCM downscaling approach and the new approach with orographic factor. But, for stations at an identifiable distance from a hill, improvements can be seen. Furthermore, from terrain analysis it was found that the topographical variations are more in the nearby areas of the stations Cherrapunjee, Shillong and Imphal as compared to the other two stations. This indicates that the orographic factor is still not entirely able to capture the local orographic effect for these stations.

Moreover, the R^2 values still could not meet the targeted value i.e. 0.8. Hence, there is a need for more specific approaches that can further strengthen the orographic factor and give better results. For example, daily variation of wind flow direction may be included in formulating the orographic factor, instead of annual or seasonal wind direction.

5.5. SUMMARY

Following are the major outcomes of the analysis:

- Inclusion of orographic factor does impart some changes in the R^2 value. For stations at an identifiable distance from a hill (Guwahati, Silchar) the orographic factor has exhibited increase in the R^2 value to a decent extent. It appears that moisture bearing monsoonal wind while crossing the Khasi-Jayantia hill range moves high and get condensed. As seen in the figure 5.9 this moisture bearing wind after crossing the hills influence the rain of Guwahati station. A part of this moisture bearing rain moves in the right direction towards Silchar and influences the precipitation there.
- However, the difference in R^2 values before addition of orographic factor and after addition is not very high.
- For stations located on the hill (Cherrapunjee, Shillong, Imphal), the orographic factor could not improve the R^2 value. Hence, not much difference is found in conventional GCM downscaling approach and the new approach with orographic factor.
- More specific approaches are required to strengthen the orographic factor to get better results. Daily variation of wind flow direction may be included in formulating the orographic factor, instead of annual or seasonal wind direction.

Chapter 6

COMPARISON OF DIFFERENT MODEL SIMULATIONS AND ASSESSMENT OF ASSOCIATED UNCERTAINTIES

6.1. INTRODUCTION

Climate projections are usually associated with various uncertainties arising from a number of known and unknown sources. For example, uncertainty due to emission scenarios, downscaling methods, inter-model uncertainty, intra-model uncertainty etc. Hence, to get a reliable result there is need to assess the uncertainties associated with a model simulation. In this study, 3GCM models (HadCM3, ESM2G, ESM2M) with 3 different scenarios (A2, RCP2.6, RCP8.5) are used to project future conditions. Two different methods are used to downscale the models: MLR and SDSM. Hence there is a need to compare the projection results also. This chapter summarizes the results of future projection for all three GCM models for their respective scenarios. Uncertainly analysis for average daily maximum and minimum temperature and average annual rainfall has also been presented and comparisons were made with the new model.

6.2. METHODOLOGY

From previous studies, it has already been established that use of single GCM or single scenario leads to unreliable projections. Hence, researchers usually go for multi-model projections; average of all the models under consideration is then analysed to get the best possible projection. Nevertheless, it is always necessary to highlight the uncertainties among the models to understand the extent of the unreliability. This kind of uncertainty analysis is termed as inter-model uncertainty.

The intra-model uncertainty occurs due to different realization of the same GCM. In this study, different possible scenarios (or ensembles) of synthetic daily weather series has been generated during simulation (20 ensemble series) and prediction (50 ensemble series), using SDSM software. Each individual ensemble member is considered as equally possible climate projection. The extent to which these ensembles vary from each other depends on the significance of the stochastic as well as the deterministic components of the regression model Wilby and Dawson (2007). Unconditional or independent parameters like temperature,

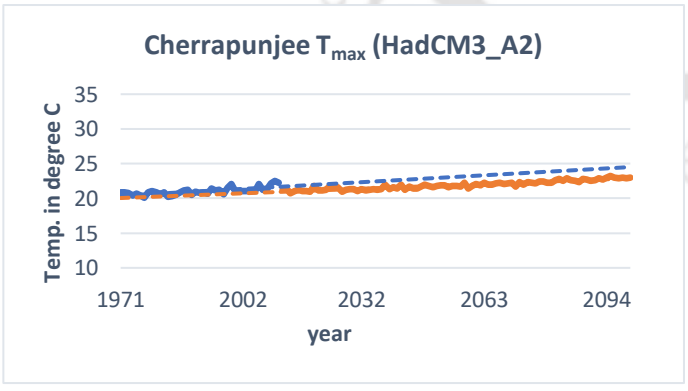
humidity etc. produces ensembles with less variations. But, in case of conditional variables like precipitation, evaporation, runoff etc, the difference among the ensemble members is large.

In the present study, inter-model uncertainty is not assessed. Instead, the intra-model uncertainty is assessed to understand the influence of the orographic factor on the inherent uncertainty of the downscaling process. At first, the simulated and predicted series from different ensemble members has been assessed to understand the uncertainty associated with it. Then, the uncertainty of the simulated and projected series of the new model, with the use of orographic factor, has been evaluated. The purpose of this uncertainty analysis is to assess the extent to which the orographic factor has been successful in influencing the intra-model uncertainty. However, comparison among different model simulations has been made to understand the inter-model variances.

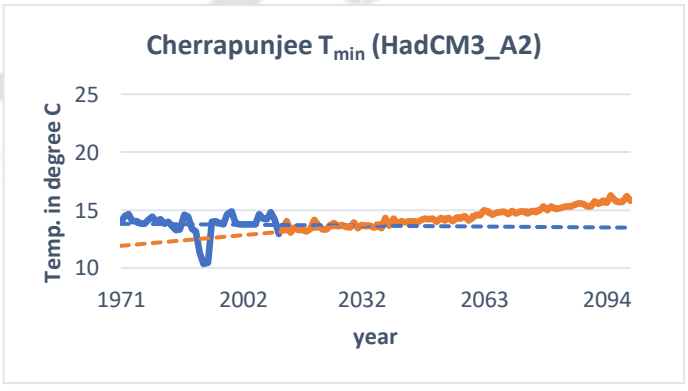
6.3. RESULTS AND DISCUSSION

6.3.1. Future projection of HadCM3 A2 scenario

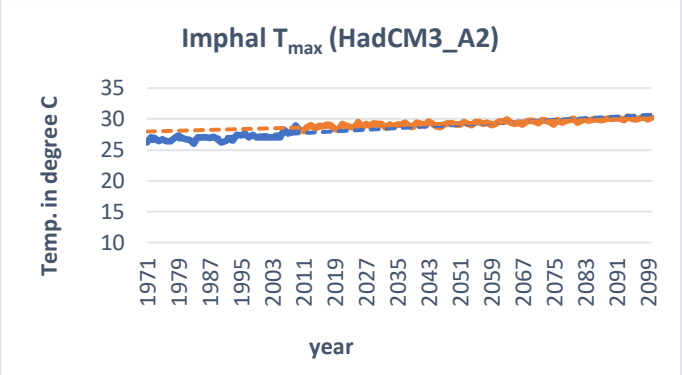
The downscaling models developed in the previous sections were applied accordingly and projected to the year 2100. The simulations and future projections for HadCM3 model were done by multiple linear regression. Fig. 6.1 presents the projected future series (line plots) after bias correction of average daily maximum temperature (T_{max}), average daily minimum temperature (T_{min}) and average annual rainfall for observed (1971–2010) and projected period (2011–2100) for model HadCM3. All the future projections have shown increasing trend, except annual rainfall of Cherrapunjee, that has shown a slightly decreasing trend.



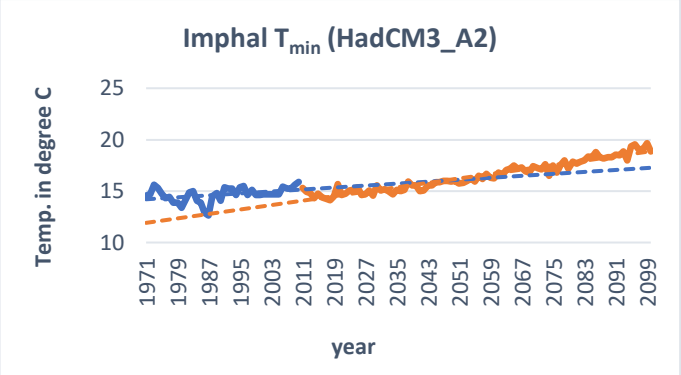
(a)



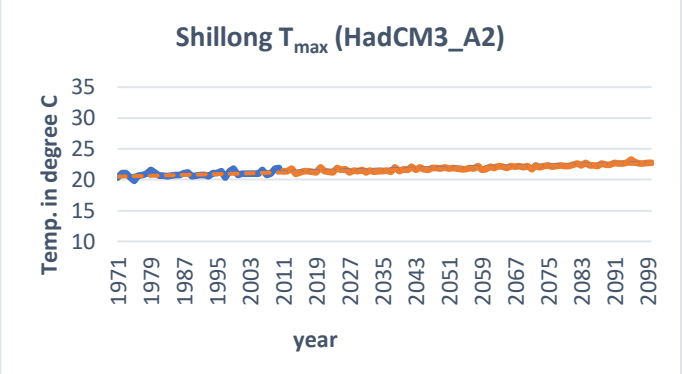
(b)


Imphal T_{max} (HadCM3_A2)
Temp. in degree C
35
30
25
20
15
10
1971 1979 1987 1995 2003 2011 2019 2027 2035 2043 2051 2059 2067 2075 2083 2091 2099
year

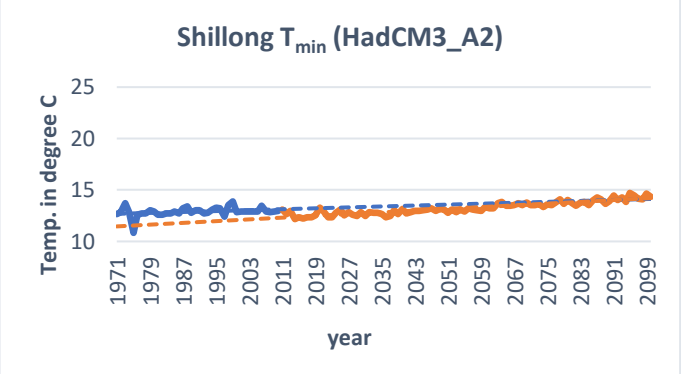
(c)


Imphal T_{min} (HadCM3_A2)
Temp. in degree C
25
20
15
10
1971 1979 1987 1995 2003 2011 2019 2027 2035 2043 2051 2059 2067 2075 2083 2091 2099
year

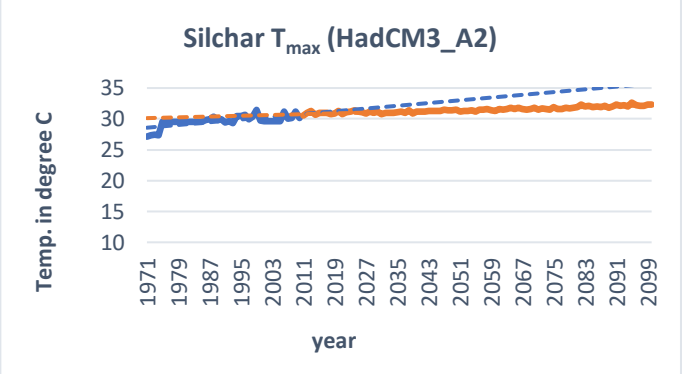
(d)


Shillong T_{max} (HadCM3_A2)
Temp. in degree C
35
30
25
20
15
10
1971 1979 1987 1995 2003 2011 2019 2027 2035 2043 2051 2059 2067 2075 2083 2091 2099
year

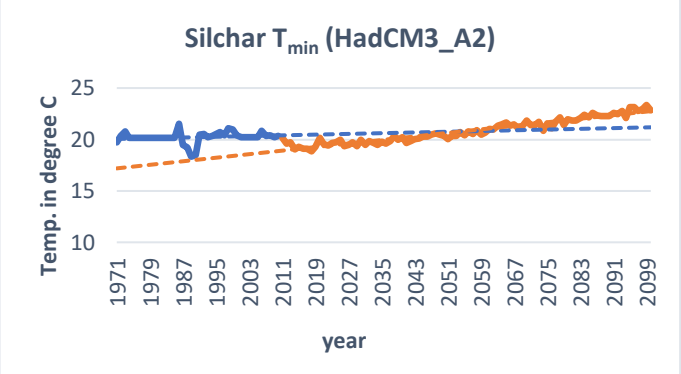
(e)


Shillong T_{min} (HadCM3_A2)
Temp. in degree C
25
20
15
10
1971 1979 1987 1995 2003 2011 2019 2027 2035 2043 2051 2059 2067 2075 2083 2091 2099
year

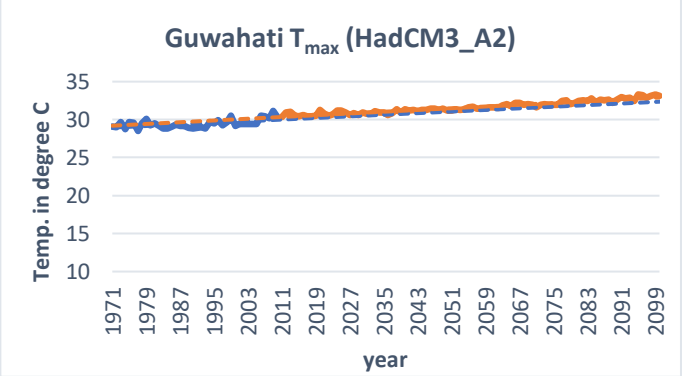
(f)


Silchar T_{max} (HadCM3_A2)
Temp. in degree C
35
30
25
20
15
10
1971 1979 1987 1995 2003 2011 2019 2027 2035 2043 2051 2059 2067 2075 2083 2091 2099
year

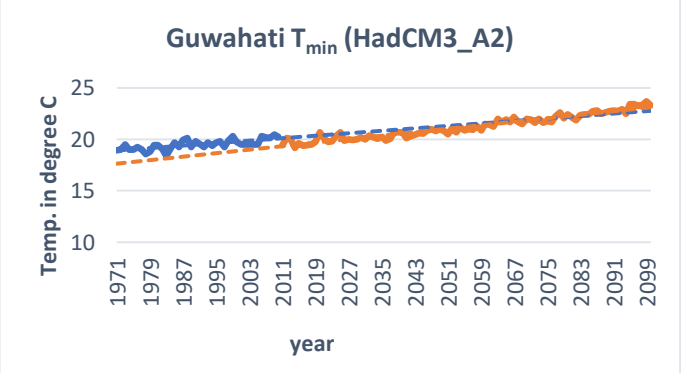
(g)


Silchar T_{min} (HadCM3_A2)
Temp. in degree C
25
20
15
10
1971 1979 1987 1995 2003 2011 2019 2027 2035 2043 2051 2059 2067 2075 2083 2091 2099
year

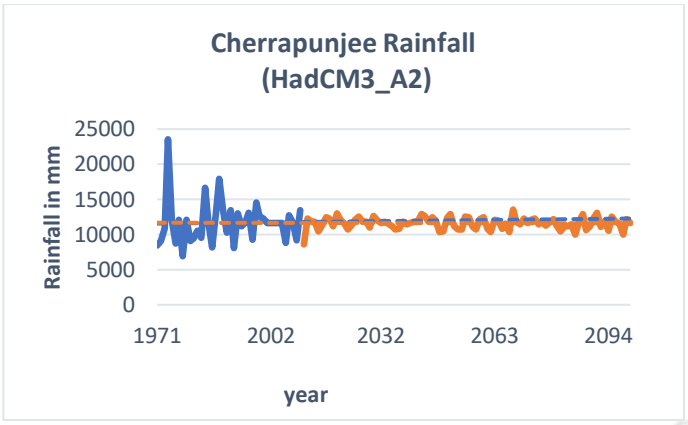
(h)


Guwahati T_{max} (HadCM3_A2)
Temp. in degree C
35
30
25
20
15
10
1971 1979 1987 1995 2003 2011 2019 2027 2035 2043 2051 2059 2067 2075 2083 2091 2099
year

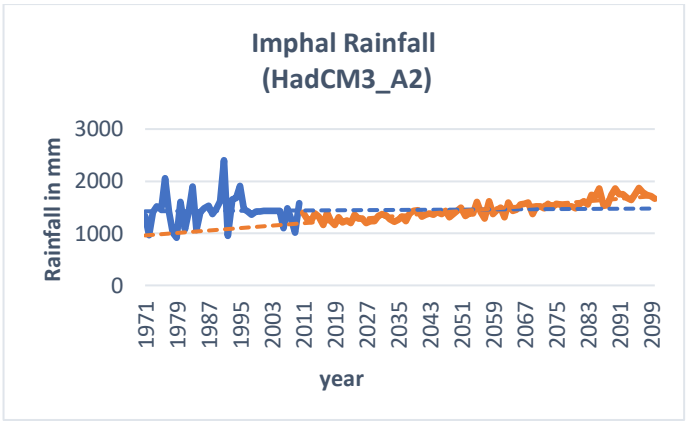
(i)


Guwahati T_{min} (HadCM3_A2)
Temp. in degree C
25
20
15
10
1971 1979 1987 1995 2003 2011 2019 2027 2035 2043 2051 2059 2067 2075 2083 2091 2099
year

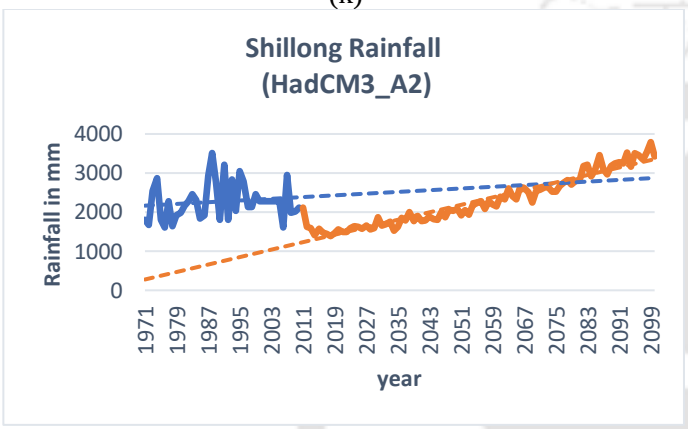
(j)



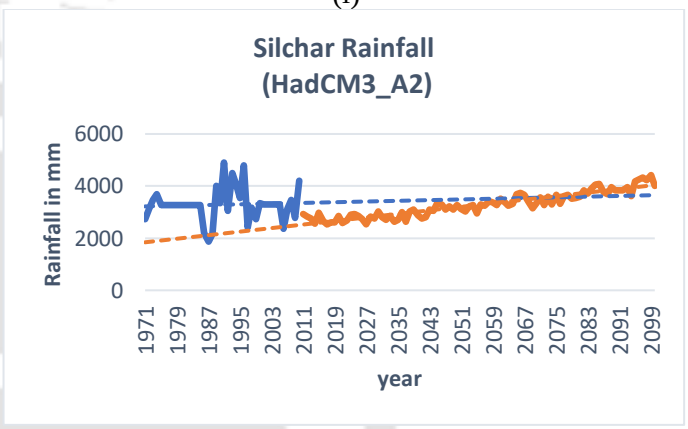
(k)



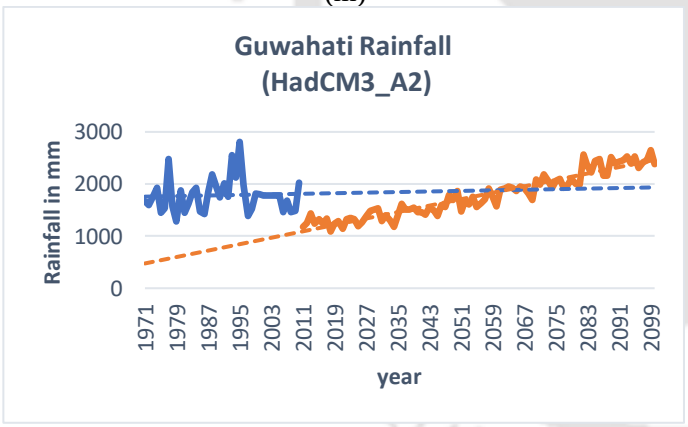
(l)

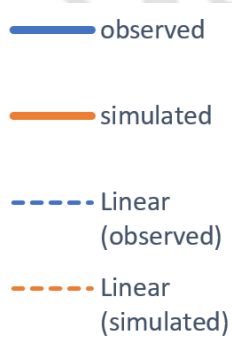


(m)



(n)





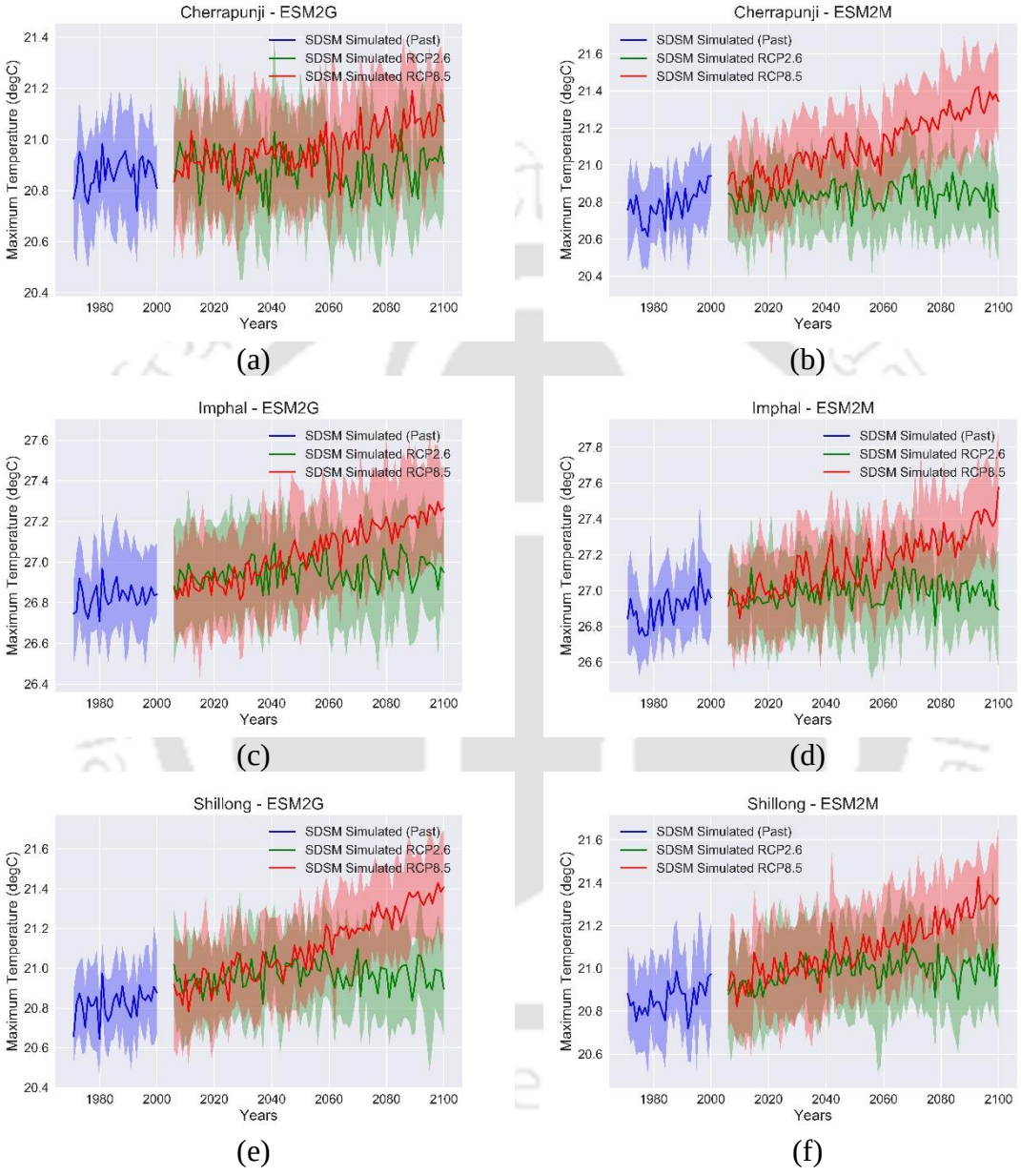
(o)

Figure 6. 1 Line plots of average daily maximum temperature (T_{max}), average daily minimum temperature (T_{min}) and average annual rainfall for observed (1971–2010) and projected period (2011–2100) for model HadCM3

6.3.2. Future projection of ESM2G and ESM2M and their uncertainty analysis

Based on the models developed by SDSM software at a particular site, future scenarios were generated for average daily maximum temperature (T_{max}), average daily minimum temperature (T_{min}), and average annual rainfall after bias correction. The line plots with uncertainty are shown in Fig. 6.2, 6.3 and 6.4 for RCP2.6 and RCP8.5 and for both ESM2G and ESM2M model.

Trend analysis of future projected series for all three parameters (T_{max}, T_{min} and rainfall) has been done, which is presented in section 6.3.4. Also, the uncertainty analysis has been done and percentage uncertainties for all future projections have been evaluated and discussed in section 6.3.5.



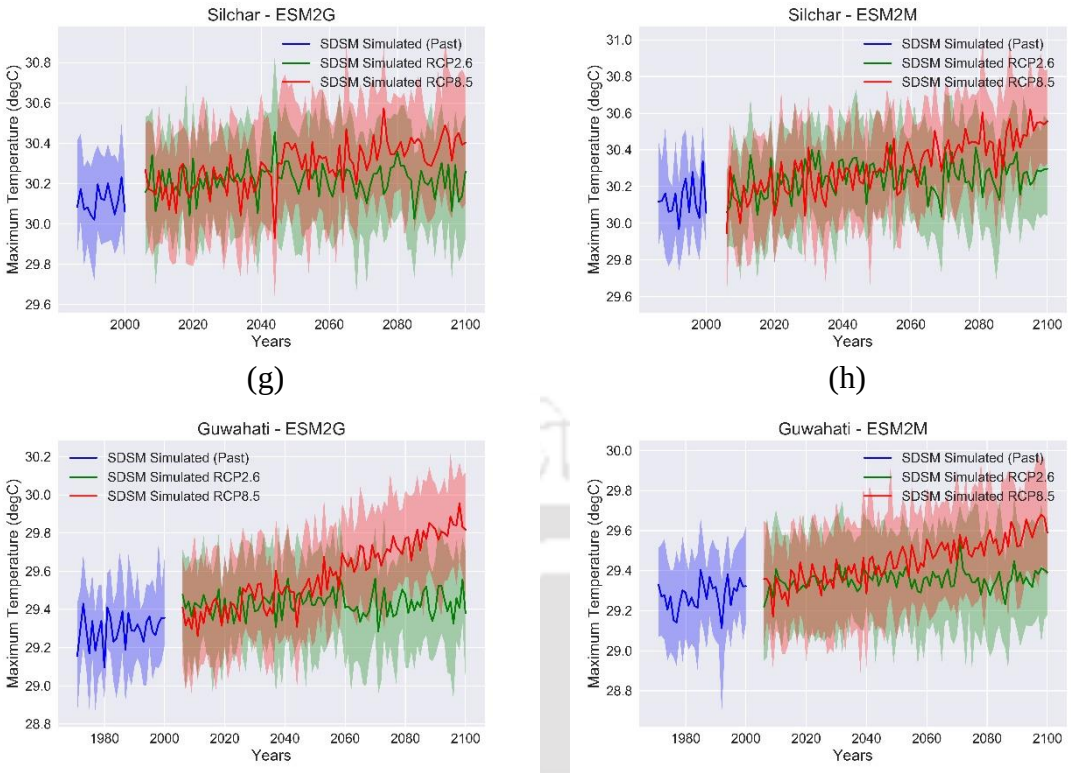
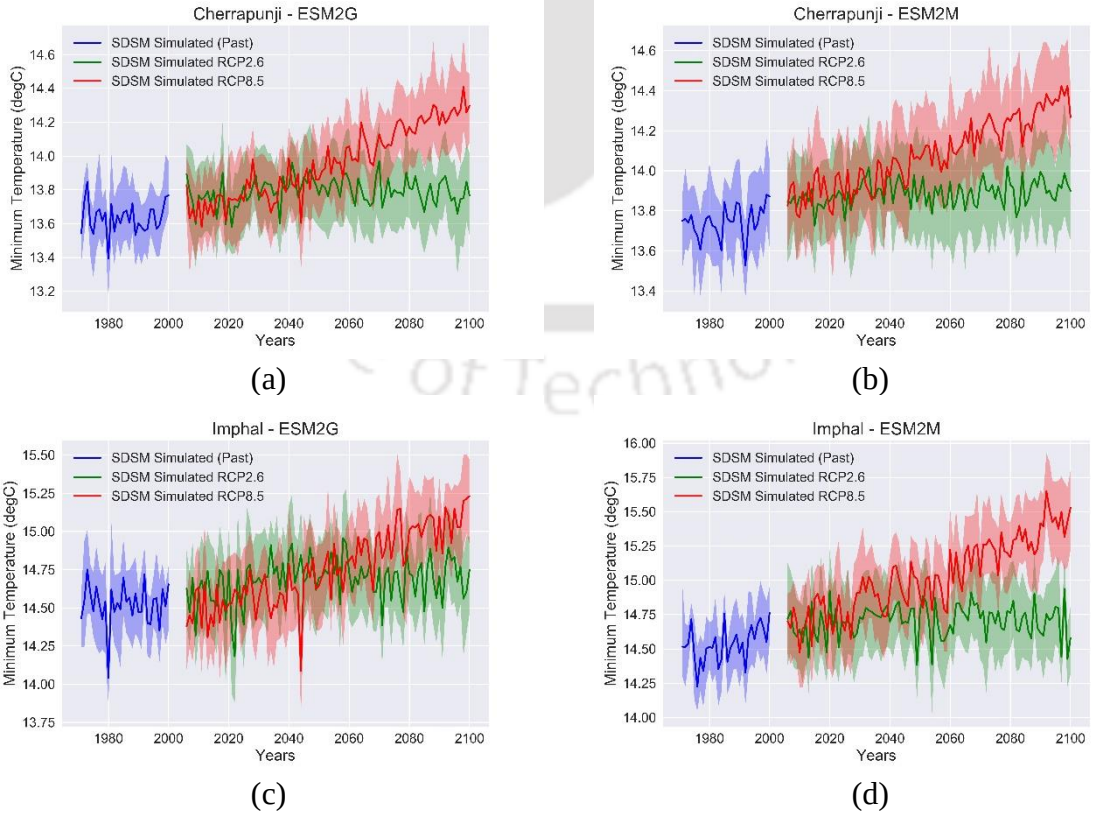


Figure 6. 2 Line plots with uncertainty of average daily maximum temperature (T_{max}) for observed (1971–2000) and projected period (2006–2100) for models ESM2G and ESM2M



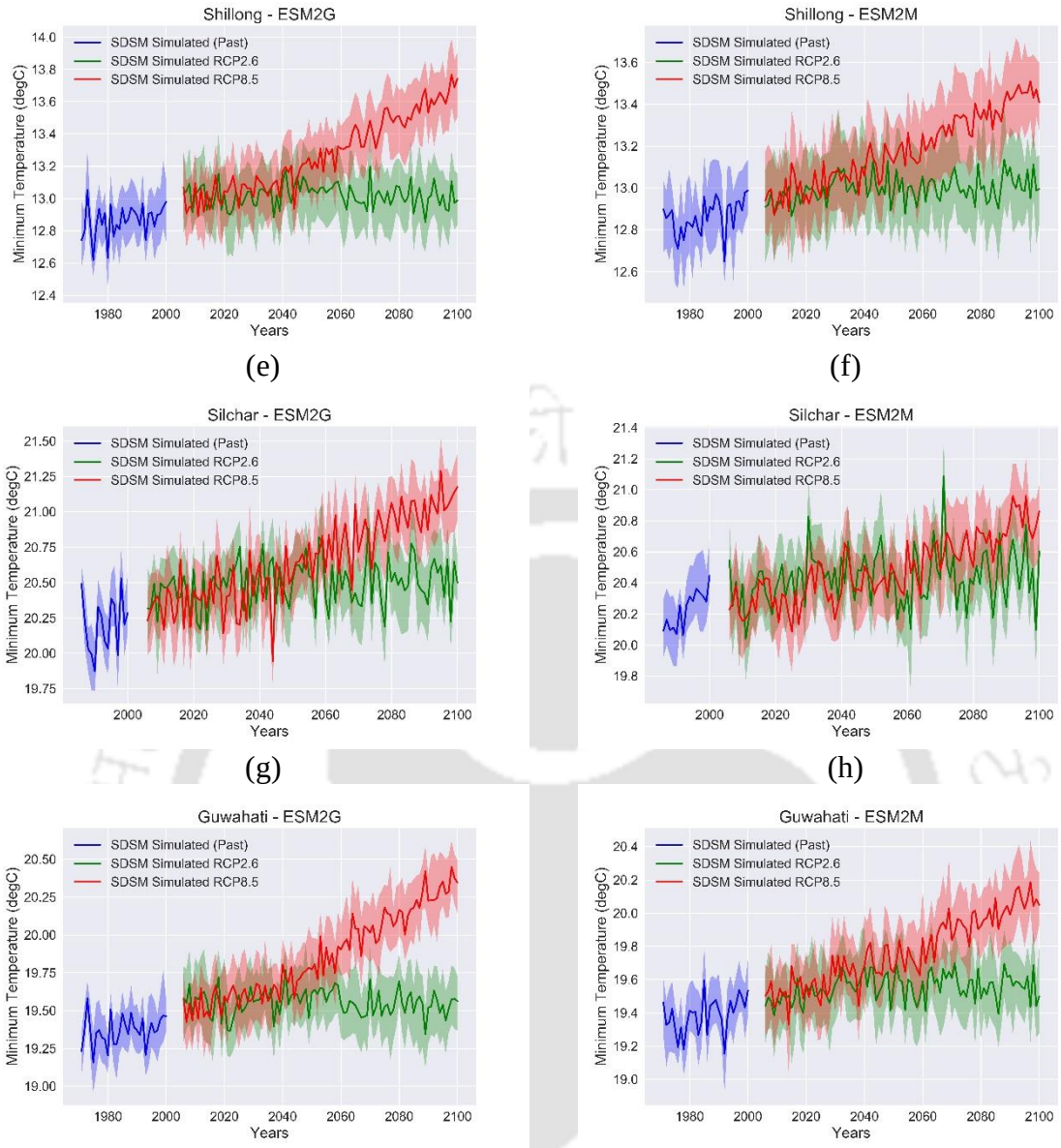
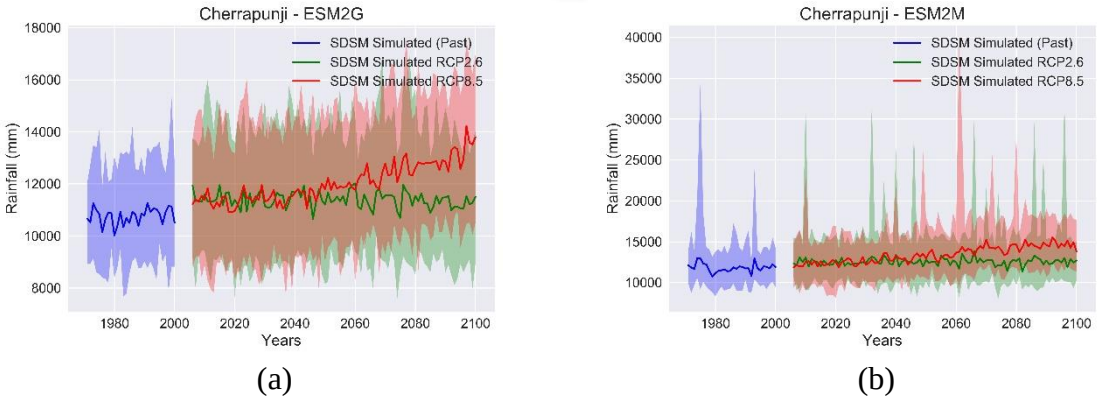


Figure 6. 3 Line plots with uncertainty of average daily minimum temperature (T_{min}) for observed (1971–2000) and projected period (2006–2100) for models ESM2G and ESM2M



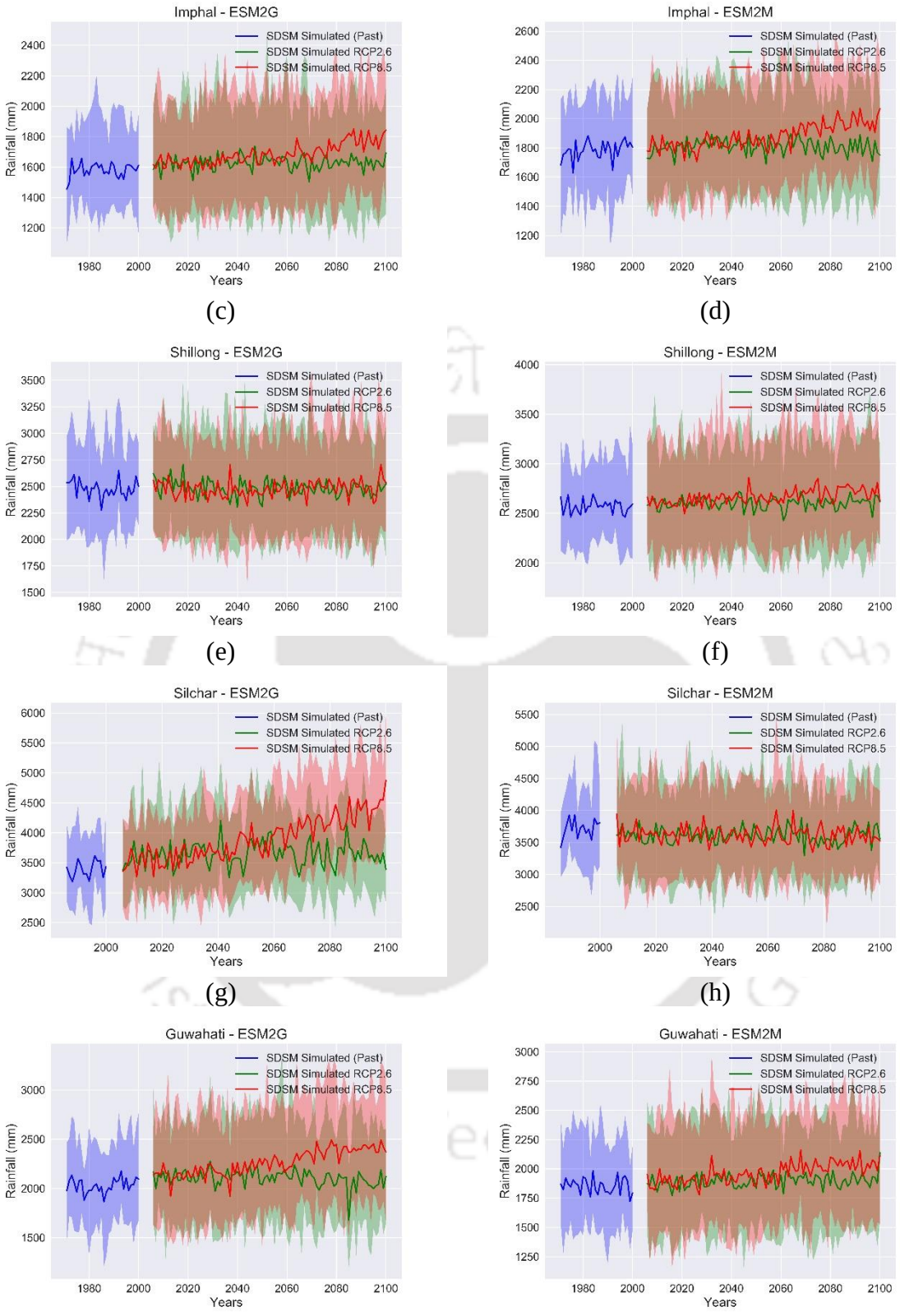
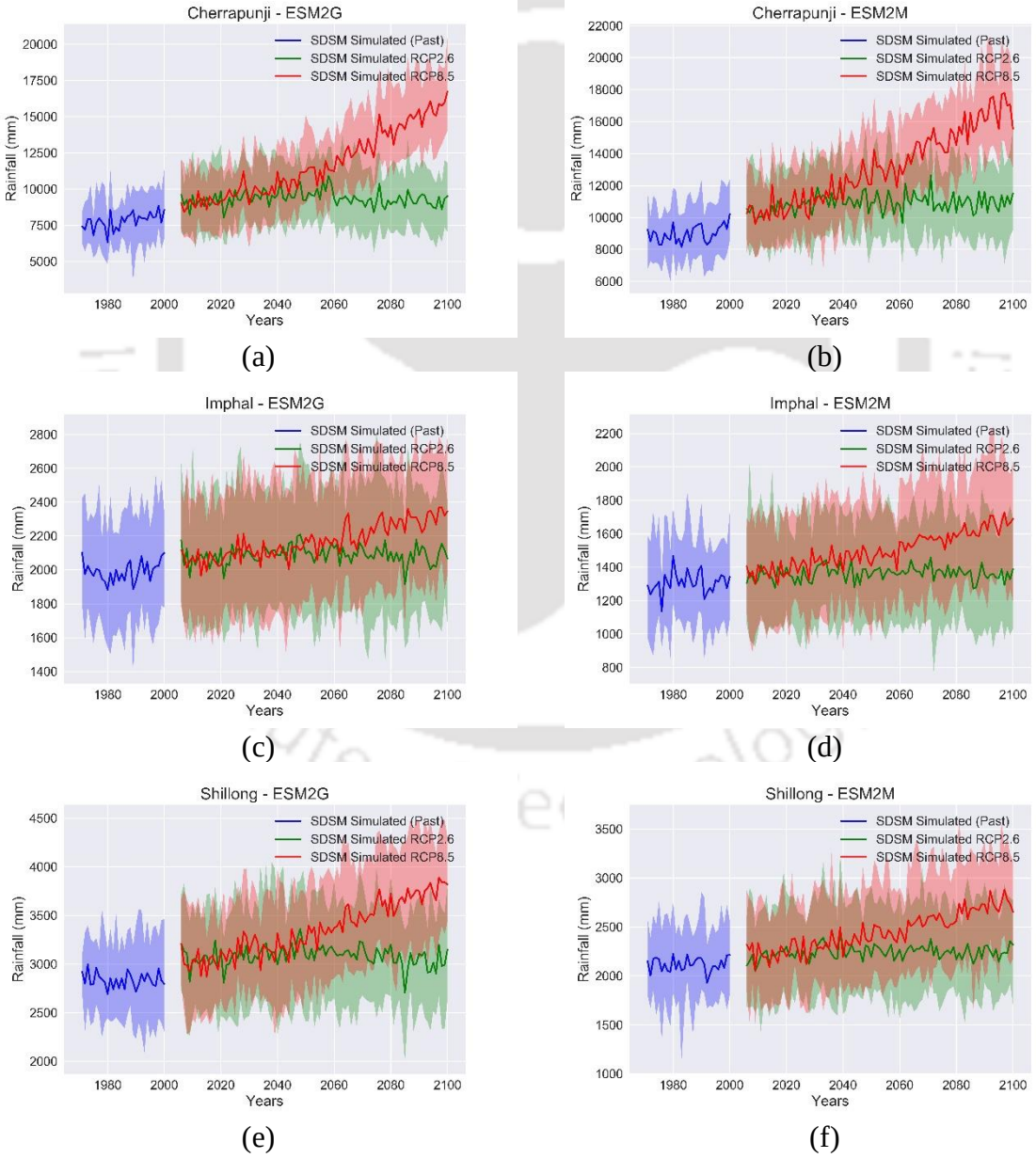


Figure 6. 4 Line plots with uncertainty of average annual rainfall for observed (1971–2000) and projected period (2006–2100) for models ESM2G and ESM2M

6.3.3. Future projection of ESM2G and ESM2M (with orographic factor) and their uncertainty analysis

Future scenarios were generated for RCP2.6 and RCP8.5 for the proposed downscaling model with the inclusion of orographic factor. Fig. 6.5 presents the projected future series after bias correction and uncertainties associated with the projections for ESM2G and ESM2M model, with inclusion of orographic factor. Trend analysis of future projected series of annual rainfall has been done for the new model, which is presented in section 6.3.4. Percentage uncertainties for all future projections have also been evaluated and discussed in section 6.3.5.



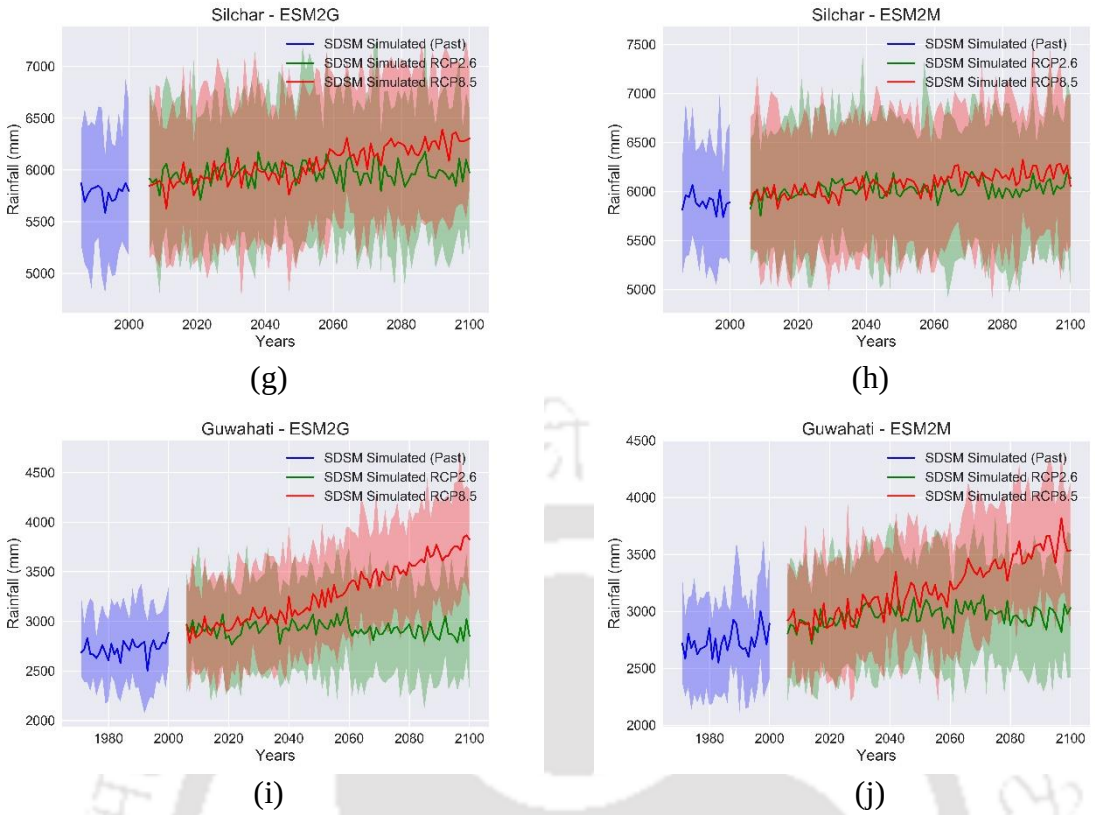


Figure 6. 5 Line plots with uncertainty of annual rainfall for observed (1971–2000) and projected period (2006–2100) for models ESM2G and ESM2M, with inclusion of orographic factor

6.3.4. Comparison of different model projections and trend analysis

In the previous sections, Fig.6.1 to Fig. 6.5 show the time series plots of future projections. For better understanding of the variations in the projected period, it was divided into three time slices: 2006-2040, 2041-2070 and 2071-2100. Mean values for each projected time slice is estimated and plotted for each station. Trend analysis provides better understanding of the significance of the variations. In the present study, Mann Kendall trend test has been used to perform trend analysis for all three parameters at every station, for both observed and projected period.

Fig. 6.6 shows the variations of average daily maximum temperature (T_{max}) for the models ESM2G (Fig. 6.6(a), (b)) and ESM2M (Fig. 6.6(c), (d)) under RCP2.6 and RCP8.5, for the three projected time slices. Increase in maximum temperature has been found for all the stations in the projected period till the year 2100, as compared to the observed period with a highest rise of about 0.5 degree C for ESM2G and 0.49 degree C for ESM2M. The MK trend test (Table 6.1) suggests that all stations have non-significant increasing trend (decreasing for Cherrapunjee) under RCP2.6 for model ESM2G, whereas all stations have significant

increasing trend (non-significant for Cherrapunjee) under RCP2.6 for model ESM2M. The decreasing trend for Cherrapunjee in the projected period indicates that although the temperature magnitude has increased, the rise is high during 2041-2070 after which it declines in 2071-2100. However, all the stations have shown significant increasing trend under RCP8.5 for both the models. For HadCM3 A2 scenario, all stations have shown significant increasing trend. This is justifiable, as A2 scenario has similar trajectory to RCP8.5.

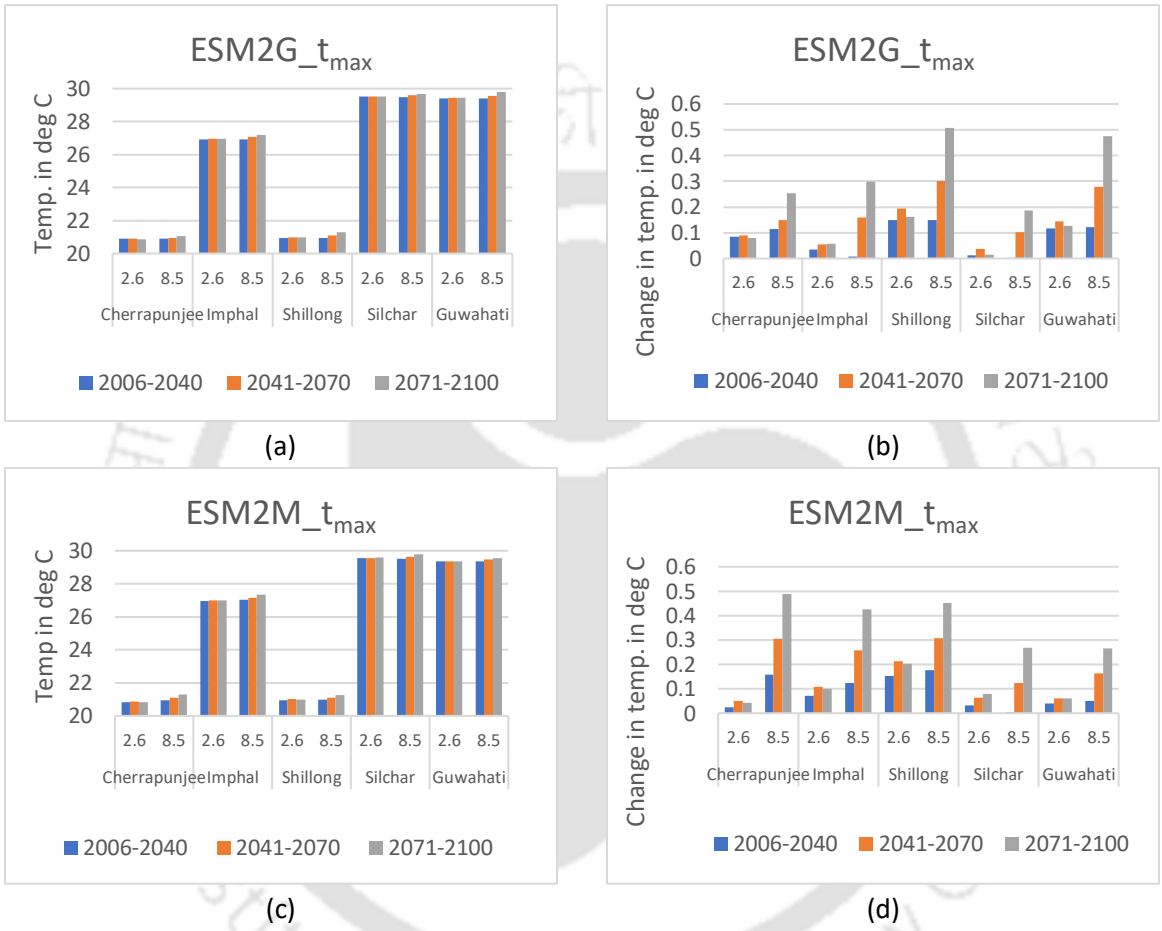


Figure 6. 6 Average daily maximum temperature and change in average daily maximum temperature w.r.t. observed mean in future projections (2006-2040), (2041-2070) and (2071-2100)

Table 6. 1 Mann Kendall trend test for average daily maximum temperature

Meteorological Parameters	Period	Model	Scenario	Cherrapunjee	Imphal	Shillong	Silchar	Guwahati
T_{max}	past	observed		++	++	++	++	+
	projected	RCP2.6	ESM2G	-	+	+	+	+
			ESM2M	+	++	++	++	++
		RCP8.5	ESM2G	++	++	++	++	++
			ESM2M	++	++	++	++	++
		HadCM3	A2	++	++	++	++	++

+ represents non- significant increasing trend (P > 0.05)
- represents non- significant decreasing trend (P > 0.05)
++ represents significant increasing trend (P < 0.05; MK Z-value > 1.96).
-- represents significant decreasing trend (P < 0.05; MK Z-value < -1.96).

Fig. 6.7 shows the variations of average daily minimum temperature (T_{min}) for the models ESM2G (Fig. 6.7(a), (b)) and ESM2M (Fig. 6.7(c), (d)) under RCP2.6 and RCP8.5, for the three projected time slices. Increase in minimum temperature has been found for all the stations in the projected period till the year 2100, as compared to the observed period with a highest rise of about 0.82 degree C for ESM2G and 0.83 degree C for ESM2M. However mixed trends have been found for different stations in the projected period till the year 2100. The MK trend test (Table 6.2) suggests that under RCP2.6 scenario, Cherrapunjee will have non-significant increasing trend in minimum temperature, when projected for ESM2G model, whereas ESM2M model suggests significant increasing trend under the same scenario. For stations Imphal and Silchar, ESM2G model suggests significant increasing trend ESM2M gives non-significant increasing trend, under RCP2.6. In case of Shillong and Guwahati, non-significant decreasing trends have been found for ESM2G model, whereas significant increasing trends have been found for ESM2M under RCP2.6. However, all the stations have shown significant increasing trend under RCP8.5 or both the models as well as for HadCM3 A2 scenario.

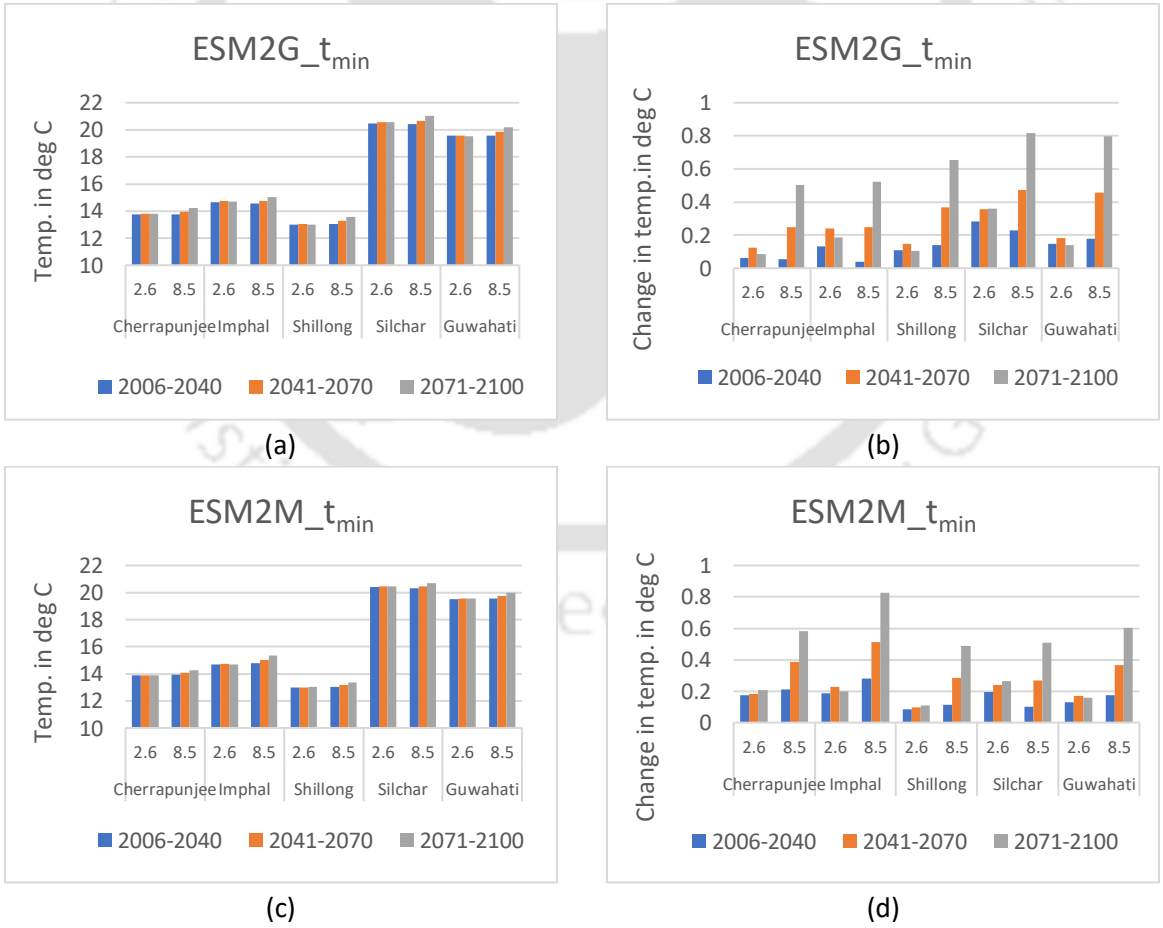


Figure 6. 7 Average daily minimum temperature and change in average daily minimum temperature w.r.t. observed mean in future projections (2006-2040), (2041-2070) and (2071-2100)

Table 6. 2 Mann Kendall trend test for average daily minimum temperature

Meteorological Parameters	Period	Model	Scenario	Cherrapunjee	Imphal	Shillong	Silchar	Guwahati
T_{min}	past	observed		-	+	++	+	++
	projected	RCP2.6	ESM2G	+	++	-	++	-
			ESM2M	++	+	++	+	++
		RCP8.5	ESM2G	++	++	++	++	++
			ESM2M	++	++	++	++	++
		HadCM3	A2	++	++	++	++	++

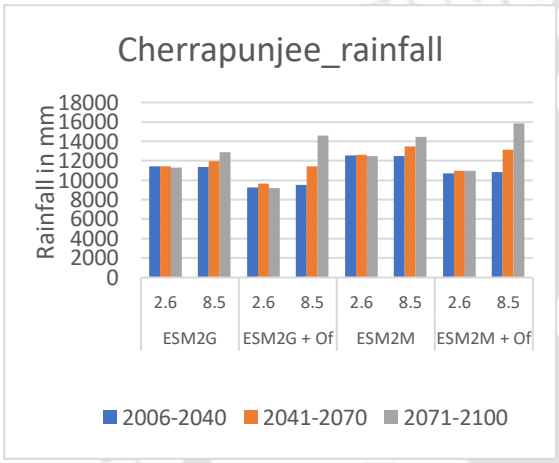
+ represents non- significant increasing trend (P > 0.05)
- represents non- significant decreasing trend (P > 0.05)
++ represents significant increasing trend (P < 0.05; MK Z-value > 1.96).
-- represents significant decreasing trend (P < 0.05; MK Z-value < -1.96).

Fig. 6.8 shows the variations of average annual rainfall under RCP2.6 and RCP8.5, for the three projected time slices. The variations have been estimated for both ESM2G and ESM2M as well the new models with inclusion of the orographic factor (ESM2G+O_f and ESM2M+O_f). The new models have shown significant variations in rainfall magnitudes with the previous models for Silchar and Guwahati. However, for other stations the differences are not significant. Percentage increase of rainfall by the end of 3rd time slice (2071-2100) has been evaluated for previous (ESM2G and ESM2M) and the new models (ESM2G+O_f and ESM2M+O_f). Results show that, the percentage has increased from 22.9% to 33.6% for Cherrapunjee when new models were used for future projection. Similarly, results have been found for Imphal (35.5% to 57%) and Shillong (18.6% to 60.1%). For Silchar the percentage increase has changed from 4.3% to 88.8%, which is quite high. For Guwahati, the percentage increase has changed from 32.4% to almost twice the amount for average rainfall during historical period of record.

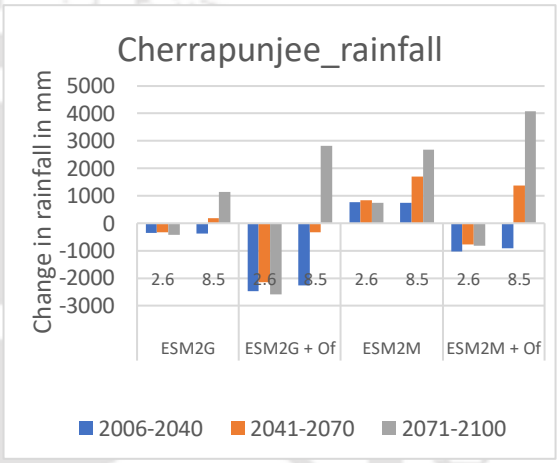
The interesting finding here is that, significant changes have been found for those stations (Silchar and Guwahati) for which R^2 values have been improved during validation (refer Chapter 5) due to inclusion of the orographic factor.

The MK trend test (Table 6.3) suggests non-significant increasing or decreasing trend under scenario RCP2.6. While some stations (Imphal and Guwahati) have shown similar trend for both ESM2G and ESM2M, other stations have shown contrasting trends (Cherrapunjee, Shillong and Silchar). Under scenario RCP8.5, all the stations have shown significant increasing trend, except Silchar for ESM2M model projection. HadCM3 A2 scenario, however, has suggested significant increasing trend for all the stations, except Cherrapunjee with non-significant decreasing trend.

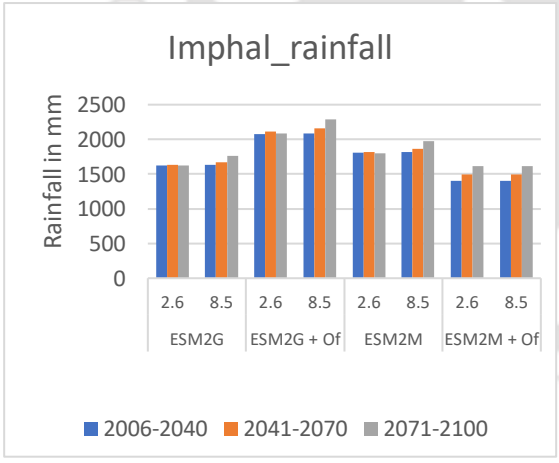
In case of new models with orographic factor, it has been realized that the trend analysis results are consistent with the previous models (without inclusion of the orographic factor) for stations Cherrapunjee, Imphal and Shillong. Whereas for stations Silchar and Guwahati, some inconsistency can be seen. While Silchar has shown inconsistency in ESM2M model projections for both RCP2.6 and RCP8.5, Guwahati has revealed contradictory results for ESM2G model under RCP2.6. However, the dissimilarities are non-significant for RCP2.6 scenario.



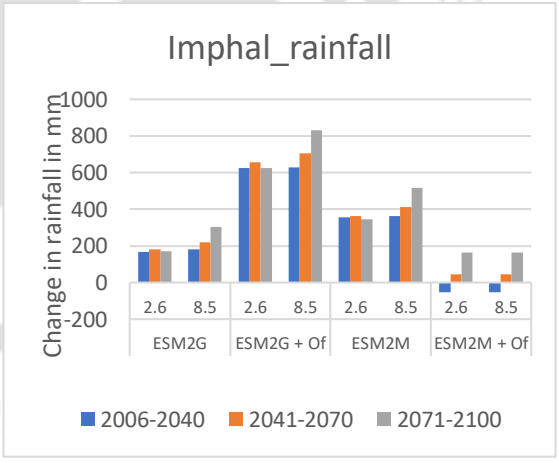
(a)



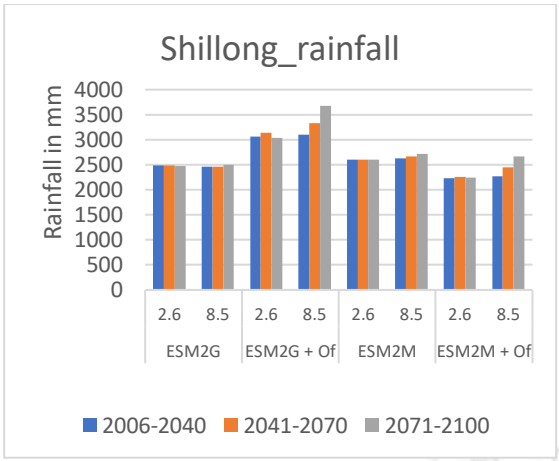
(b)



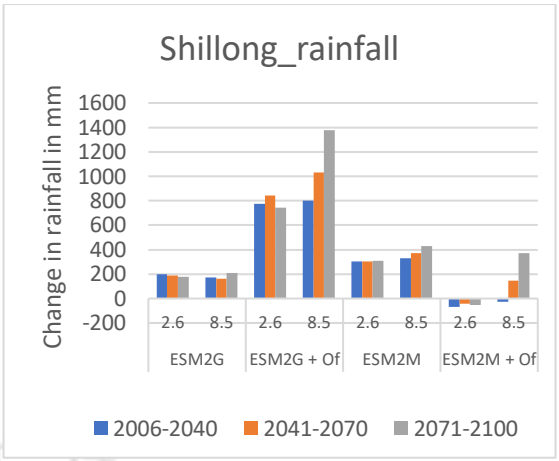
(c)



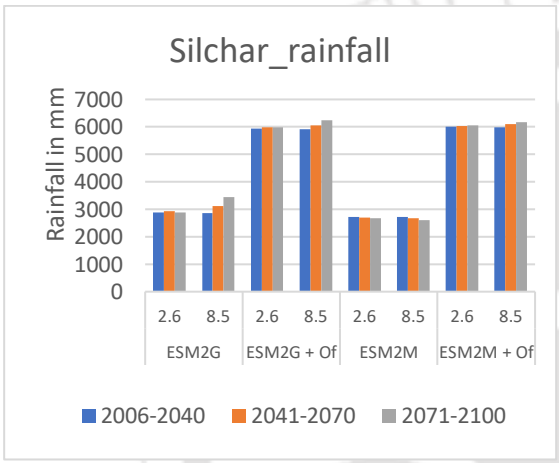
(d)



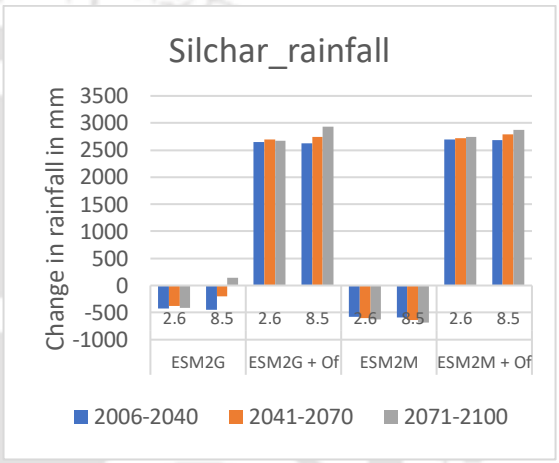
(e)



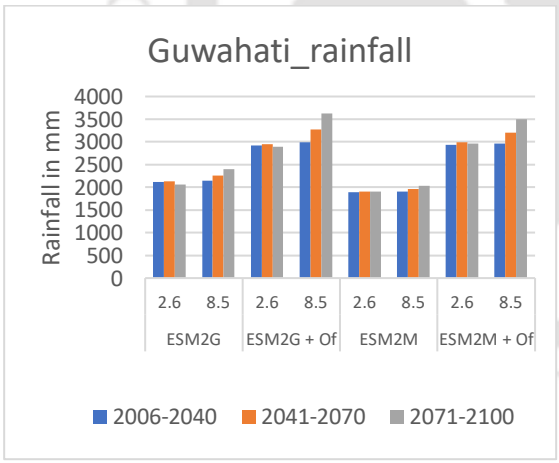
(f)



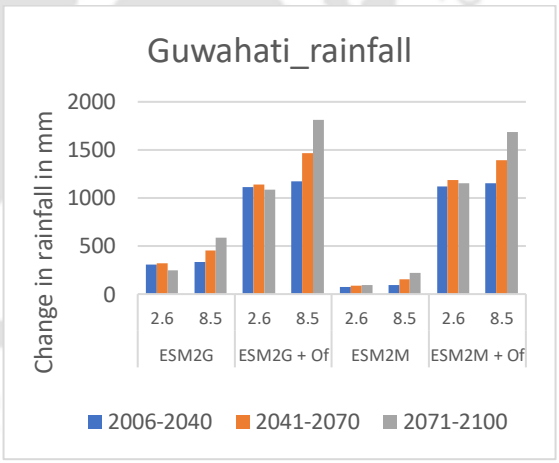
(g)



(h)



(i)



(j)

Figure 6. 8 Average annual rainfall and change in average annual rainfall w.r.t. observed mean in future projections (2006-2040), (2041-2070) and (2071-2100)

Table 6. 3 Mann Kendall trend test for average annual rainfall

Meteorological Parameters	Period	Model	Scenario	Cherrapunjee	Imphal	Shillong	Silchar	Guwahati
Rainfall	past	observed		+	+	++	+	+
	projected	RCP2.6	ESM2G	-	+	-	+	+
			ESM2G+O_f	-	+	-	+	-

			ESM2M	+	+	+	-	+
			ESM2M+O_f	++	+	+	++	++
		RCP8.5	ESM2G	++	++	++	++	++
			ESM2G+O_f	++	++	++	++	++
			ESM2M	++	++	++	--	++
			ESM2M+O_f	++	++	++	++	++
		HadCM3	A2	-	++	++	++	++

+ represents non- significant increasing trend (P > 0.05)
- represents non- significant decreasing trend (P > 0.05)
++ represents significant increasing trend (P < 0.05; MK Z-value > 1.96).
-- represents significant decreasing trend (P < 0.05; MK Z-value < -1.96).

6.3.5. Uncertainty analysis of the downscaling process

In this study, intra-model uncertainty has been assessed for all three parameters to understand the underlying uncertainties of the downscaling process. Fig. 6.2 to Fig. 6.4 present the line plot of the observed as well as projected future series after bias correction and uncertainties associated with the projections for ESM2G and ESM2M model for both RCPs. The size of the uncertainty band indicates the maximum and minimum value of all possible ensemble series (synthetic daily weather series) generated during simulation and prediction. In other words, it displays the extent to which these ensembles vary from each other.

For the unconditional parameters i.e. average daily maximum temperature (Fig. 6.2) and average daily minimum temperature (Fig. 6.3), variations are less. Whereas for annual average rainfall, large variations among individual ensembles have been observed (Fig. 6.4). To quantify the variations, percentage uncertainty based on maximum-minimum range has been calculated. The percentage uncertainties for average daily maximum temperature, average daily minimum temperature and average annual rainfall were tabulated below in table 6.4, table 6.5 and table 6.6 respectively. It was observed that the percentage uncertainty for maximum temperature ranges from 0.9% to 1.2%, which can be considered as insignificant. Similarly, for minimum temperature, the range is 1.0% - 1.9%. This justifies that, local scale orographic arrangement is not required to be used for simulation as well as for future prediction of temperature.

Table 6. 4 Percentage uncertainty in average daily maximum temperature (T_{max})

Meteorological Parameters	Period	Model	Scenario	Cherrapunjee	Imphal	Shillong	Silchar	Guwahati
T_{max}	projected	RCP2.6	ESM2G	**1.2**	**1.0**	**1.2**	**1.0**	**0.9**
			ESM2M	**1.1**	**1.0**	**1.1**	**1.0**	**0.9**
		RCP8.5	ESM2G	**1.1**	**1.0**	**1.1**	**1.0**	**0.9**
			ESM2M	**1.2**	**1.0**	**1.1**	**1.0**	**0.9**

Table 6. 5 Percentage uncertainty in average daily minimum temperature (T_{min})

Meteorological Parameters	Period	Model	Scenario	Cherrapunjee	Imphal	Shillong	Silchar	Guwahati
T_{min}	projected	RCP2.6	ESM2G	**1.6**	**1.9**	**1.5**	**1.2**	**1.1**
			ESM2M	**1.7**	**1.9**	**1.6**	**1.1**	**1.1**
		RCP8.5	ESM2G	**1.6**	**1.8**	**1.5**	**1.2**	**1.0**
			ESM2M	**1.6**	**1.9**	**1.5**	**1.2**	**1.0**

On the other hand, the percentage uncertainties for average annual rainfall, without considering the orographic factor, ranges from 20.7% - 28.3%, which is quite high (Table 6.6). This uncertainty can be considered as significant as it may deviate the results to a large extent. Particularly, if a researcher uses a single ensemble series for future prediction and further studies are done based on those results, it may produce huge differences with another study based on the same simulation and prediction results but with a different ensemble series. To understand the influence of the orographic factor in intra-model uncertainty, percentage uncertainties have also been calculated for the new models (Table 6.6). Fig. 6.5 present the line plot of the observed as well as projected future series after bias correction and uncertainties associated with the projections for ESM2G and ESM2M models with orographic factor. It has been observed that inclusion of the orographic factor has reduced the percentage uncertainty to a decent extent in most of the cases. The reduction of percentage uncertainty ranges from 0.2% (insignificant) to as high as 11% (significant). Although in some cases (highlighted in red for stations Cherrapunjee and Imphal) percentage uncertainty has shown an increment, but the range is within 1.1% to 1.9%, which is insignificant. Here also, interestingly for stations at an identifiable distance from a hill (Guwahati, Silchar), the orographic factor exhibited significant reduction in the percentage uncertainty. These results are consistent with the previous results of Chapter 5, where R^2 values were improved by the orographic factor for those stations.

Table 6. 6 Percentage uncertainty in average annual rainfall

Meteorological Parameters	Period	Model	Scenario	Cherrapunjee	Imphal	Shillong	Silchar	Guwahati
Rainfall	projected	RCP2.6	ESM2G	**24.2**	**24.4**	**22.1**	**21.4**	**27.6**
			ESM2G+O_f	**25.3**	**19.4**	**18.7**	**12.5**	**18.0**
			ESM2M	**28.3**	**22.6**	**23.2**	**23.4**	**27.1**
			ESM2M+O_f	**23.5**	**23.9**	**23.0**	**12.4**	**18.0**
		RCP8.5	ESM2G	**23.4**	**24.0**	**22.1**	**20.7**	**26.8**
			ESM2G+O_f	**22.2**	**18.3**	**17.1**	**12.5**	**16.6**
			ESM2M	**26.4**	**22.0**	**23.1**	**23.2**	**27.0**
			ESM2M+O_f	**20.8**	**23.9**	**21.8**	**12.4**	**16.8**

6.4. SUMMARY

Following conclusions could be drawn from the analysis:

- It is evident from the results that there is increase in both rainfall and temperature in most of the cases for RCP8.5. In case of RCP2.6, much increment is not exhibited by both the models.
- However, there are some variations between the model projections for ESM2G and ESM2M.
- Overall, the projections for rainfall have higher uncertainty compared to temperature. This can be attributed to the complexity of rainfall process.
- The increase in projected annual rainfall is found to be more in case of new projections where orographic factor is incorporated.
- Inclusion of orographic factor has been able to reduce the uncertainty of the downscaling process to a decent extent.

Chapter 7

SUMMARY AND CONCLUSIONS

The primary goal of the study was to determine homogeneous hydroclimatic regions in northeast India and to analyse impacts of climate change on temperature and rainfall projections, with inclusion of orographic influence. As a first step, the region was delineated into hydrologically homogeneous clusters using fuzzy clustering approach. Observed rainfall data were collected from IMD for 10 IMD stations situated in the north-eastern region of India. Also, gridded dataset of size $0.5^{\circ} \times 0.5^{\circ}$ were collected from the department. GCM datasets were collected from IPCC AR4 and AR5. Various approaches were tried to find out the best possible clustering of the region. Following are the major outcomes of the analysis:

- In most of the cases, all the stations are found to be in the same cluster while using nearest stations' rainfall data as input. Hence, it contradicts with the whole purpose of regionalization.
- Difference in rainfall magnitudes among the stations affects the output clustering pattern. Hence, rainfall data alone is not effective for clustering.
- Similarly, usage of LSAVs (from GCM data) alone is also not effective because of its coarse resolution.
- Inclusion of geographical location parameters (latitude, longitude and elevation) effectively reduces the above limitations, enhancing the chance of getting satisfactory clustering results.
- A hybrid approach with GCM grid data (interpolated to station) along with rainfall data and location parameters (latitude, longitude and elevation) was tried for clustering and gave the best result among all other approaches**.**
- The hybrid approach exhibited three clusters in the northeastern region
    - **Cluster 1:** Region covering the stations Guwahati, Dhubri and Tezpur
    - **Cluster 2**: Region covering the stations Passighat, North-Lakhimpur and Mohanbari
    - **Cluster 3:** Region covering the stations Shillong, Silchar, Imphal and Cherrapunjee

However, there are some limitations to large scale clustering, such as use of limited number of rainfall stations for analysis, and course resolution of GCM grid data. To address these issues, sub-clustering of one of the clusters formed was done by utilizing local rainfall data (collected from 83 stations situated in various tea gardens of Assam), along with location parameters. Again, various combinations were tried. Following are the outcomes of the analysis:

- The best result was exhibited by the approach, in which total annual rainfall, standard deviation of total annual rainfall and location parameters (latitude, longitude and elevation) were used as attributes.
- Hence, this approach can be used for regionalization of small-scale areas.
- Two sub-clusters could be found in the cluster 2 by using this approach
    - **Cluster 2a:** region on the west side of cluster 2
    - **Cluster 2b:** region on the east side of cluster 2.

In the next step, five stations were selected to develop statistical downscaling models for predicting temperature and rainfall. Main objective was to understand the applicability of a specific downscaling model within a cluster. Historical data trends were analysed for rainfall and temperature for all five stations. Three different GCM models were used for the study: (i) AR4: HadCM3 with scenario A2 and (ii) AR5: GFDL-ESM2G and GFDL-ESM2G for RCP2.6 and RCP8.5. Various combinations of calibration and validation datasets were examined for HadCM3 model due to limitation in the observed dataset and the best combination was used for further analysis. For ESM2G and ESM2G models three different sub-models of SDSM (SDSM-A, SDSM-M and SDSM-S) were examined. The models developed were then validated to assess their applicability. Following observations were made from the analysis:

- In case of rainfall, having a good correlation (more than 0.8) using daily rainfall data is difficult. Hence, there is a need to develop better correlated models to reduce the uncertainties.
- Although the same GCM variables were used for homogeneous clustering as well for downscaling, the correlation coefficients do not show any consistency within a cluster.
- The regression models are different for different stations within a cluster, although the areal distance between those stations was not very large.
- This indicates that orography plays a very important role in rainfall variations at a local scale.

- In case of daily maximum and minimum temperature, local scale orographic arrangement is not showing significant influence and therefore orographic influence may not be used in developing downscaling model for temperature in the studied region.

To understand the effect of orography, topographical characteristics of the region were thoroughly studied. SRTM DEM data were collected from NASA servers, wind datasets were collected from IMD, more rainfall data were collected from various open sources. Terrain analysis of the region exhibited that the area covering the five stations has high orographic variations along with topographical and meteorological variations. Further, wind datasets were analysed to understand the wind flow variations and the major directions of wind flow existing in the five stations.

Based on the analysis, a new downscaling approach was attempted with inclusion of an orographic factor, formulated by using location and height of hills near the stations, distance of a station from hill, elevation of the stations, wind flow direction and seasonality factor. The variable was then included in the downscaling models and its efficiency was assessed. Following results were found from the analysis:

- Inclusion of orographic factor does impart some changes in the R^2 value.
- For stations at an identifiable distance from a hill (Guwahati, Silchar) the orographic factor has exhibited increase in the R^2 value to a decent extent. It appears that moisture bearing monsoonal wind while crossing the khasi-Jayantia hill range moves high and get condensed. As seen in the figure 5.9 this moisture bearing wind after crossing the hills influence the rain of Guwahati station. A part of this moisture bearing rain moves in the right direction towards Silchar and influences the precipitation there.
- However, the difference in R^2 values before addition of orographic factor and after addition is not very high.
- For stations located on a hill, the orographic factor could not improve the R^2 value.

Finally, uncertainty analysis was done and the future rainfall series were projected for respective model. Following results were found from the analysis:

- Both rainfall and temperature exhibited increasing trend in most of the cases for RCP8.5. In case of RCP2.6, much increment is not exhibited.
- The increase in projected annual rainfall is found to be more in case of new projections where orographic factor is incorporated.
- Overall, the projections for rainfall have higher uncertainty compared to temperature, which can be attributed to the complexity of rainfall process.
- Finally, from the analysis it was found that, inclusion of orographic factor has been able to reduce the uncertainty of the downscaling process to a decent extent.

6.5. FUTURE SCOPE OF THE WORK

Based on the limitations of this study, this section provides some possible future directions for research:

- Homogeneous clustering of the region was done based on one meteorological parameter (rainfall). Further study can be done by using multiple parameters to understand the clustering pattern for different parameters and further analysis on climate change impact may be done based on those results.
- This study used limited observed dataset for rainfall and temperature due to lack of reliable data resources. Future study may be done with the use of larger dataset which may improve the results.
- More specific approaches are required to strengthen the orographic factor to get better results. Hence, daily variation of wind flow direction may be included in formulating the orographic factor, instead of annual or seasonal wind direction.
- Further study may also be done by incorporating more orographic variables and with emphasis on monthly/daily variations of those variables.
- In this study our main objective was to assess the impact of orographic factor on intra-model uncertainty. Hence, inter-model uncertainly was not assessed. Further study could be done by incorporating multiple models with inclusion of orographic factor to examine inter-model uncertainty.

ANNEXURE A

LIST OF PREDICTORS

1. *List of HadCM3 predictors*

Sl. NO.	*ACRONYMS*	*PREDICTORS*
1	hur200	Relative Humidity @ 200hpa
2	hur500	Relative Humidity @ 500hpa
3	hur850	Relative Humidity @ 850hpa
4	mrso	Total Soil Moisture Content
5	pr	Total Precipitation
6	prc	Convective Precipitation
7	psl	Sea Level Pressure
8	rsds	Surface Downwelling Shortwave Radiation
9	snm	Surface Snow Melt
10	ta200	Air Temperature @ 200hpa
11	ta500	Air Temperature @ 500hpa
12	ta850	Air Temperature @ 850hpa
13	tas	Surface Air Temperature
14	ts	Surface Temperature
15	ua200	Zonal Wind Speed @ 200hpa
16	ua500	Zonal Wind Speed @ 500hpa
17	ua850	Zonal Wind Speed @ 850hpa
18	uas	Zonal Surface Wind Speed
19	va200	Meridional Wind Speed@200hpa
20	va500	Meridional Wind Speed@500hpa
21	va850	Meridional Wind Speed@850hpa
22	vas	Meridional Surface Wind Speed
23	zg200	Geopotential Height @ 200hpa
24	Zg500	Geopotential Height @ 500hpa
25	Zg850	Geopotential Height @ 850hpa

2. *List of ESM2G and ESM2M predictors*

Sl. NO.	*ACRONYMS*	*PREDICTORS*
1	clt	Total Cloud Cover
2	hfls	Surface Upward Latent Heat Flux
3	hfss	Surface Upward Sensible Heat Flux
4	huss	Near Surface Specific Humidity
5	pr	Precipitation
6	prc	Convective Precipitation
7	psl	Sea Level Pressure
8	rhs	Near Surface Relative Humidity
9	rhs_max	Maximum RHS
10	rhs_min	Minimum RHS
11	rlds	Surface Downwelling Longwave Radiation
12	rlus	Surface Upwelling Longwave Radiation
13	rlut	Total Outgoing Longwave Radiation
14	rsds	Surface Downwelling Shortwave Radiation
15	rsus	Surface Upwelling Shortwave Radiation
16	sfc wind	Daily Mean Near Surface Wind Speed
17	sfc wind_max	Daily Maximum Near Surface Wind Speed
18	tas	Near Surface Air Temperature
19	tas_max	Maximum TAS
20	tas_min	Minimum TAS
21	uas	Eastward Near Surface Temperature
22	vas	Westward Near Surface Temperature

ANNEXURE B

VALIDITY INDICES RESULTS FOR UPPER BRAHMAPUTRA VALLEY REGION

1. Values of the validity indices with total monthly rainfall as attributes

fuzzifier(m)	Cluster no.	V_{PC}	V_{PE}	V_{FPI}	V_{NCE}	V_{XB}	V_{FS}	V_K
	no. of cluster=2	0.98	0.05	0.04	0.05	0.34	161.43	28.68
	no. of cluster=3	0.98	0.06	0.04	0.04	0.62	-37.80	52.35
	no. of cluster=4	0.95	0.12	0.06	0.06	0.77	-107.16	65.07
	no. of cluster=5	0.97	0.08	0.04	0.04	0.67	-217.12	57.88
1.1	no. of cluster=6	0.97	0.09	0.04	0.04	1.31	-264.16	113.66
	no. of cluster=7	0.97	0.10	0.04	0.03	1.39	-305.09	119.63
	no. of cluster=8	0.97	0.09	0.04	0.03	1.29	-355.93	111.92
	no. of cluster=9	0.96	0.12	0.05	0.04	2.38	-356.80	204.41
	ıo. of cluster=10	0.98	0.06	0.02	0.02	1.07	-457.16	94.60
	no. of cluster=2	0.95	0.12	0.09	0.12	0.34	160.52	28.29
	no. of cluster=3	0.93	0.19	0.11	0.12	0.60	-29.63	50.82
	no. of cluster=4	0.92	0.21	0.11	0.10	0.67	-139.08	57.03
	no. of cluster=5	0.89	0.29	0.14	0.13	0.78	-197.88	66.44
1.2	no. of cluster=6	0.92	0.22	0.10	0.09	0.59	-287.84	52.76
	no. of cluster=7	0.86	0.38	0.16	0.14	1.40	-288.23	121.43
	no. of cluster=8	0.88	0.35	0.14	0.12	1.31	-344.05	113.88
	no. of cluster=9	0.87	0.35	0.14	0.11	1.23	-382.97	107.62
	ıo. of cluster=10	0.87	0.39	0.15	0.12	1.55	-401.27	135.11
	no. of cluster=2	0.92	0.19	0.15	0.19	0.33	161.26	27.87
	no. of cluster=3	0.87	0.33	0.20	0.21	0.58	-16.83	49.00
	no. of cluster=4	0.86	0.37	0.19	0.19	0.63	-124.31	53.49
	no. of cluster=5	0.81	0.51	0.24	0.22	0.78	-179.91	67.26
1.3	no. of cluster=6	0.78	0.61	0.27	0.24	1.46	-221.77	126.85
	no. of cluster=7	0.77	0.64	0.26	0.23	1.14	-245.23	98.80
	no. of cluster=8	0.75	0.71	0.28	0.24	1.38	-289.98	119.86
	no. of cluster=9	0.72	0.82	0.31	0.26	1.29	-295.11	112.33
	ıo. of cluster=10	0.74	0.76	0.29	0.23	1.53	-356.73	134.40
	no. of cluster=2	0.89	0.26	0.22	0.26	0.33	163.09	27.44
	no. of cluster=3	0.81	0.48	0.28	0.31	0.56	-2.52	47.56
	no. of cluster=4	0.79	0.57	0.28	0.28	0.60	-104.16	51.04
	no. of cluster=5	0.73	0.73	0.34	0.31	0.78	-157.82	66.70
1.4	no. of cluster=6	0.66	0.93	0.41	0.36	1.67	-186.74	144.61
	no. of cluster=7	0.65	0.98	0.40	0.35	1.54	-225.64	133.63
	no. of cluster=8	0.62	1.12	0.43	0.37	1.41	-240.07	122.59
	no. of cluster=9	0.60	1.19	0.45	0.38	1.73	-256.52	150.87
	ıo. of cluster=10	0.59	1.29	0.46	0.39	1.59	-264.43	138.14

	no. of cluster=2	0.86	0.34	0.28	0.34	0.32	165.32	26.98
	no. of cluster=3	0.75	0.63	0.37	0.40	0.55	12.08	46.49
	no. of cluster=4	0.64	0.93	0.48	0.47	0.98	-41.43	82.78
	no. of cluster=5	0.65	0.96	0.43	0.41	0.76	-129.61	64.98
1.5	no. of cluster=6	0.60	1.15	0.49	0.45	1.01	-152.42	87.44
	no. of cluster=7	0.56	1.26	0.51	0.45	1.66	-187.26	144.59
	no. of cluster=8	0.55	1.34	0.51	0.45	1.50	-207.78	130.91
	no. of cluster=9	0.51	1.50	0.55	0.47	1.95	-215.54	170.24
	ıo. of cluster=10	0.49	1.61	0.56	0.48	1.70	-225.43	149.33
	no. of cluster=2	0.83	0.41	0.35	0.41	0.32	167.47	26.49
	no. of cluster=3	0.70	0.77	0.45	0.49	0.54	26.16	45.71
	no. of cluster=4	0.58	1.11	0.57	0.55	1.00	-19.35	85.18
	no. of cluster=5	0.57	1.20	0.54	0.52	0.73	-92.40	62.57
1.6	no. of cluster=6	0.50	1.41	0.59	0.54	2.09	-121.53	180.72
	no. of cluster=7	0.48	1.52	0.60	0.54	1.77	-143.59	153.90
	no. of cluster=8	0.45	1.68	0.63	0.56	1.48	-160.07	129.08
	no. of cluster=9	0.43	1.77	0.64	0.56	2.08	-175.24	183.24
	ıo. of cluster=10	0.42	1.88	0.65	0.57	3.11	-182.76	273.71
	no. of cluster=2	0.80	0.48	0.41	0.48	0.31	169.21	25.96
	no. of cluster=3	0.64	0.89	0.53	0.56	0.53	38.87	45.07
	no. of cluster=4	0.52	1.25	0.64	0.63	1.04	-0.43	87.90
	no. of cluster=5	0.47	1.46	0.66	0.63	1.06	-36.66	90.33
1.7	no. of cluster=6	0.44	1.61	0.67	0.62	2.33	-89.40	201.57
	no. of cluster=7	0.42	1.75	0.68	0.62	1.85	-106.26	161.01
	no. of cluster=8	0.40	1.85	0.68	0.62	1.59	-130.87	139.67
	no. of cluster=9	0.37	2.00	0.71	0.63	2.33	-138.42	206.04
	ıo. of cluster=10	0.34	2.16	0.73	0.65	6.37	-139.05	563.95
	no. of cluster=2	0.77	0.54	0.47	0.54	0.30	170.35	25.40
	no. of cluster=3	0.60	1.00	0.60	0.63	0.53	49.56	44.42
	no. of cluster=4	0.48	1.37	0.70	0.69	1.08	15.10	91.86
	no. of cluster=5	0.42	1.61	0.72	0.69	1.24	-14.73	105.78
1.8	no. of cluster=6	0.39	1.79	0.74	0.69	2.81	-59.56	243.51
	no. of cluster=7	0.36	1.94	0.74	0.69	1.95	-74.90	170.45
	no. of cluster=8	0.33	2.11	0.77	0.70	1.53	-83.57	134.61
	no. of cluster=9	0.32	2.20	0.76	0.69	3.02	-107.76	268.84
	ıo. of cluster=10	0.30	2.35	0.78	0.71	1946.01	-110.53	173867.57
	no. of cluster=2	0.74	0.60	0.52	0.60	0.30	170.73	24.81
	no. of cluster=3	0.56	1.09	0.66	0.69	0.52	57.97	43.77
	no. of cluster=4	0.44	1.47	0.75	0.74	1.14	26.99	96.89
1.9	no. of cluster=5	0.38	1.73	0.77	0.74	1.44	1.36	123.59
	no. of cluster=6	0.33	1.97	0.80	0.76	6.12	-23.94	530.27
	no. of cluster=7	0.32	2.09	0.79	0.75	2.23	-48.35	195.81
	no. of cluster=8	0.29	2.27	0.82	0.76	1.73	-55.20	153.25
	no. of cluster=9	0.28	2.37	0.81	0.75	11.68	-81.27	1050.73

	ıo. of cluster=10	0.26	2.50	0.82	0.75	2431.17	-87.65	219845.24
	no. of cluster=2	0.71	0.65	0.57	0.65	0.29	170.29	24.19
	no. of cluster=3	0.53	1.16	0.71	0.73	0.51	64.04	42.96
	no. of cluster=4	0.41	1.56	0.79	0.78	1.21	35.44	103.10
	no. of cluster=5	0.35	1.82	0.81	0.78	1.70	12.70	146.56
2	no. of cluster=6	0.29	2.09	0.85	0.81	3.18	4.01	275.33
	no. of cluster=7	0.28	2.23	0.84	0.79	3.77	-25.12	332.04
	no. of cluster=8	0.25	2.40	0.85	0.80	2.09	-37.44	186.55
	no. of cluster=9	0.23	2.55	0.86	0.80	656.11	-41.08	59120.64
	ıo. of cluster=10	0.22	2.69	0.87	0.81	2266.69	-45.07	205681.05
	no. of cluster=2	0.69	0.69	0.62	0.69	0.28	169.02	23.55
	no. of cluster=3	0.50	1.22	0.75	0.77	0.50	67.93	41.99
	no. of cluster=4	0.38	1.62	0.82	0.81	1.31	41.00	111.71
	no. of cluster=5	0.32	1.90	0.84	0.82	2.09	20.46	179.94
2.1	no. of cluster=6	0.27	2.17	0.87	0.84	2.40	11.76	209.09
	no. of cluster=7	0.24	2.37	0.88	0.84	927.77	5.54	81409.91
	no. of cluster=8	0.22	2.53	0.89	0.84	16.52	-15.00	1476.02
	no. of cluster=9	0.21	2.65	0.89	0.83	151604.79	-30.00	13841433.58
	ıo. of cluster=10	0.20	2.79	0.89	0.84	5444.82	-34.01	501748.26
	no. of cluster=2	0.67	0.73	0.66	0.73	0.27	166.95	22.89
	no. of cluster=3	0.48	1.27	0.78	0.80	0.48	69.92	40.80
	no. of cluster=4	0.36	1.68	0.85	0.84	1.43	44.07	122.27
	no. of cluster=5	0.30	1.97	0.87	0.85	2.74	25.49	236.88
2.2	no. of cluster=6	0.25	2.23	0.90	0.86	2.42	16.62	211.78
	no. of cluster=7	0.23	2.43	0.90	0.87	709.98	9.89	62753.32
	no. of cluster=8	0.20	2.63	0.92	0.88	769.40	7.02	68586.17
	no. of cluster=9	0.18	2.79	0.92	0.88	861.21	4.57	77858.67
	ıo. of cluster=10	0.17	2.90	0.92	0.87	681.94	-11.95	63245.63
	no. of cluster=2	0.65	0.76	0.70	0.76	0.26	164.17	22.21
	no. of cluster=3	0.46	1.31	0.81	0.83	0.47	70.40	39.63
	no. of cluster=4	0.35	1.72	0.87	0.86	1.61	45.25	137.68
	no. of cluster=5	0.29	2.02	0.89	0.87	4.40	28.44	382.21
2.3	no. of cluster=6	0.24	2.28	0.91	0.88	2.23	18.64	196.23
	no. of cluster=7	0.21	2.49	0.92	0.89	496.61	12.30	44285.35
	no. of cluster=8	0.19	2.68	0.93	0.89	701.94	8.77	63298.87
	no. of cluster=9	0.17	2.84	0.94	0.90	1176.78	6.56	107831.69
	ıo. of cluster=10	0.16	2.98	0.94	0.90	3308.85	2.86	307424.63
	no. of cluster=2	0.64	0.79	0.73	0.79	0.26	160.71	21.52
	no. of cluster=3	0.44	1.35	0.83	0.85	0.45	69.59	38.19
	no. of cluster=4	0.33	1.76	0.89	0.88	1.82	44.89	156.54
2.4	no. of cluster=5	0.27	2.07	0.91	0.89	58.16	29.90	5073.55
	no. of cluster=6	0.23	2.33	0.93	0.90	2.72	19.83	241.19
	no. of cluster=7	0.20	2.53	0.93	0.90	285.33	13.84	25700.40
	no. of cluster=8	0.18	2.73	0.94	0.91	218.94	9.73	20014.81

	no. of cluster=9	0.16	2.89	0.95	0.91	1053.64	7.46	98262.53
	ıo. of cluster=10	0.15	3.03	0.95	0.91	1839.17	4.33	174471.09
	no. of cluster=2	0.62	0.81	0.76	0.81	0.25	156.72	20.82
	no. of cluster=3	0.43	1.38	0.86	0.87	0.43	67.86	36.80
	no. of cluster=4	0.32	1.79	0.91	0.90	2.28	43.59	196.23
	no. of cluster=5	0.26	2.10	0.92	0.90	76.62	28.90	6717.17
2.5	no. of cluster=6	0.22	2.36	0.94	0.91	4.06	20.12	362.44
	no. of cluster=7	0.19	2.57	0.94	0.92	410.50	14.34	37390.13
	no. of cluster=8	0.17	2.77	0.95	0.92	1040.04	10.76	96508.97
	no. of cluster=9	0.15	2.93	0.96	0.92	1705.10	8.55	162179.86
	ıo. of cluster=10	0.14	3.08	0.96	0.93	7253.88	5.78	704597.97
	no. of cluster=2	0.61	0.83	0.78	0.83	0.24	152.28	20.12
	no. of cluster=3	0.42	1.40	0.87	0.88	0.42	65.44	35.52
	no. of cluster=4	0.31	1.82	0.92	0.91	3.31	41.58	286.00
	no. of cluster=5	0.25	2.13	0.93	0.92	319.96	27.28	28214.36
2.6	no. of cluster=6	0.21	2.39	0.95	0.93	28.02	19.60	2523.81
	no. of cluster=7	0.18	2.60	0.95	0.93	220.93	13.94	20378.92
	no. of cluster=8	0.16	2.80	0.96	0.93	69.20	10.67	6533.40
	no. of cluster=9	0.15	2.96	0.96	0.93	872.02	7.71	84605.67
	ıo. of cluster=10	0.13	3.11	0.96	0.94	16635.83	5.94	1664223.28
	no. of cluster=2	0.60	0.85	0.80	0.85	0.23	147.46	19.41
	no. of cluster=3	0.41	1.42	0.89	0.90	0.40	62.48	33.93
	no. of cluster=4	0.30	1.84	0.93	0.92	7.00	39.06	608.14
	no. of cluster=5	0.25	2.15	0.94	0.93	858.64	25.25	76223.41
2.7	no. of cluster=6	0.21	2.42	0.95	0.94	9.15	18.33	832.47
	no. of cluster=7	0.18	2.63	0.96	0.94	1092.12	12.99	102210.58
	no. of cluster=8	0.16	2.83	0.96	0.94	258.61	10.12	24886.97
	no. of cluster=9	0.14	2.99	0.97	0.94	752.95	7.60	74908.71
	ıo. of cluster=10	0.13	3.14	0.97	0.94	23365.97	6.39	2410798.87
	no. of cluster=2	0.59	0.86	0.82	0.86	0.22	142.37	18.71
	no. of cluster=3	0.40	1.44	0.90	0.91	0.38	59.23	32.71
	no. of cluster=4	0.30	1.86	0.94	0.93	27.34	36.10	2384.38
	no. of cluster=5	0.24	2.17	0.95	0.94	1047.80	23.05	93734.60
2.8	no. of cluster=6	0.20	2.44	0.96	0.94	389.07	16.73	35832.68
	no. of cluster=7	0.17	2.65	0.96	0.94	2793.91	11.84	265895.13
	no. of cluster=8	0.15	2.85	0.97	0.95	88.56	9.45	8708.43
	no. of cluster=9	0.14	3.01	0.97	0.95	4070.15	7.28	415822.43
	ıo. of cluster=10	0.12	3.17	0.98	0.95	11085.55	6.22	1180933.44
	no. of cluster=2	0.58	0.88	0.84	0.88	0.21	137.08	18.01
	no. of cluster=3	0.39	1.46	0.91	0.92	0.37	55.77	31.32
2.9	no. of cluster=4	0.29	1.87	0.94	0.94	122.43	32.93	10733.98
	no. of cluster=5	0.24	2.19	0.95	0.94	1137.23	21.61	102586.06
	no. of cluster=6	0.20	2.45	0.96	0.95	145.34	14.97	13571.59
	no. of cluster=7	0.17	2.67	0.97	0.95	985.74	10.51	95715.68

	no. of cluster=8	0.15	2.87	0.97	0.96	7558.00	8.36	762802.54
	no. of cluster=9	0.13	3.03	0.98	0.96	2478.66	6.51	261618.44
	no. of cluster=10	0.12	3.19	0.98	0.96	8700.59	5.48	964376.13
	no. of cluster=2	0.57	0.89	0.85	0.89	0.21	131.67	17.32
	no. of cluster=3	0.39	1.47	0.92	0.93	0.35	52.21	29.71
	no. of cluster=4	0.29	1.89	0.95	0.94	211.75	29.84	18673.61
	no. of cluster=5	0.23	2.20	0.96	0.95	939.73	18.65	85680.45
3	no. of cluster=6	0.19	2.47	0.97	0.95	2214.16	13.23	210105.33
	no. of cluster=7	0.17	2.69	0.97	0.96	2142.84	9.26	212922.13
	no. of cluster=8	0.15	2.88	0.98	0.96	951.21	7.31	98986.66
	no. of cluster=9	0.13	3.05	0.98	0.96	5369.26	5.73	590884.46
	no. of cluster=10	0.12	3.20	0.98	0.96	2266.69	-45.07	205681.05

2. Values of the validity indices with standard deviation of total monthly rainfall as attributes

fuzzifier(m)	Cluster no.	V_{PC}	V_{PE}	V_{FPI}	V_{NCE}	V_{XB}	V_{FS}	V_K
	no. of cluster=2	0.99	0.02	0.02	0.02	0.25	338.45	20.74
	no. of cluster=3	0.97	0.09	0.05	0.05	0.96	125.03	81.22
	no. of cluster=4	0.94	0.16	0.08	0.08	1.66	52.56	140.21
	no. of cluster=5	0.94	0.16	0.08	0.07	1.38	-117.46	121.00
1.1	no. of cluster=6	0.94	0.17	0.08	0.07	1.53	-162.50	132.99
	no. of cluster=7	0.93	0.19	0.08	0.07	1.69	-259.07	153.61
	no. of cluster=8	0.95	0.15	0.06	0.05	1.41	-262.17	123.34
	no. of cluster=9	0.96	0.12	0.05	0.04	1.24	-314.53	108.75
	no. of cluster=10	0.96	0.13	0.05	0.04	1.53	-344.30	133.51
	no. of cluster=2	0.97	0.07	0.05	0.07	0.25	341.76	21.48
	no. of cluster=3	0.90	0.24	0.14	0.15	0.99	134.51	83.95
	no. of cluster=4	0.83	0.43	0.22	0.21	1.80	69.69	152.33
	no. of cluster=5	0.85	0.39	0.19	0.17	1.48	-118.30	130.20
1.2	no. of cluster=6	0.83	0.42	0.20	0.16	1.43	-221.95	131.43
	no. of cluster=7	0.80	0.51	0.23	0.18	1.53	-259.04	139.50
	no. of cluster=8	0.80	0.55	0.22	0.18	1.44	-291.92	130.44
	no. of cluster=9	0.82	0.51	0.20	0.16	1.36	-338.81	123.19
	no. of cluster=10	0.80	0.60	0.22	0.18	1.75	-343.06	157.46
	no. of cluster=2	0.94	0.15	0.12	0.15	0.28	349.36	23.39
	no. of cluster=3	0.84	0.40	0.24	0.25	1.00	149.10	84.26
	no. of cluster=4	0.74	0.68	0.35	0.34	1.79	91.85	151.60
	no. of cluster=5	0.75	0.63	0.31	0.27	1.51	-101.72	133.32
1.3	no. of cluster=6	0.75	0.66	0.30	0.25	1.43	-214.16	131.99
	no. of cluster=7	0.70	0.82	0.35	0.29	1.53	-247.07	140.60
	no. of cluster=8	0.66	0.96	0.39	0.32	2.00	-272.43	182.92
	no. of cluster=9	0.67	0.94	0.37	0.30	1.81	-326.83	166.24
	no. of cluster=10	0.69	0.90	0.34	0.27	1.70	-371.00	156.64
1.4	no. of cluster=2	0.88	0.29	0.25	0.29	0.35	362.99	29.44

	no. of cluster=3	0.77	0.56	0.34	0.36	1.00	163.83	84.48
	no. of cluster=4	0.65	0.90	0.47	0.45	1.79	111.55	151.56
	no. of cluster=5	0.64	0.96	0.45	0.41	1.74	16.02	150.15
	no. of cluster=6	0.58	1.14	0.50	0.44	2.06	-53.38	183.85
	no. of cluster=7	0.60	1.11	0.47	0.40	1.67	-226.87	154.59
	no. of cluster=8	0.54	1.30	0.52	0.43	2.25	-248.23	207.56
	no. of cluster=9	0.53	1.41	0.53	0.45	2.56	-272.33	234.80
	no. of cluster=10	0.55	1.39	0.50	0.42	2.13	-323.00	196.99
	no. of cluster=2	0.82	0.42	0.36	0.42	0.44	372.57	36.62
	no. of cluster=3	0.71	0.72	0.44	0.45	1.01	176.26	85.40
	no. of cluster=4	0.57	1.09	0.57	0.55	1.82	126.33	154.03
	no. of cluster=5	0.55	1.20	0.56	0.52	1.77	52.01	152.25
1.5	no. of cluster=6	0.49	1.45	0.62	0.56	2.16	28.67	186.07
	no. of cluster=7	0.42	1.68	0.67	0.60	3.81	16.89	327.39
	no. of cluster=8	0.40	1.84	0.69	0.61	4.77	5.47	410.04
	no. of cluster=9	0.36	2.00	0.72	0.63	5.13	-7.76	440.94
	no. of cluster=10	0.40	1.91	0.67	0.58	3.35	-140.46	298.45
	no. of cluster=2	0.78	0.51	0.44	0.51	0.47	372.35	39.23
	no. of cluster=3	0.65	0.86	0.53	0.55	1.03	185.72	87.23
	no. of cluster=4	0.51	1.26	0.65	0.63	1.88	136.20	159.37
	no. of cluster=5	0.43	1.55	0.71	0.67	2.99	106.54	254.18
1.6	no. of cluster=6	0.41	1.68	0.71	0.65	2.15	52.62	184.58
	no. of cluster=7	0.35	1.92	0.76	0.68	5.50	40.30	473.74
	no. of cluster=8	0.32	2.10	0.78	0.70	6.50	29.41	559.60
	no. of cluster=9	0.28	2.28	0.80	0.72	146.33	17.98	12616.04
	no. of cluster=10	0.28	2.38	0.80	0.72	6.80	2.43	587.63
	no. of cluster=2	0.74	0.59	0.51	0.59	0.49	369.12	41.13
	no. of cluster=3	0.59	0.99	0.61	0.63	1.07	192.13	90.13
	no. of cluster=4	0.46	1.39	0.72	0.70	1.98	141.56	168.13
	no. of cluster=5	0.38	1.71	0.77	0.74	3.59	113.26	305.39
1.7	no. of cluster=6	0.34	1.88	0.79	0.73	3.87	69.41	333.07
	no. of cluster=7	0.30	2.10	0.82	0.75	7.07	53.71	610.75
	no. of cluster=8	0.26	2.33	0.85	0.78	10.04	45.52	868.02
	no. of cluster=9	0.23	2.49	0.86	0.79	75.32	35.39	6518.26
	no. of cluster=10	0.22	2.63	0.87	0.79	20.09	20.14	1742.02
	no. of cluster=2	0.71	0.66	0.58	0.66	0.51	363.66	42.82
	no. of cluster=3	0.54	1.10	0.68	0.70	1.12	195.63	94.61
	no. of cluster=4	0.42	1.51	0.77	0.75	2.14	142.85	181.23
	no. of cluster=5	0.34	1.83	0.82	0.79	3.83	115.42	326.12
1.8	no. of cluster=6	0.28	2.11	0.87	0.82	1943.11	97.10	166112.14
	no. of cluster=7	0.26	2.25	0.86	0.80	8.17	60.81	708.19
	no. of cluster=8	0.22	2.47	0.89	0.82	38.64	50.71	3355.18
	no. of cluster=9	0.20	2.63	0.90	0.83	245.61	39.38	21381.32
	no. of cluster=10	0.20	2.72	0.89	0.82	10.86	18.55	953.89

	no. of cluster=2	0.68	0.71	0.65	0.71	0.53	356.21	44.39
	no. of cluster=3	0.50	1.20	0.75	0.76	1.20	196.40	101.38
	no. of cluster=4	0.39	1.60	0.82	0.80	2.35	140.58	199.54
	no. of cluster=5	0.31	1.93	0.86	0.83	3.33	113.43	283.37
1.9	no. of cluster=6	0.25	2.21	0.90	0.85	8595.85	95.46	735453.73
	no. of cluster=7	0.22	2.42	0.91	0.86	548.66	77.90	47228.88
	no. of cluster=8	0.20	2.56	0.91	0.85	10799.28	50.47	943528.14
	no. of cluster=9	0.18	2.73	0.92	0.86	22201.31	36.25	1948643.61
	no. of cluster=10	0.17	2.89	0.93	0.87	143593.32	25.29	12648018.81
	no. of cluster=2	0.65	0.76	0.70	0.76	0.55	346.99	45.87
	no. of cluster=3	0.47	1.28	0.80	0.81	1.32	194.68	111.40
	no. of cluster=4	0.36	1.67	0.85	0.84	2.73	135.73	231.70
	no. of cluster=5	0.29	2.01	0.89	0.87	2.94	108.21	250.54
2	no. of cluster=6	0.24	2.28	0.92	0.88	196309.49	90.49	16809218.04
	no. of cluster=7	0.20	2.50	0.93	0.89	45682.84	75.19	3937692.68
	no. of cluster=8	0.18	2.69	0.94	0.90	17901.68	61.32	1555618.91
	no. of cluster=9	0.17	2.81	0.93	0.88	15170.37	30.83	1345009.26
	no. of cluster=10	0.15	2.96	0.94	0.89	208719.00	18.79	18610043.59
	no. of cluster=2	0.63	0.80	0.75	0.80	0.57	336.29	47.24
	no. of cluster=3	0.44	1.35	0.84	0.85	1.49	190.64	125.69
	no. of cluster=4	0.34	1.73	0.88	0.87	3.34	128.82	283.39
	no. of cluster=5	0.27	2.07	0.92	0.89	3.86	100.93	329.72
2.1	no. of cluster=6	0.22	2.34	0.93	0.90	21311.73	82.86	1828002.25
	no. of cluster=7	0.19	2.57	0.95	0.91	23622.57	70.60	2035095.96
	no. of cluster=8	0.17	2.75	0.95	0.92	219957.66	59.05	19130670.55
	no. of cluster=9	0.15	2.91	0.95	0.92	112016.31	34.94	9921205.69
	no. of cluster=10	0.15	3.02	0.95	0.91	76658.97	13.66	6928840.92
	no. of cluster=2	0.61	0.84	0.79	0.84	0.58	324.41	48.51
	no. of cluster=3	0.41	1.40	0.88	0.88	1.72	184.23	144.27
	no. of cluster=4	0.32	1.78	0.91	0.89	4.71	120.83	400.25
	no. of cluster=5	0.25	2.12	0.93	0.91	8.87	92.64	758.48
2.2	no. of cluster=6	0.21	2.38	0.95	0.92	8827.94	74.44	759045.81
	no. of cluster=7	0.18	2.61	0.96	0.93	6773.67	62.67	585263.98
	no. of cluster=8	0.16	2.81	0.96	0.94	125971.39	53.67	10943163.27
	no. of cluster=9	0.14	2.98	0.97	0.94	262441.27	47.11	22916510.51
	no. of cluster=10	0.13	3.09	0.96	0.93	871887.76	13.01	78950670.34
	no. of cluster=2	0.59	0.86	0.82	0.86	0.60	311.66	49.67
	no. of cluster=3	0.40	1.44	0.90	0.91	1.99	175.56	167.35
	no. of cluster=4	0.31	1.83	0.92	0.91	9.40	112.21	798.48
2.3	no. of cluster=5	0.24	2.15	0.95	0.93	60.64	83.65	5192.00
	no. of cluster=6	0.20	2.42	0.96	0.94	23974.16	65.90	2067276.88
	no. of cluster=7	0.17	2.65	0.97	0.94	5050.02	54.62	438091.64
	no. of cluster=8	0.15	2.84	0.97	0.95	11215.70	46.13	979052.87
	no. of cluster=9	0.13	3.01	0.97	0.95	535128.15	39.88	46999347.53

	no. of cluster=10	0.12	3.17	0.98	0.95	4472.72	34.86	395192.61
	no. of cluster=2	0.58	0.89	0.85	0.89	0.61	298.36	50.73
	no. of cluster=3	0.38	1.47	0.92	0.93	2.33	165.06	195.44
	no. of cluster=4	0.30	1.86	0.94	0.93	78.57	103.22	6676.17
	no. of cluster=5	0.24	2.18	0.95	0.94	127.70	74.74	10953.37
2.4	no. of cluster=6	0.20	2.45	0.96	0.95	91672.43	57.68	7930461.70
	no. of cluster=7	0.17	2.67	0.97	0.95	5104.38	46.97	444927.76
	no. of cluster=8	0.15	2.87	0.98	0.96	5055.86	39.00	444142.53
	no. of cluster=9	0.13	3.03	0.98	0.96	19411.85	32.90	1733661.57
	no. of cluster=10	0.12	3.19	0.98	0.96	278037.88	28.83	24795364.51
	no. of cluster=2	0.56	0.90	0.87	0.90	0.62	284.73	51.69
	no. of cluster=3	0.37	1.49	0.94	0.94	2.92	153.59	245.19
	no. of cluster=4	0.29	1.88	0.95	0.94	543.28	94.11	46138.95
	no. of cluster=5	0.23	2.21	0.96	0.95	121.36	66.42	10427.28
2.5	no. of cluster=6	0.19	2.47	0.97	0.95	26916.61	50.04	2336798.43
	no. of cluster=7	0.16	2.69	0.98	0.96	3243.20	40.15	284190.71
	no. of cluster=8	0.14	2.89	0.98	0.96	199253.14	32.77	17630585.34
	no. of cluster=9	0.13	3.06	0.98	0.97	308705.82	27.75	27572461.21
	no. of cluster=10	0.11	3.21	0.98	0.97	282659.85	23.66	25489394.26
	no. of cluster=2	0.55	0.92	0.89	0.92	0.63	271.02	52.59
	no. of cluster=3	0.37	1.51	0.95	0.95	3.44	141.63	288.22
	no. of cluster=4	0.28	1.91	0.96	0.95	42062.18	85.79	3566492.23
	no. of cluster=5	0.22	2.23	0.97	0.96	6267.63	58.69	539276.88
2.6	no. of cluster=6	0.19	2.49	0.97	0.96	163764.75	43.23	14269606.29
	no. of cluster=7	0.16	2.71	0.98	0.97	6674.83	34.08	588321.86
	no. of cluster=8	0.14	2.90	0.98	0.97	8419.19	27.43	750954.16
	no. of cluster=9	0.12	3.08	0.98	0.97	27592.90	22.92	2490294.74
	no. of cluster=10	0.11	3.23	0.99	0.97	442511.89	19.34	40415311.67
	no. of cluster=2	0.55	0.93	0.91	0.93	0.64	257.36	53.40
	no. of cluster=3	0.36	1.52	0.96	0.96	4.62	129.92	387.72
	no. of cluster=4	0.27	1.93	0.97	0.96	2040.98	78.09	172537.21
	no. of cluster=5	0.22	2.24	0.97	0.97	7031.27	51.67	605465.46
2.7	no. of cluster=6	0.18	2.50	0.98	0.97	45334.67	37.25	3964479.71
	no. of cluster=7	0.16	2.73	0.98	0.97	81513.69	28.81	7230032.57
	no. of cluster=8	0.14	2.92	0.98	0.97	26554.56	22.89	2389342.18
	no. of cluster=9	0.12	3.09	0.99	0.97	106335.87	18.86	9709382.78
	no. of cluster=10	0.11	3.24	0.99	0.98	39366.47	15.76	3647378.61
	no. of cluster=2	0.54	0.94	0.92	0.94	0.65	243.95	54.20
	no. of cluster=3	0.36	1.53	0.97	0.97	9.59	118.80	803.97
	no. of cluster=4	0.27	1.94	0.97	0.97	6419.36	70.00	541658.29
2.8	no. of cluster=5	0.22	2.26	0.98	0.97	12311.02	45.60	1058467.23
	no. of cluster=6	0.18	2.51	0.98	0.97	42088.20	32.03	3691641.14
	no. of cluster=7	0.16	2.74	0.98	0.98	53790.21	24.30	4802004.86
	no. of cluster=8	0.14	2.93	0.99	0.98	3535.84	19.05	321154.54

	no. of cluster=9	0.12	3.10	0.99	0.98	17759.35	15.49	1642537.47
	no. of cluster=10	0.11	3.25	0.99	0.98	442858.19	12.81	41694808.67
	no. of cluster=2	0.53	0.95	0.93	0.95	0.66	230.83	54.96
	no. of cluster=3	0.35	1.54	0.97	0.97	16.27	108.03	1364.64
	no. of cluster=4	0.27	1.95	0.98	0.97	81454.04	61.83	6878364.66
	no. of cluster=5	0.21	2.27	0.98	0.98	3116.00	40.62	265914.80
2.9	no. of cluster=6	0.18	2.53	0.99	0.98	2350.88	28.62	202470.06
	no. of cluster=7	0.15	2.75	0.99	0.98	70454.70	20.49	6324558.82
	no. of cluster=8	0.13	2.95	0.99	0.98	51407.99	16.54	4551524.72
	no. of cluster=9	0.12	3.12	0.99	0.98	31824.42	13.25	2867009.16
	no. of cluster=10	0.11	3.26	0.99	0.98	99116.27	10.43	9492035.95
	no. of cluster=2	0.53	0.96	0.94	0.96	0.67	218.09	55.69
	no. of cluster=3	0.35	1.55	0.97	0.98	46.19	98.04	3874.74
	no. of cluster=4	0.26	1.96	0.98	0.98	12147.97	54.51	1026786.30
	no. of cluster=5	0.21	2.28	0.99	0.98	3503.90	35.01	299596.48
3	no. of cluster=6	0.18	2.54	0.99	0.98	546739.66	24.23	47255585.23
	no. of cluster=7	0.15	2.76	0.99	0.98	31439.48	17.82	2763563.87
	no. of cluster=8	0.13	2.96	0.99	0.99	689.77	13.70	61700.84
	no. of cluster=9	0.12	3.13	0.99	0.99	22347.96	10.77	2034613.91
	no. of cluster=10	0.11	3.27	0.99	0.99	235556.73	8.45	23023351.58

3. *Values of the validity indices with latitude, longitude, elevation, total annual rainfall and standard deviation of total monthly rainfall as attributes*

fuzzifier(m)	**Cluster no.**	**V_{PC}**	**V_{PE}**	**V_{FPI}**	**V_{NCE}**	**V_{XB}**	**V_{FS}**	**V_{K}**
	no. of cluster=2	0.98	0.04	0.04	0.04	0.39	92.76	32.22
	no. of cluster=3	0.97	0.08	0.05	0.05	0.89	26.40	74.96
	no. of cluster=4	0.98	0.04	0.02	0.02	0.48	-67.31	42.73
	no. of cluster=5	0.97	0.07	0.04	0.03	0.72	-121.62	64.73
1.1	no. of cluster=6	0.98	0.07	0.03	0.03	1.25	-136.49	110.73
	no. of cluster=7	0.98	0.05	0.02	0.02	0.95	-170.66	85.25
	no. of cluster=8	0.99	0.04	0.02	0.01	1.09	-180.47	96.65
	no. of cluster=9	0.98	0.06	0.02	0.02	1.05	-185.56	93.27
	no. of cluster=10	1.00	0.02	0.00	0.00	0.47	-310.70	50.24
	no. of cluster=2	0.97	0.07	0.06	0.07	0.38	89.99	31.76
	no. of cluster=3	0.93	0.17	0.10	0.11	0.91	23.95	76.50
	no. of cluster=4	0.92	0.19	0.10	0.09	0.83	-52.51	69.89
	no. of cluster=5	0.94	0.16	0.08	0.07	0.73	-106.60	63.97
1.2	no. of cluster=6	0.93	0.19	0.08	0.07	1.39	-120.73	120.46
	no. of cluster=7	0.93	0.19	0.08	0.07	1.02	-151.74	90.99
	no. of cluster=8	0.94	0.16	0.06	0.05	0.89	-186.31	80.40
	no. of cluster=9	0.94	0.18	0.07	0.06	0.87	-199.61	78.08
	no. of cluster=10	0.95	0.14	0.06	0.04	0.77	-213.97	69.80
1.3	no. of cluster=2	0.95	0.12	0.10	0.12	0.37	85.98	31.01

	no. of cluster=3	0.90	0.27	0.16	0.17	0.91	19.24	76.30
	no. of cluster=4	0.89	0.30	0.15	0.15	0.78	-54.76	65.91
	no. of cluster=5	0.89	0.29	0.14	0.13	0.70	-91.19	59.81
	no. of cluster=6	0.89	0.31	0.13	0.12	0.86	-107.98	73.62
	no. of cluster=7	0.87	0.36	0.15	0.13	1.00	-151.85	89.45
	no. of cluster=8	0.89	0.31	0.12	0.10	0.86	-187.08	77.52
	no. of cluster=9	0.91	0.28	0.11	0.09	0.80	-198.06	72.25
	no. of cluster=10	0.90	0.29	0.11	0.09	0.72	-213.72	65.88
	no. of cluster=2	0.93	0.18	0.14	0.18	0.36	81.99	30.14
	no. of cluster=3	0.85	0.38	0.22	0.24	0.90	15.41	75.26
	no. of cluster=4	0.84	0.42	0.21	0.21	0.73	-56.53	61.57
	no. of cluster=5	0.82	0.51	0.22	0.22	1.04	-72.48	88.54
1.4	no. of cluster=6	0.84	0.47	0.20	0.18	0.82	-108.31	70.39
	no. of cluster=7	0.84	0.49	0.19	0.17	0.75	-120.20	64.03
	no. of cluster=8	0.83	0.49	0.19	0.16	0.81	-184.03	73.89
	no. of cluster=9	0.83	0.53	0.20	0.17	1.04	-187.65	94.94
	no. of cluster=10	0.82	0.56	0.20	0.17	0.96	-199.58	87.72
	no. of cluster=2	0.90	0.25	0.19	0.25	0.35	78.57	29.27
	no. of cluster=3	0.85	0.41	0.22	0.26	0.45	0.71	38.31
	no. of cluster=4	0.79	0.56	0.27	0.28	0.68	-55.73	57.44
	no. of cluster=5	0.75	0.70	0.31	0.30	1.04	-69.64	88.66
1.5	no. of cluster=6	0.77	0.67	0.28	0.26	0.81	-131.32	73.83
	no. of cluster=7	0.76	0.74	0.28	0.27	0.95	-116.88	81.53
	no. of cluster=8	0.77	0.71	0.27	0.24	0.77	-175.69	70.88
	no. of cluster=9	0.73	0.81	0.30	0.25	0.90	-181.64	83.44
	no. of cluster=10	0.74	0.80	0.28	0.24	0.99	-195.24	92.89
	no. of cluster=2	0.87	0.32	0.25	0.32	0.34	75.75	28.40
	no. of cluster=3	0.80	0.54	0.29	0.34	0.45	3.18	38.47
	no. of cluster=4	0.74	0.71	0.34	0.35	0.63	-52.56	53.55
	no. of cluster=5	0.73	0.79	0.34	0.34	0.54	-79.44	46.48
1.6	no. of cluster=6	0.68	0.94	0.39	0.36	0.85	-88.10	74.32
	no. of cluster=7	0.66	1.04	0.40	0.37	0.93	-97.37	81.29
	no. of cluster=8	0.68	1.00	0.37	0.33	0.78	-128.19	69.18
	no. of cluster=9	0.66	1.07	0.38	0.34	0.84	-135.49	74.70
	no. of cluster=10	0.65	1.13	0.39	0.34	1.01	-138.51	90.12
	no. of cluster=2	0.84	0.39	0.32	0.39	0.33	73.38	27.53
	no. of cluster=3	0.75	0.67	0.37	0.42	0.46	6.70	38.79
	no. of cluster=4	0.69	0.85	0.41	0.42	0.58	-47.72	49.85
	no. of cluster=5	0.61	1.10	0.49	0.47	1.22	-53.42	103.98
1.7	no. of cluster=6	0.59	1.19	0.49	0.46	0.99	-71.69	86.02
	no. of cluster=7	0.60	1.21	0.47	0.43	0.82	-98.84	72.60
	no. of cluster=8	0.57	1.35	0.49	0.45	1.02	-95.06	90.07
	no. of cluster=9	0.56	1.42	0.49	0.45	0.94	-100.68	83.25
	no. of cluster=10	0.59	1.36	0.46	0.41	0.71	-126.90	63.49

	no. of cluster=2	0.81	0.46	0.38	0.46	0.32	71.30	26.66
	no. of cluster=3	0.68	0.83	0.48	0.52	0.78	12.40	65.63
	no. of cluster=4	0.64	0.98	0.47	0.49	0.54	-41.97	46.30
	no. of cluster=5	0.59	1.16	0.51	0.50	0.65	-53.91	55.95
1.8	no. of cluster=6	0.53	1.38	0.57	0.53	0.97	-58.66	84.79
	no. of cluster=7	0.54	1.41	0.54	0.50	0.78	-83.89	68.84
	no. of cluster=8	0.54	1.48	0.53	0.49	0.68	-96.14	60.48
	no. of cluster=9	0.49	1.66	0.57	0.52	0.97	-84.38	86.91
	no. of cluster=10	0.51	1.63	0.55	0.49	0.79	-105.78	71.63
	no. of cluster=2	0.78	0.52	0.43	0.52	0.31	69.38	25.77
	no. of cluster=3	0.66	0.90	0.52	0.57	0.47	14.70	39.97
	no. of cluster=4	0.60	1.10	0.53	0.55	0.50	-35.91	42.90
	no. of cluster=5	0.51	1.40	0.61	0.60	1.14	-37.70	98.43
1.9	no. of cluster=6	0.48	1.54	0.63	0.60	0.93	-47.60	81.44
	no. of cluster=7	0.46	1.66	0.64	0.59	0.90	-57.09	79.79
	no. of cluster=8	0.45	1.74	0.62	0.58	0.78	-68.77	70.55
	no. of cluster=9	0.47	1.75	0.60	0.55	0.59	-87.68	54.00
	no. of cluster=10	0.43	1.93	0.64	0.58	0.95	-74.82	87.21
	no. of cluster=2	0.76	0.57	0.49	0.57	0.30	67.52	24.88
	no. of cluster=3	0.60	1.00	0.59	0.63	0.71	13.63	59.67
	no. of cluster=4	0.56	1.20	0.59	0.60	0.46	-30.00	39.65
	no. of cluster=5	0.47	1.51	0.66	0.65	1.10	-30.63	95.10
2	no. of cluster=6	0.43	1.68	0.68	0.65	0.88	-37.78	77.73
	no. of cluster=7	0.41	1.81	0.68	0.64	0.78	-46.68	70.14
	no. of cluster=8	0.39	1.97	0.70	0.66	0.75	-48.78	66.97
	no. of cluster=9	0.38	2.03	0.69	0.64	0.74	-57.88	67.97
	no. of cluster=10	0.40	2.05	0.67	0.62	0.58	-73.27	54.89
	no. of cluster=2	0.73	0.62	0.54	0.62	0.29	65.66	23.99
	no. of cluster=3	0.57	1.07	0.64	0.68	0.67	14.23	56.81
	no. of cluster=4	0.52	1.30	0.63	0.65	0.42	-24.48	36.56
	no. of cluster=5	0.45	1.58	0.69	0.68	1.03	-25.80	89.64
2.1	no. of cluster=6	0.39	1.80	0.73	0.70	0.83	-29.04	73.76
	no. of cluster=7	0.37	1.96	0.73	0.70	0.63	-35.32	56.69
	no. of cluster=8	0.36	2.05	0.73	0.68	0.55	-43.66	50.62
	no. of cluster=9	0.34	2.19	0.74	0.69	0.63	-44.09	58.22
	no. of cluster=10	0.35	2.24	0.73	0.67	0.55	-52.70	52.47
	no. of cluster=2	0.71	0.66	0.58	0.66	0.28	63.75	23.09
	no. of cluster=3	0.55	1.13	0.68	0.71	0.64	14.72	54.08
	no. of cluster=4	0.49	1.38	0.68	0.69	0.39	-19.51	33.64
	no. of cluster=5	0.42	1.68	0.73	0.72	0.94	-20.31	82.26
2.2	no. of cluster=6	0.36	1.90	0.76	0.74	0.78	-22.28	69.43
	no. of cluster=7	0.34	2.05	0.76	0.73	0.63	-28.94	57.20
	no. of cluster=8	0.33	2.17	0.76	0.72	0.47	-35.56	43.91
	no. of cluster=9	0.31	2.33	0.78	0.73	0.57	-35.10	53.55

	no. of cluster=10	0.30	2.42	0.78	0.73	0.54	-36.93	52.26
	no. of cluster=2	0.69	0.69	0.62	0.69	0.26	61.80	22.19
	no. of cluster=3	0.52	1.18	0.72	0.75	0.61	15.10	51.51
	no. of cluster=4	0.46	1.45	0.72	0.73	0.36	-15.18	30.91
	no. of cluster=5	0.38	1.78	0.78	0.77	0.92	-14.69	81.05
2.3	no. of cluster=6	0.33	2.00	0.80	0.77	1.17	-15.27	104.75
	no. of cluster=7	0.30	2.20	0.82	0.78	1.31	-17.24	119.72
	no. of cluster=8	0.29	2.31	0.81	0.77	1.08	-24.77	102.01
	no. of cluster=9	0.28	2.42	0.81	0.76	0.47	-29.74	45.57
	no. of cluster=10	0.25	2.61	0.84	0.79	0.94	-21.43	90.62
	no. of cluster=2	0.67	0.73	0.66	0.73	0.25	59.80	21.30
	no. of cluster=3	0.50	1.23	0.75	0.78	0.58	15.34	49.13
	no. of cluster=4	0.44	1.52	0.75	0.76	0.33	-11.50	28.37
	no. of cluster=5	0.36	1.84	0.81	0.79	0.85	-11.05	74.87
2.4	no. of cluster=6	0.32	2.06	0.82	0.80	0.65	-12.79	59.41
	no. of cluster=7	0.27	2.27	0.85	0.81	1.29	-12.85	119.27
	no. of cluster=8	0.25	2.47	0.86	0.82	1.50	-12.42	141.32
	no. of cluster=9	0.25	2.53	0.84	0.80	0.41	-19.51	40.47
	no. of cluster=10	0.25	2.62	0.83	0.79	0.39	-25.74	39.89
	no. of cluster=2	0.66	0.75	0.69	0.75	0.24	57.74	20.42
	no. of cluster=3	0.48	1.27	0.78	0.80	0.55	15.43	46.95
	no. of cluster=4	0.42	1.57	0.78	0.79	0.30	-8.44	26.02
	no. of cluster=5	0.34	1.90	0.83	0.82	0.77	-8.11	68.62
2.5	no. of cluster=6	0.30	2.12	0.84	0.82	0.59	-9.38	54.07
	no. of cluster=7	0.25	2.35	0.87	0.84	1.45	-8.39	136.52
	no. of cluster=8	0.25	2.47	0.86	0.82	1.09	-11.92	106.54
	no. of cluster=9	0.23	2.61	0.87	0.82	0.99	-11.94	99.61
	no. of cluster=10	0.21	2.76	0.88	0.83	1.25	-11.68	128.78
	no. of cluster=2	0.64	0.78	0.72	0.78	0.23	55.64	19.55
	no. of cluster=3	0.47	1.30	0.80	0.82	0.53	15.37	44.94
	no. of cluster=4	0.40	1.62	0.80	0.81	0.27	-5.94	23.87
	no. of cluster=5	0.32	1.94	0.85	0.84	0.69	-5.83	62.34
2.6	no. of cluster=6	0.28	2.18	0.86	0.84	0.60	-6.64	56.16
	no. of cluster=7	0.25	2.37	0.87	0.84	0.40	-7.78	38.38
	no. of cluster=8	0.22	2.59	0.90	0.86	1.21	-6.68	118.77
	no. of cluster=9	0.20	2.72	0.90	0.86	1.31	-7.41	135.05
	no. of cluster=10	0.20	2.80	0.88	0.84	1.01	-11.13	108.86
	no. of cluster=2	0.63	0.80	0.74	0.80	0.22	53.51	18.70
	no. of cluster=3	0.45	1.34	0.82	0.84	0.50	15.19	43.14
	no. of cluster=4	0.38	1.66	0.82	0.83	0.25	-3.94	21.90
2.7	no. of cluster=5	0.31	1.99	0.87	0.86	0.62	-4.08	56.24
	no. of cluster=6	0.26	2.22	0.88	0.86	0.68	-4.51	63.94
	no. of cluster=7	0.23	2.45	0.90	0.87	1.10	-4.08	106.92
	no. of cluster=8	0.22	2.59	0.90	0.86	0.86	-6.63	89.05

	no. of cluster=9	0.19	2.77	0.91	0.87	35.62	-5.38	3820.21
	no. of cluster=10	0.17	2.93	0.92	0.88	12.05	-6.10	1334.22
	no. of cluster=2	0.62	0.82	0.76	0.82	0.21	51.37	17.86
	no. of cluster=3	0.44	1.36	0.84	0.86	0.48	14.89	41.56
	no. of cluster=4	0.37	1.70	0.84	0.85	0.23	-2.36	20.36
	no. of cluster=5	0.29	2.02	0.88	0.87	0.55	-2.71	50.66
2.8	no. of cluster=6	0.25	2.27	0.90	0.88	0.88	-2.91	84.35
	no. of cluster=7	0.22	2.49	0.91	0.89	0.92	-2.64	91.66
	no. of cluster=8	0.20	2.63	0.91	0.88	0.74	-4.67	79.20
	no. of cluster=9	0.18	2.82	0.92	0.89	13.68	-3.82	1532.35
	no. of cluster=10	0.16	2.97	0.93	0.89	10.73	-4.94	1257.02
	no. of cluster=2	0.61	0.83	0.78	0.83	0.20	49.22	17.05
	no. of cluster=3	0.43	1.39	0.86	0.87	0.47	14.49	40.11
	no. of cluster=4	0.35	1.73	0.86	0.86	0.22	-1.09	19.90
	no. of cluster=5	0.28	2.05	0.89	0.88	0.49	-1.66	45.75
2.9	no. of cluster=6	0.24	2.30	0.91	0.89	0.87	-1.82	85.32
	no. of cluster=7	0.21	2.53	0.92	0.90	0.74	-1.65	76.29
	no. of cluster=8	0.19	2.70	0.93	0.90	0.78	-1.99	85.22
	no. of cluster=9	0.17	2.86	0.93	0.90	54.48	-2.52	6384.16
	no. of cluster=10	0.15	3.01	0.94	0.91	46.39	-2.74	5704.69
	no. of cluster=2	0.60	0.85	0.80	0.85	0.19	47.10	16.26
	no. of cluster=3	0.42	1.41	0.87	0.89	0.45	14.00	38.87
	no. of cluster=4	0.34	1.75	0.87	0.88	0.21	-0.12	19.32
	no. of cluster=5	0.27	2.08	0.91	0.90	0.43	-0.88	40.66
3	no. of cluster=6	0.23	2.33	0.92	0.90	0.37	-1.17	37.10
	no. of cluster=7	0.20	2.56	0.93	0.91	0.58	-1.01	61.12
	no. of cluster=8	0.17	2.76	0.95	0.92	6.35	-0.77	701.28
	no. of cluster=9	0.16	2.89	0.94	0.91	333.22	-1.70	41258.30
	no. of cluster=10	0.14	3.06	0.95	0.92	34.86	-1.09	4469.13

ANNEXURE C

SELECTED PREDICTORS FOR STEPWISE MULTIPLE LINEAR REGRESSION (HadCM3 MODEL) DURING SCREENING

1. *Pearson coefficient (PC): Maximum temperature*

pearson coeff	hur200	hur500	hur850	mrso	pr	prc	psl	rsds	snm	ta200	ta500	ta850	tas	ts	ua200	ua500	ua850	uas	va200	va500	va850	vas	zg200	zg500	zg850
Cherrapunjee	0.55	0.61	0.63	0.47	0.67	0.67	-0.70	0.21	-0.30	0.53	0.71	0.81	0.84	0.84	-0.72	-0.76	-0.35	0.23	-0.39	0.37	0.04	0.33	0.70	0.83	-0.51
Imphal	0.64	0.71	0.65	0.37	0.75	0.76	-0.81	0.37	-0.30	0.57	0.72	0.90	0.91	0.91	-0.76	-0.77	0.46	0.68	-0.52	0.14	0.51	0.57	0.71	0.77	-0.65
Shillong	0.64	0.73	0.62	0.40	0.76	0.78	-0.87	0.33	-0.30	0.61	0.76	0.93	0.95	0.94	-0.81	-0.81	-0.18	0.43	-0.53	0.22	0.34	0.60	0.75	0.81	-0.73
Silchar	0.62	0.70	0.62	0.38	0.72	0.73	-0.77	0.36	-0.31	0.53	0.67	0.86	0.87	0.87	-0.72	-0.76	0.07	0.57	-0.48	0.29	0.35	0.49	0.67	0.73	-0.63
Guwahati	0.64	0.73	0.64	0.47	0.76	0.78	-0.83	0.15	-0.31	0.62	0.77	0.88	0.90	0.90	-0.82	-0.80	-0.30	0.24	-0.47	0.32	0.32	0.56	0.77	0.81	-0.69

2. *Pearson coefficient (PC): Minimum Temperature*

pearson coeff	hur200	hur500	hur850	mrso	pr	prc	psl	rsds	snm	ta200	ta500	ta850	tas	ts	ua200	ua500	ua850	uas	va200	va500	va850	vas	zg200	zg500	zg850
Cherrapunjee	0.75	0.80	0.73	0.56	0.83	0.84	-0.86	0.19	-0.37	0.72	0.85	0.88	0.92	0.91	-0.90	-0.90	-0.43	0.20	-0.55	0.35	0.08	0.40	0.86	0.86	-0.73
Imphal	0.80	0.83	0.76	0.56	0.85	0.87	-0.89	0.27	-0.23	0.75	0.88	0.94	0.95	0.96	-0.92	-0.93	0.37	0.65	-0.61	0.14	0.46	0.53	0.89	0.89	-0.74
Shillong	0.76	0.81	0.74	0.57	0.83	0.85	-0.88	0.16	-0.40	0.73	0.87	0.91	0.94	0.94	-0.92	-0.92	-0.39	0.25	-0.51	0.31	0.17	0.46	0.87	0.88	-0.74
Silchar	0.76	0.79	0.75	0.56	0.83	0.84	-0.85	0.26	-0.25	0.72	0.86	0.91	0.93	0.93	-0.90	-0.92	-0.12	0.47	-0.56	0.29	0.24	0.41	0.87	0.89	-0.70
Guwahati	0.77	0.83	0.78	0.65	0.84	0.85	-0.86	-0.02	-0.45	0.76	0.90	0.87	0.91	0.91	-0.94	-0.92	-0.50	0.05	-0.51	0.31	0.19	0.46	0.91	0.92	-0.69

3. *Pearson coefficient (PC): Annual Rainfall*

pearson coeff	hur200	hur500	hur850	mrso	pr	prc	psl	rsds	snm	ta200	ta500	ta850	tas	ts	ua200	ua500	ua850	uas	va200	va500	va850	vas	zg200	zg500	zg850
Cherrapunjee	0.64	0.67	0.52	0.39	0.67	0.69	-0.76	0.18	-0.31	0.68	0.68	0.69	0.70	0.70	-0.74	-0.69	-0.27	0.18	-0.45	0.10	0.20	0.42	0.69	0.59	-0.74
Imphal	0.59	0.62	0.66	0.36	0.72	0.70	-0.78	0.16	-0.19	0.65	0.68	0.72	0.73	0.73	-0.72	-0.72	0.44	0.61	-0.54	-0.15	0.54	0.58	0.67	0.62	-0.72
Shillong	0.62	0.68	0.61	0.53	0.75	0.77	-0.76	0.05	-0.35	0.74	0.74	0.72	0.74	0.73	-0.78	-0.73	-0.38	0.08	-0.38	0.13	0.13	0.36	0.76	0.68	-0.70
Silchar	0.71	0.74	0.68	0.41	0.79	0.81	-0.80	0.26	-0.19	0.66	0.74	0.80	0.82	0.82	-0.81	-0.79	0.06	0.50	-0.61	0.02	0.43	0.53	0.76	0.73	-0.71
Guwahati	0.65	0.70	0.55	0.42	0.72	0.75	-0.79	0.09	-0.36	0.67	0.73	0.77	0.78	0.78	-0.78	-0.75	-0.31	0.10	-0.46	0.07	0.33	0.51	0.74	0.70	-0.72

ANNEXURE D

CROSS-CORRELATION TABLE FOR CHERRAPUNJEE

	hur200	*hur500*	*hur850*	*mrso*	*pr*	*prc*	*psl*	*rsds*	*snm*	*ta200*	*ta500*	*ta850*	*tas*	*ts*	*ua200*	*ua500*	*ua850*	*uas*	*va200*	*va500*	*va850*	*vas*	*zg200*	*zg500*	*zg850*
hur200	1.00																								
hur500	1.00	1.00																							
hur850	0.78	0.78	1.00																						
mrso	0.74	0.74	0.79	1.00																					
pr	0.83	0.83	0.78	0.64	1.00																				
prc	0.85	0.85	0.79	0.64	0.99	1.00																			
psl	-0.69	-0.69	-0.54	-0.35	-0.77	-0.80	1.00																		
rsds	-0.19	-0.19	-0.44	-0.52	-0.18	-0.14	-0.39	1.00																	
snm	-0.28	-0.28	-0.27	-0.30	-0.20	-0.21	0.27	-0.01	1.00																
ta200	0.68	0.68	0.70	0.63	0.73	0.75	-0.74	-0.08	-0.34	1.00															
ta500	0.79	0.79	0.79	0.70	0.75	0.77	-0.78	-0.05	-0.47	0.86	1.00														
ta850	0.59	0.59	0.50	0.30	0.66	0.69	-0.93	0.49	-0.34	0.63	0.74	1.00													
tas	0.64	0.64	0.59	0.38	0.69	0.73	-0.93	0.41	-0.34	0.66	0.79	0.99	1.00												
ts	0.64	0.64	0.59	0.38	0.69	0.72	-0.93	0.42	-0.35	0.66	0.79	0.99	1.00	1.00											
ua200	-0.84	-0.84	-0.83	-0.72	-0.81	-0.84	0.82	0.03	0.40	-0.87	-0.95	-0.77	-0.82	-0.82	1.00										
ua500	-0.83	-0.83	-0.83	-0.71	-0.80	-0.82	0.77	0.06	0.39	-0.80	-0.87	-0.76	-0.81	-0.81	0.95	1.00									
ua850	-0.48	-0.48	-0.58	-0.74	-0.45	-0.44	0.17	0.53	0.35	-0.56	-0.58	-0.15	-0.20	-0.20	0.56	0.59	1.00								
uas	0.01	0.01	-0.08	-0.36	0.02	0.06	-0.40	0.66	0.14	-0.07	-0.02	0.42	0.40	0.39	-0.03	-0.01	0.75	1.00							
va200	-0.43	-0.43	-0.40	-0.21	-0.44	-0.47	0.57	-0.17	0.11	-0.42	-0.49	-0.49	-0.52	-0.51	0.49	0.43	0.00	-0.34	1.00						
va500	0.49	0.49	0.41	0.61	0.32	0.31	-0.13	-0.24	-0.07	0.24	0.33	0.14	0.19	0.19	-0.37	-0.40	-0.52	-0.25	0.15	1.00					
va850	-0.09	-0.09	-0.24	-0.53	0.06	0.08	-0.42	0.71	0.15	-0.09	-0.11	0.40	0.35	0.34	0.07	0.11	0.76	0.83	-0.36	-0.53	1.00				
vas	0.16	0.16	0.05	-0.29	0.31	0.34	-0.68	0.69	0.02	0.18	0.20	0.66	0.63	0.62	-0.24	-0.20	0.56	0.84	-0.49	-0.34	0.94	1.00			
zg200	0.82	0.82	0.81	0.74	0.77	0.79	-0.77	-0.09	-0.46	0.89	0.99	0.74	0.78	0.78	-0.96	-0.90	-0.62	-0.06	-0.45	0.39	-0.15	0.16	1.00		
zg500	0.67	0.67	0.74	0.63	0.63	0.66	-0.68	0.02	-0.52	0.70	0.90	0.79	0.83	0.83	-0.87	-0.86	-0.54	0.00	-0.36	0.32	-0.12	0.18	0.90	1.00	
zg850	-0.65	-0.65	-0.46	-0.28	-0.75	-0.77	0.96	-0.35	0.18	-0.72	-0.69	-0.80	-0.79	-0.78	0.74	0.66	0.12	-0.38	0.56	-0.05	-0.45	-0.67	-0.67	-0.49	1.00

CROSS-CORRELATION TABLE FOR IMPHAL

	hur200	*hur500*	*hur850*	*mrso*	*pr*	*prc*	*psl*	*rsds*	*snm*	*ta200*	*ta500*	*ta850*	*tas*	*ts*	*ua200*	*ua500*	*ua850*	*uas*	*va200*	*va500*	*va850*	*vas*	*zg200*	*zg500*	*zg850*
hur200	1.00																								
hur500	0.95	1.00																							
hur850	0.69	0.77	1.00																						
mrso	0.72	0.72	0.59	1.00																					
pr	0.76	0.80	0.75	0.57	1.00																				
prc	0.80	0.83	0.78	0.58	0.99	1.00																			
psl	-0.75	-0.77	-0.71	-0.40	-0.79	-0.82	1.00																		
rsds	0.07	0.05	-0.18	-0.30	-0.12	-0.07	-0.34	1.00																	
snm	-0.16	-0.19	-0.27	-0.06	-0.18	-0.18	0.20	-0.08	1.00																
ta200	0.70	0.74	0.67	0.64	0.68	0.70	-0.76	0.01	-0.13	1.00															
ta500	0.81	0.78	0.73	0.70	0.70	0.73	-0.81	0.07	-0.11	0.86	1.00														
ta850	0.72	0.74	0.66	0.43	0.68	0.73	-0.91	0.43	-0.30	0.67	0.81	1.00													
tas	0.76	0.79	0.76	0.47	0.72	0.76	-0.92	0.38	-0.31	0.69	0.83	0.99	1.00												
ts	0.76	0.79	0.76	0.48	0.71	0.76	-0.91	0.38	-0.30	0.69	0.83	0.99	1.00	1.00											
ua200	-0.85	-0.87	-0.78	-0.71	-0.74	-0.78	0.84	-0.11	0.17	-0.87	-0.95	-0.84	-0.86	-0.87	1.00										
ua500	-0.85	-0.87	-0.75	-0.69	-0.75	-0.79	0.81	-0.11	0.27	-0.80	-0.88	-0.85	-0.87	-0.87	0.95	1.00									
ua850	0.12	0.18	0.32	-0.29	0.29	0.31	-0.55	0.47	-0.34	0.11	0.08	0.46	0.46	0.45	-0.16	-0.17	1.00								
uas	0.32	0.38	0.49	-0.04	0.45	0.48	-0.73	0.43	-0.37	0.35	0.37	0.70	0.69	0.68	-0.45	-0.47	0.90	1.00							
va200	-0.48	-0.52	-0.58	-0.25	-0.46	-0.49	0.63	-0.20	0.15	-0.47	-0.55	-0.55	-0.61	-0.60	0.55	0.51	-0.41	-0.47	1.00						
va500	0.32	0.34	0.02	0.49	0.12	0.12	-0.03	-0.04	0.16	0.12	0.20	0.15	0.13	0.14	-0.23	-0.25	-0.46	-0.28	0.27	1.00					
va850	0.38	0.45	0.49	-0.09	0.51	0.53	-0.68	0.37	-0.32	0.31	0.25	0.53	0.57	0.56	-0.34	-0.35	0.89	0.81	-0.56	-0.37	1.00				
vas	0.39	0.45	0.53	-0.08	0.52	0.55	-0.74	0.40	-0.34	0.34	0.31	0.61	0.64	0.63	-0.39	-0.41	0.92	0.89	-0.56	-0.36	0.98	1.00			
zg200	0.84	0.83	0.73	0.74	0.71	0.74	-0.79	0.05	-0.13	0.89	0.98	0.81	0.83	0.83	-0.96	-0.91	0.05	0.34	-0.51	0.26	0.23	0.29	1.00		
zg500	0.71	0.69	0.66	0.64	0.59	0.63	-0.68	0.11	-0.21	0.69	0.90	0.85	0.84	0.85	-0.87	-0.86	0.06	0.38	-0.42	0.24	0.15	0.24	0.91	1.00	
zg850	-0.67	-0.69	-0.62	-0.30	-0.75	-0.77	0.96	-0.29	0.12	-0.72	-0.70	-0.76	-0.77	-0.76	0.73	0.68	-0.58	-0.70	0.61	0.06	-0.73	-0.76	-0.67	-0.47	1.00

CROSS-CORRELATION TABLE FOR SHILLONG

	hur200	*hur500*	*hur850*	*mrso*	*pr*	*prc*	*psl*	*rsds*	*snm*	*ta200*	*ta500*	*ta850*	*tas*	*ts*	*ua200*	*ua500*	*ua850*	*uas*	*va200*	*va500*	*va850*	*vas*	*zg200*	*zg500*	*zg850*
hur200	1.00																								
hur500	0.95	1.00																							
hur850	0.78	0.81	1.00																						
mrso	0.74	0.73	0.82	1.00																					
pr	0.82	0.87	0.77	0.63	1.00																				
prc	0.85	0.89	0.78	0.62	0.99	1.00																			
psl	-0.71	-0.75	-0.57	-0.39	-0.79	-0.83	1.00																		
rsds	-0.23	-0.20	-0.47	-0.56	-0.18	-0.14	-0.33	1.00																	
snm	-0.28	-0.20	-0.27	-0.30	-0.19	-0.21	0.28	0.00	1.00																
ta200	0.68	0.73	0.71	0.64	0.71	0.74	-0.74	-0.12	-0.34	1.00															
ta500	0.79	0.77	0.80	0.72	0.74	0.77	-0.80	-0.09	-0.46	0.86	1.00														
ta850	0.62	0.66	0.52	0.34	0.69	0.73	-0.94	0.44	-0.34	0.64	0.76	1.00													
tas	0.67	0.71	0.61	0.41	0.72	0.76	-0.93	0.36	-0.34	0.67	0.80	0.99	1.00												
ts	0.66	0.71	0.61	0.41	0.72	0.75	-0.93	0.37	-0.35	0.67	0.80	0.99	1.00	1.00											
ua200	-0.85	-0.87	-0.84	-0.73	-0.81	-0.84	0.84	0.08	0.40	-0.87	-0.95	-0.79	-0.84	-0.84	1.00										
ua500	-0.83	-0.86	-0.84	-0.72	-0.79	-0.82	0.78	0.10	0.39	-0.80	-0.87	-0.78	-0.82	-0.83	0.95	1.00									
ua850	-0.45	-0.42	-0.57	-0.72	-0.39	-0.38	0.13	0.59	0.34	-0.53	-0.53	-0.11	-0.15	-0.16	0.52	0.55	1.00								
uas	0.02	0.08	-0.08	-0.33	0.06	0.10	-0.43	0.67	0.12	-0.04	0.02	0.46	0.43	0.43	-0.07	-0.03	0.77	1.00							
va200	-0.43	-0.46	-0.39	-0.22	-0.44	-0.47	0.57	-0.14	0.11	-0.42	-0.49	-0.49	-0.52	-0.51	0.49	0.42	-0.04	-0.36	1.00						
va500	0.46	0.47	0.39	0.58	0.28	0.27	-0.12	-0.26	-0.05	0.22	0.30	0.14	0.19	0.19	-0.34	-0.36	-0.49	-0.25	0.17	1.00					
va850	-0.01	0.05	-0.19	-0.45	0.16	0.18	-0.47	0.72	0.13	-0.03	-0.03	0.45	0.41	0.40	-0.01	0.04	0.76	0.86	-0.40	-0.44	1.00				
vas	0.22	0.29	0.09	-0.21	0.38	0.41	-0.70	0.67	0.01	0.21	0.26	0.68	0.65	0.65	-0.30	-0.24	0.56	0.86	-0.51	-0.26	0.94	1.00			
zg200	0.82	0.81	0.82	0.76	0.76	0.79	-0.79	-0.13	-0.46	0.89	0.99	0.75	0.80	0.80	-0.96	-0.90	-0.58	-0.02	-0.45	0.35	-0.07	0.22	1.00		
zg500	0.69	0.68	0.76	0.66	0.64	0.67	-0.70	-0.02	-0.51	0.71	0.91	0.80	0.84	0.84	-0.88	-0.86	-0.50	0.04	-0.36	0.30	-0.05	0.22	0.91	1.00	
zg850	-0.65	-0.70	-0.46	-0.29	-0.76	-0.79	0.96	-0.32	0.18	-0.71	-0.69	-0.80	-0.79	-0.79	0.73	0.66	0.05	-0.43	0.56	-0.04	-0.52	-0.71	-0.67	-0.50	1.00

CROSS-CORRELATION TABLE FOR SILCHAR

	hur200	*hur500*	*hur850*	*mrso*	*pr*	*prc*	*psl*	*rsds*	*snm*	*ta200*	*ta500*	*ta850*	*tas*	*ts*	*ua200*	*ua500*	*ua850*	*uas*	*va200*	*va500*	*va850*	*vas*	*zg200*	*zg500*	*zg850*
hur200	1.00																								
hur500	0.95	1.00																							
hur850	0.76	0.81	1.00																						
mrso	0.73	0.73	0.69	1.00																					
pr	0.81	0.84	0.78	0.60	1.00																				
prc	0.84	0.87	0.80	0.61	0.99	1.00																			
psl	-0.71	-0.74	-0.65	-0.35	-0.79	-0.81	1.00																		
rsds	-0.02	-0.03	-0.25	-0.39	-0.11	-0.06	-0.40	1.00																	
snm	-0.16	-0.19	-0.27	-0.06	-0.18	-0.19	0.21	-0.09	1.00																
ta200	0.69	0.74	0.71	0.62	0.72	0.74	-0.75	-0.01	-0.12	1.00															
ta500	0.80	0.77	0.78	0.69	0.73	0.76	-0.79	0.04	-0.11	0.86	1.00														
ta850	0.65	0.67	0.62	0.35	0.69	0.72	-0.92	0.48	-0.31	0.65	0.78	1.00													
tas	0.70	0.73	0.71	0.42	0.72	0.76	-0.92	0.42	-0.31	0.68	0.81	0.99	1.00												
ts	0.70	0.73	0.71	0.42	0.72	0.75	-0.92	0.42	-0.31	0.68	0.81	0.99	1.00	1.00											
ua200	-0.85	-0.86	-0.82	-0.70	-0.79	-0.82	0.83	-0.07	0.17	-0.87	-0.95	-0.80	-0.84	-0.84	1.00										
ua500	-0.84	-0.86	-0.81	-0.69	-0.78	-0.82	0.79	-0.06	0.27	-0.81	-0.88	-0.81	-0.85	-0.85	0.95	1.00									
ua850	-0.25	-0.20	-0.19	-0.62	-0.12	-0.11	-0.18	0.57	-0.21	-0.28	-0.31	0.15	0.12	0.11	0.27	0.28	1.00								
uas	0.20	0.26	0.27	-0.19	0.29	0.32	-0.64	0.61	-0.35	0.20	0.23	0.64	0.63	0.62	-0.30	-0.31	0.78	1.00							
va200	-0.46	-0.49	-0.52	-0.23	-0.47	-0.49	0.61	-0.21	0.15	-0.46	-0.54	-0.53	-0.58	-0.57	0.53	0.48	-0.19	-0.43	1.00						
va500	0.42	0.44	0.24	0.57	0.23	0.24	-0.08	-0.13	0.12	0.19	0.28	0.14	0.17	0.17	-0.31	-0.34	-0.54	-0.26	0.20	1.00					
va850	0.16	0.22	0.17	-0.33	0.32	0.34	-0.59	0.54	-0.27	0.14	0.08	0.49	0.48	0.47	-0.15	-0.15	0.84	0.85	-0.49	-0.46	1.00				
vas	0.26	0.32	0.30	-0.23	0.42	0.44	-0.70	0.56	-0.31	0.25	0.23	0.63	0.62	0.62	-0.30	-0.30	0.79	0.91	-0.53	-0.38	0.98	1.00			
zg200	0.83	0.82	0.79	0.73	0.75	0.78	-0.78	0.01	-0.13	0.89	0.98	0.77	0.81	0.81	-0.96	-0.91	-0.35	0.19	-0.49	0.34	0.06	0.20	1.00		
zg500	0.69	0.67	0.72	0.63	0.62	0.64	-0.68	0.09	-0.22	0.69	0.90	0.82	0.83	0.84	-0.86	-0.86	-0.29	0.25	-0.40	0.29	0.03	0.19	0.90	1.00	
zg850	-0.66	-0.69	-0.56	-0.27	-0.76	-0.78	0.96	-0.35	0.13	-0.73	-0.70	-0.78	-0.78	-0.78	0.74	0.68	-0.22	-0.60	0.59	0.00	-0.62	-0.71	-0.67	-0.48	1.00

CROSS-CORRELATION TABLE FOR GUWAHATI

	hur200	*hur500*	*hur850*	*mrso*	*pr*	*prc*	*psl*	*rsds*	*snm*	*ta200*	*ta500*	*ta850*	*tas*	*ts*	*ua200*	*ua500*	*ua850*	*uas*	*va200*	*va500*	*va850*	*vas*	*zg200*	*zg500*	*zg850*
hur200	1.00																								
hur500	0.95	1.00																							
hur850	0.77	0.79	1.00																						
mrso	0.74	0.73	0.86	1.00																					
pr	0.82	0.88	0.74	0.63	1.00																				
prc	0.84	0.90	0.74	0.60	0.99	1.00																			
psl	-0.73	-0.78	-0.56	-0.42	-0.81	-0.85	1.00																		
rsds	-0.34	-0.29	-0.57	-0.63	-0.22	-0.17	-0.22	1.00																	
snm	-0.27	-0.21	-0.27	-0.29	-0.18	-0.20	0.29	0.03	1.00																
ta200	0.67	0.73	0.71	0.64	0.70	0.72	-0.75	-0.21	-0.34	1.00															
ta500	0.79	0.78	0.81	0.72	0.74	0.77	-0.83	-0.18	-0.46	0.86	1.00														
ta850	0.62	0.69	0.50	0.35	0.71	0.76	-0.94	0.36	-0.34	0.64	0.77	1.00													
tas	0.67	0.74	0.58	0.42	0.74	0.78	-0.94	0.28	-0.34	0.67	0.81	0.99	1.00												
ts	0.66	0.73	0.58	0.42	0.73	0.78	-0.93	0.28	-0.35	0.67	0.81	0.99	1.00	1.00											
ua200	-0.85	-0.88	-0.84	-0.73	-0.81	-0.84	0.86	0.18	0.39	-0.86	-0.95	-0.80	-0.85	-0.85	1.00										
ua500	-0.82	-0.86	-0.83	-0.71	-0.78	-0.82	0.79	0.19	0.37	-0.79	-0.85	-0.78	-0.82	-0.83	0.95	1.00									
ua850	-0.49	-0.47	-0.66	-0.73	-0.42	-0.40	0.19	0.63	0.35	-0.56	-0.55	-0.15	-0.20	-0.21	0.55	0.60	1.00								
uas	-0.10	-0.03	-0.25	-0.40	-0.03	0.02	-0.30	0.68	0.18	-0.13	-0.06	0.35	0.32	0.31	0.04	0.11	0.82	1.00							
va200	-0.42	-0.46	-0.33	-0.22	-0.43	-0.45	0.54	-0.07	0.09	-0.39	-0.47	-0.47	-0.49	-0.49	0.47	0.40	-0.04	-0.33	1.00						
va500	0.45	0.46	0.40	0.55	0.28	0.26	-0.15	-0.28	-0.03	0.20	0.27	0.15	0.20	0.19	-0.32	-0.31	-0.42	-0.23	0.16	1.00					
va850	0.05	0.13	-0.18	-0.36	0.23	0.27	-0.51	0.70	0.09	0.03	0.08	0.51	0.47	0.47	-0.10	-0.01	0.68	0.85	-0.42	-0.25	1.00				
vas	0.26	0.34	0.09	-0.11	0.42	0.46	-0.69	0.59	-0.02	0.25	0.33	0.69	0.67	0.66	-0.35	-0.25	0.48	0.79	-0.49	-0.08	0.95	1.00			
zg200	0.82	0.82	0.83	0.76	0.76	0.78	-0.81	-0.23	-0.45	0.90	0.99	0.76	0.80	0.81	-0.96	-0.89	-0.60	-0.12	-0.43	0.32	0.02	0.28	1.00		
zg500	0.69	0.70	0.77	0.67	0.65	0.68	-0.73	-0.10	-0.51	0.71	0.91	0.81	0.85	0.85	-0.88	-0.86	-0.53	-0.06	-0.34	0.27	0.03	0.27	0.91	1.00	
zg850	-0.66	-0.72	-0.42	-0.29	-0.76	-0.80	0.95	-0.24	0.18	-0.70	-0.70	-0.81	-0.80	-0.79	0.73	0.64	0.08	-0.35	0.55	-0.06	-0.57	-0.71	-0.67	-0.51	1.00

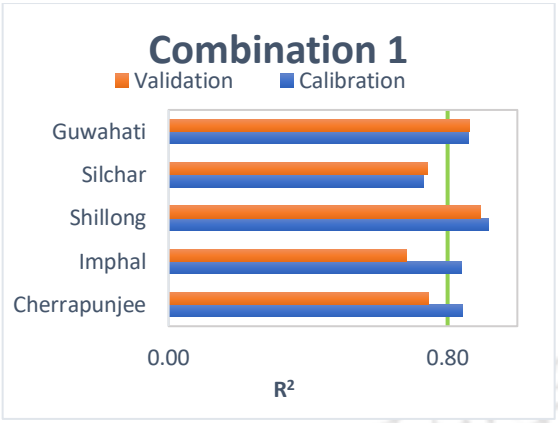
ANNEXURE E

SELECTED PREDICTORS DURING CALIBRATION-VALIDATION AND THEIR R^2 VALUES

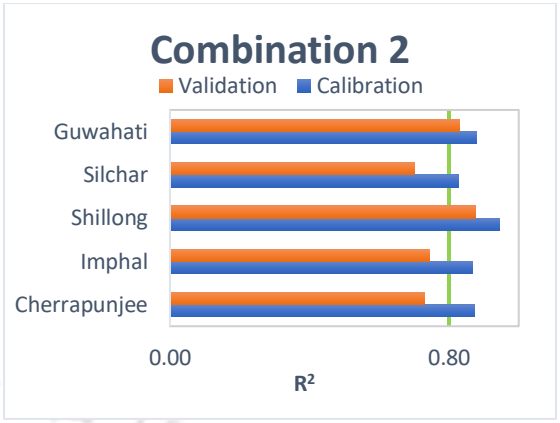
1. *Maximum Temperature*

Station		combination 1		combination 2		combination 3		combination 4		combination 5		combination 6	
		calib	valid	calib	valid	calib	valid	calib	valid	calib	valid	calib	valid
Cherrapunjee	R square	0.84	0.75	0.87	0.73	0.76	0.78	0.89	0.51	0.82	0.76	0.79	0.90
	predictors	hur200	ua500	pr		psl		hur200	ta850	hur200	zg850	hur200	zg850
		hur500	zg850	psl		ta850		pr		hur500		hur500	
		psl		zg850		zg200		prc		psl		psl	
		ta200		prc		zg850		psl		ta500		ta500	
		ta850						ta500		zg200		zg200	
Imphal	R square	0.84	0.68	0.87	0.75	0.87	0.79	0.85	0.77	0.83	0.89	0.84	0.92
	predictors	pr	zg850	prc		hur200	zg850	hur850	zg850	pr	zg850	hur200	zg850
		tas		ta500		hur500		ta500		ta500		hur500	
		ts		ta850		psl		ts		ta850		ta500	
		uas		zg500		ta500		uas		uas		ta850	
		zg500		zg850		zg200		zg500		zg500		zg200	
Shillong	R square	0.92	0.89	0.95	0.88	0.93	0.85	0.93	0.88	0.90	0.93	0.90	0.97
	predictors	hur200	ua500	hur500		psl		hur200	zg200	hur200	zg850	hur200	zg850
		hur500	zg850	psl		ta850		hur500		hur500		hur500	
		psl		zg500		zg200		psl		psl		psl	
		ta200		zg850		zg850		ta500		ta500		ta500	
		ta850						ta850		zg200		zg200	
Silchar	R square	0.73	0.74	0.83	0.70	0.71	0.82	0.87	0.65	0.78	0.77	0.76	0.82
	predictors	psl	hur500	hur500		hur500		tas		hur200		hur500	
		ua500	ta500	ta850		psl		zg850		hur500		tas	
		zg200		zg850		zg200				tas		zg200	
		zg850		zg500		zg850				psl			
		hur200		ta500									
Guwahati	R square	0.86	0.86	0.88	0.83	0.86	0.85	0.87	0.71	0.83	0.85	0.84	0.93
	predictors	psl		prc		psl		ta500		hur200	zg500	prc	
		ta850		psl		ta850		ta850		hur500		psl	
		zg500		ta500		zg500		zg850		psl		ta500	
		zg850		zg500		zg850		hur200		zg850		zg500	
				zg850				hur500		ta500		zg850	

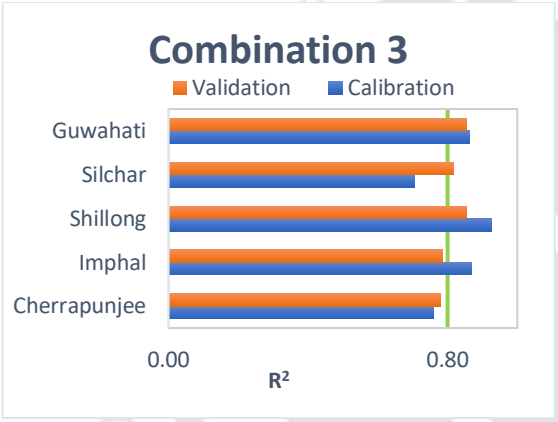
Note: calib: calibration period; valid: validation period



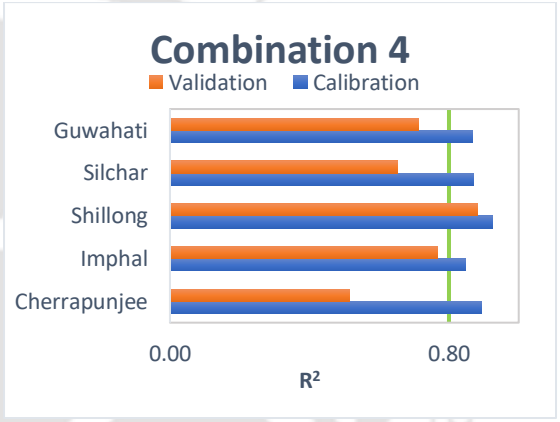
(a)



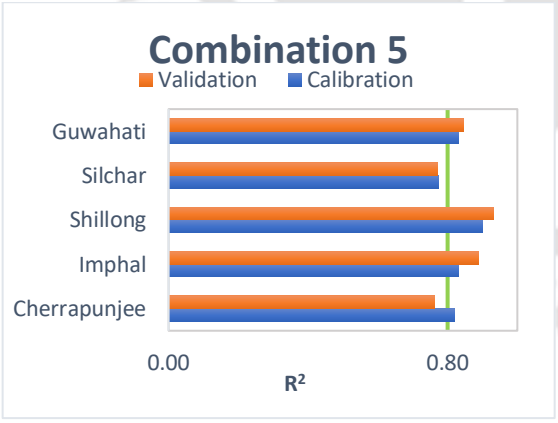
(b)



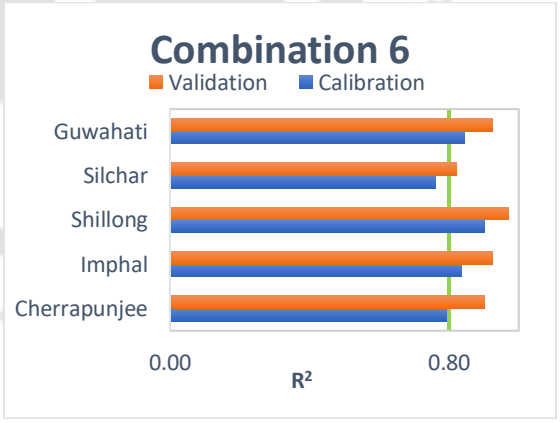
(c)



(d)



(e)



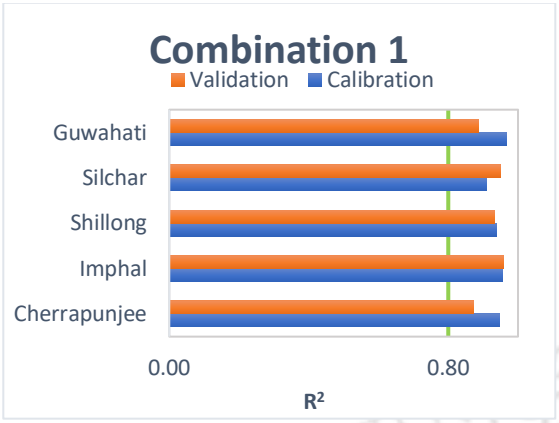
(f)

Coefficient of determination (R^2) for daily maximum temperature (T_{max})

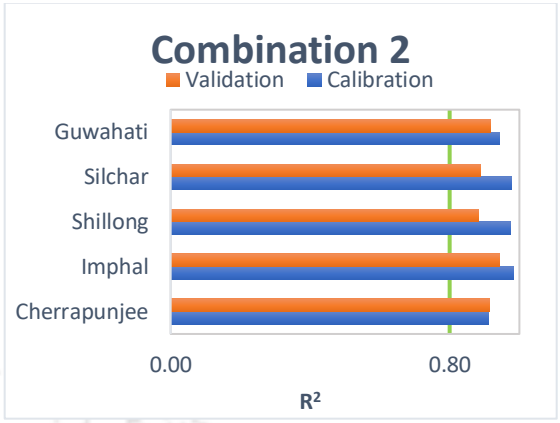
2. *Minimum Temperature*

Station		combination 1		combination 2		combination 3		combination 4		combination 5		combination 6	
		calib	valid	calib	valid	calib	valid	calib	valid	calib	valid	calib	valid
Cherrapunjee	R square	0.95	0.87	0.91	0.91	0.92	0.92	0.91	0.89	0.94	0.93	0.94	0.98
	predictors	psl		psl		pr	ta850	psl		prc		hur200	zg850
		ta850		ta850		psl		zg200		psl		hur500	
		ua500		zg850		ua200		zg850		ua500		psl	
		zg200		hur500		zg850		hur500		zg850		ta850	
		zg850				zg200						ua500	
Imphal	R square	0.96	0.96	0.98	0.94	0.96	0.94	0.97	0.95	0.96	0.96	0.96	0.98
	predictors	pr		pr		pr		pr		pr	zg200	pr	
		tas		psl		psl		psl		psl	zg500	psl	
		ts		ua200		ua500		ua200		ua500		ua500	
		hur500		zg850		zg200		zg850		zg850		zg850	
						zg850				ta500			
Shillong	R square	0.94	0.93	0.97	0.88	0.95	0.91	0.94	0.92	0.94	0.94	0.94	0.98
	predictors	psl		hur200		psl		hur850	zg500	hur500		psl	
		ta500		hur500		ta850		ta850	zg850	psl		ta850	
		ta850		ta500		ua500		ta500		ua500		ua500	
		ua500		ts		zg200		tas		zg850		zg500	
		zg850				zg850		ts		hur200		zg850	
Silchar	R square	0.91	0.95	0.98	0.89	0.94	0.91	0.95	0.91	0.95	0.94	0.93	0.98
	predictors	psl		psl		hur200		hur850		pr		pr	
		ua500		ua200		hur500		pr		psl		psl	
		zg850		zg850		psl		psl		ua500		ua500	
		pr		ta500		ua500		ua200		zg850		zg850	
				pr		zg850		zg850					
Guwahati	R square	0.97	0.89	0.94	0.92	0.93	0.94	0.97	0.90	0.96	0.96	0.95	0.98
	predictors	psl	zg850	ts		psl		hur200	zg850	pr	zg850	pr	
		ta850		zg200		ta850		hur500		psl		psl	
		ua200		zg850		zg850		psl		ta500		ta850	
		zg200		pr		ta500		ta500		ua500		ua500	
		zg500				ua200		zg500		zg500		zg850	

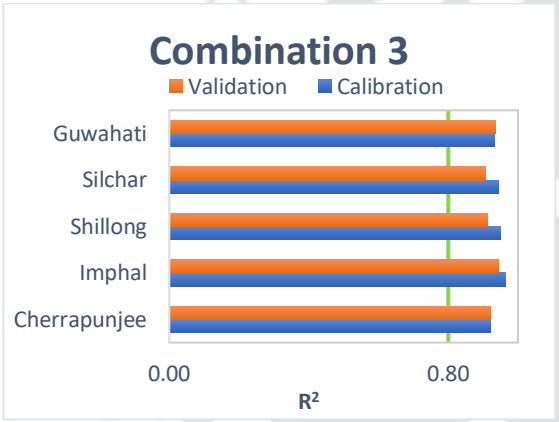
Note: calib: calibration period; valid: validation period



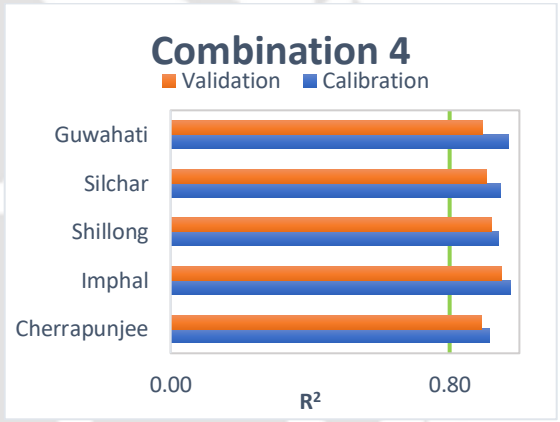
(a)



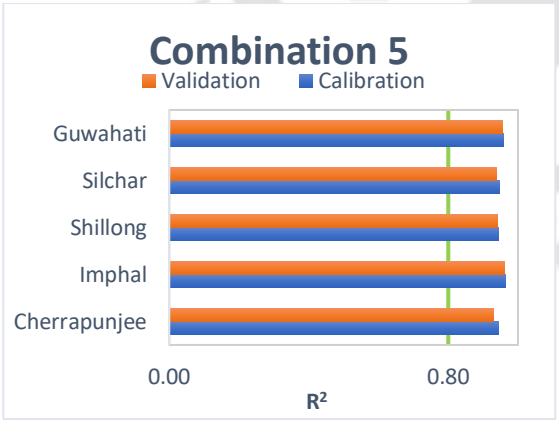
(b)



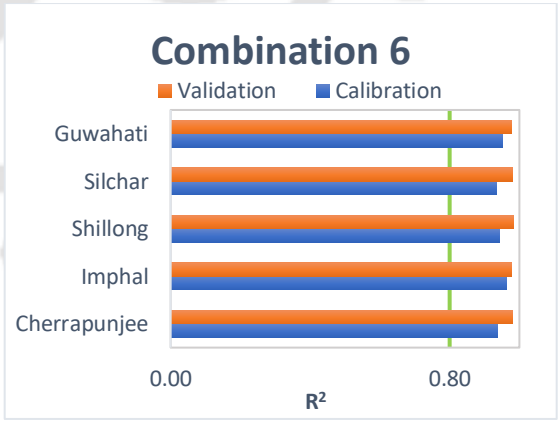
(c)



(d)



(e)



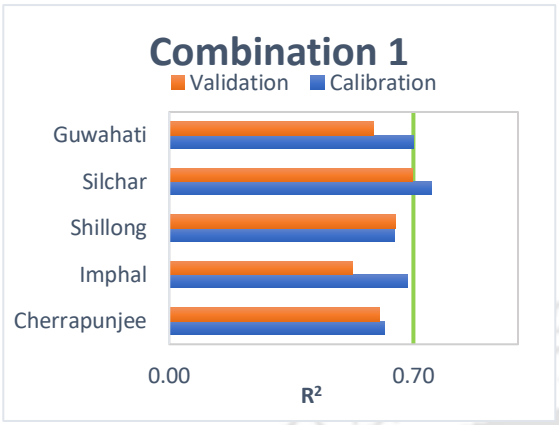
(f)

Coefficient of determination (R^2) for daily minimum temperature (T_{min})

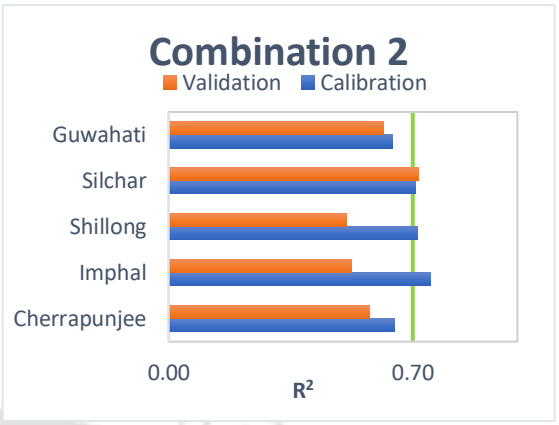
3. Annual rainfall

Station		combination 1		combination 2		combination 3		combination 4		combination 5		combination 6	
		calib	**valid**	**calib**	**valid**	**calib**	**valid**	**calib**	**valid**	**calib**	**valid**	**calib**	**valid**
Cherrapunjee	**R square**	0.62	0.60	0.65	0.58	0.63	0.53	0.70	0.42	0.64	0.55	0.63	0.70
	predictors	**hur200**		**psl**		hur200		**hur500**		**hur200**		**hur200**	
		ta200		ta500		hur500		psl		**psl**		**psl**	
		ts		zg200		**psl**		**ta200**		ta200		tas	
		zg200		prc		ta200		zg200		zg200		ts	
						ta500							
Imphal	**R square**	0.68	0.52	0.75	0.53	0.73	0.60	0.70	0.47	0.66	0.64	0.64	0.74
	predictors	**hur850**		**pr**		**pr**		**pr**		**hur850**		**pr**	
		zg200		**prc**		**prc**		**prc**		ta500		**prc**	
		zg500		**ua500**		**ua500**		**ta500**		zg500		**zg500**	
		zg850		**zg850**		**uas**		**ua200**		**zg850**		**zg850**	
						hur500							
Shillong	**R square**	0.64	0.65	0.71	0.51	0.72	0.52	0.73	0.34	0.68	0.60	0.67	0.62
	predictors	**hur500**		**hur500**		**hur500**		**pr**		**hur500**		**ta200**	
		ta200		**hur850**		**pr**		**prc**		**ta200**		**ta850**	
		ta850		**ta850**		**ua200**		**ua200**		**ta850**		hur850	
		zg850		**tas**		ts		**zg500**					
				zg850		hur200							
Silchar	**R square**	0.75	0.70	0.71	0.72	0.68	0.76	0.86	0.64	0.74	0.71	0.72	0.79
	predictors	**hur200**		pr		ta500		**prc**		hur200		**prc**	
		ta500		ta200		**zg500**		ta200		ta500		**zg500**	
		zg500		**ua200**		**zg850**		ua200		**zg500**		**zg850**	
		zg850		ta500				**zg850**		**zg850**		**ta500**	
Guwahati	**R square**	0.70	0.59	0.64	0.62	0.79	0.52	0.68	0.58	0.68	0.67	0.71	0.74
	predictors	**ta500**		**tas**		**psl**		**hur500**		hur500		psl	**zg850**
		zg500		**zg200**		**ta500**		psl		psl		ta500	
		zg850		ta500		**zg500**		tas		tas		tas	
		hur500				**zg850**		zg500		**zg500**		zg200	
								zg850		**zg850**		**zg500**	

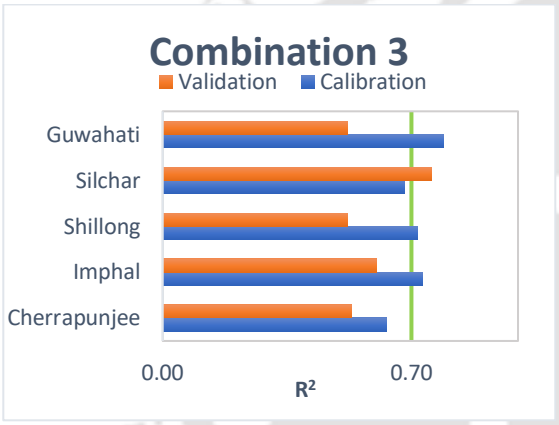
Note: calib: calibration period; valid: validation period



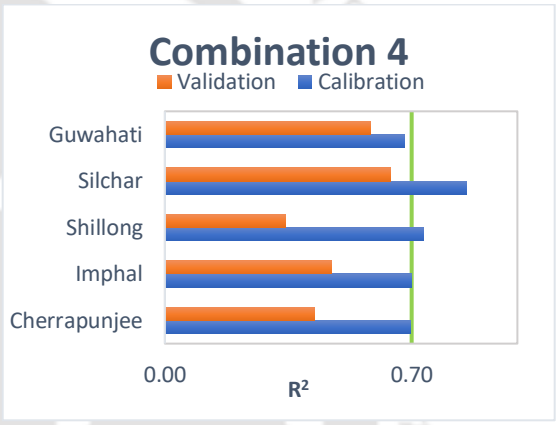
(a)



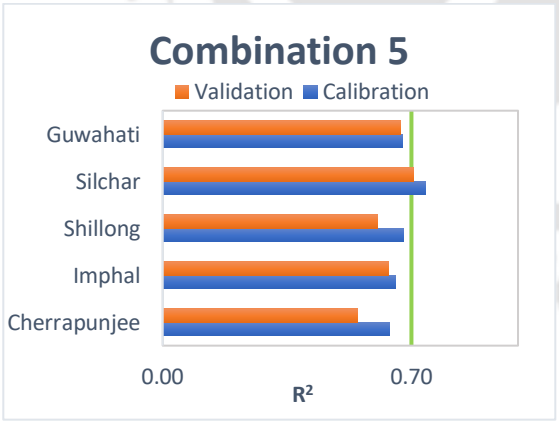
(b)



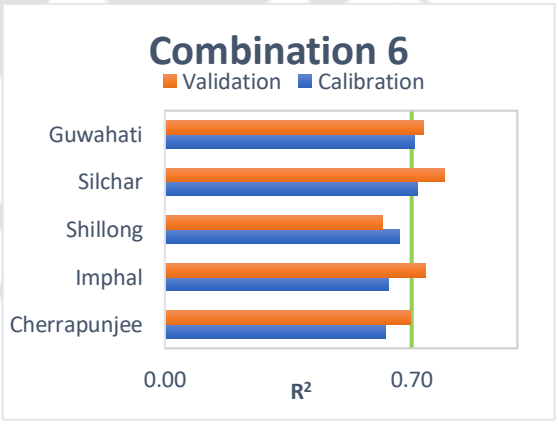
(c)



(d)



(e)



(f)

Coefficient of determination (R^2) for annual rainfall

ANNEXURE F

STATISTICAL PARAMETERS OF THE COMBINATION SETS

1. Maximum Temperature

(a) CHERRAPUNJEE							
Statistical parameters		Combinations					
		1	2	3	4	5	6
calib	mean (μ)	21.52	22.32	21.94	21.90	21.68	21.85
	stdev (б)	2.58	2.240091	2.41	2.48	2.53	2.46
	max	24.78	25.39	25.39	24.63	24.78	25.39
	min	16.28	17.05	16.28	16.68	16.28	16.28
valid	mean (μ)	22.32	21.52	21.90	21.94	22.39	22.28
	stdev (б)	2.24	2.57639	2.48	2.41	2.18	2.36
	max	25.39	24.78	24.63	25.39	25.39	24.56
	min	17.05	16.28	16.68	16.28	17.05	17.05
	Pearson	0.978666	0.978666	0.977174	0.977174	0.980632	0.963216
	R^2	0.9578	0.9578	0.9549	0.9549	0.9616	0.9278
	RMSE	0.969979	0.969979	0.502007	0.502007	0.89295	0.752846
	RE_μ	3.675828	3.545501	0.140925	0.141123	3.278605	1.965181
	RE_б	13.05313	15.01277	2.68977	2.619316	13.81907	3.755565
	%age ch in max	2.447279	2.388818	2.982211	3.073881	2.447279	3.240152
	%age ch in min	4.675119	4.466313	2.416799	2.359768	4.675119	4.675119

(b) IMPHAL							
Statistical parameters		Combinations					
		1	2	3	4	5	6
calib	mean (μ)	27.94784	28.43053	28.14572	28.23265	28.2027	28.21352
	stdev (б)	2.766329	2.461719	2.584234	2.674006	2.71408	2.620459
	max	31.50333	31.61935	31.50333	31.61935	31.61935	31.61935
	min	22.16452	22.5129	22.16452	22.5129	22.16452	22.16452
valid	mean (μ)	28.43053	27.94784	28.23265	28.14572	28.16214	28.06752
	stdev (б)	2.461719	2.766329	2.674006	2.584234	2.448575	2.67618
	max	31.61935	31.50333	31.61935	31.50333	30.97333	30.97333
	min	22.5129	22.16452	22.5129	22.16452	22.5129	22.5129
	Pearson	0.95312	0.95312	0.984686	0.984686	0.977879	0.950201
	R^2	0.9084	0.9084	0.9696	0.9696	0.9562	0.9029
	RMSE	0.947249	0.947249	0.444444	0.444444	0.554167	0.811734
	RE_μ	1.727125	1.697801	0.308877	0.307926	0.14381	0.517473
	RE_б	11.01133	12.37385	3.473802	3.35718	9.782495	2.126365

	%age ch in max	0.368283	0.366932	0.368283	0.366932	2.04312	2.04312
	%age ch in min	1.571824	1.5475	1.571824	1.5475	1.571824	1.571824

(c) SHILLONG							
Statistical parameters		Combinations					
		1	2	3	4	5	6
calib	mean (μ)	21.09015	21.68278	21.48308	21.28985	21.26481	21.39099
	stdev (б)	3.303821	3.144698	3.18311	3.291441	3.286621	3.239077
	max	24.90645	25.01	25.01	24.80645	24.90645	25.01
	min	14.95484	15.04516	15.05806	14.95484	14.95484	14.95484
valid	mean (μ)	21.68278	21.09015	21.28985	21.48308	21.62979	21.36385
	stdev (б)	3.144698	3.303821	3.291441	3.18311	3.125316	3.239831
	max	25.01	24.90645	24.80645	25.01	25.01	24.66452
	min	15.04516	14.95484	14.95484	15.05806	15.04516	15.04516
	Pearson	0.986222	0.986222	0.99107	0.99107	0.991084	0.979805
	R^2	0.9726	0.9726	0.9822	0.9822	0.9822	0.96
	RMSE	0.796988	0.796988	0.475303	0.475303	0.573066	0.624624
	RE_μ	2.809968	2.733167	0.899425	0.907588	1.716377	0.126855
	RE_б	4.816341	5.06005	3.40331	3.291297	4.907928	0.023277
	%age ch in max	0.415749	0.414028	0.813868	0.820546	0.415749	1.381383
	%age ch in min	0.603969	0.600343	0.685518	0.69025	0.603969	0.603969

(d) SILCHAR							
Statistical parameters		Combinations					
		1	2	3	4	5	6
calib	mean (μ)	30.45658	30.63575	30.48781	30.60452	30.6483	30.5453
	stdev (б)	2.847682	2.926305	2.777828	2.994282	2.890704	2.807536
	max	34.56129	34.21333	34.56129	34.21333	34.56129	34.56129
	min	23.57742	23.12581	23.57742	23.12581	23.57742	23.57742
valid	mean (μ)	30.63575	30.45658	30.60452	30.48781	30.34189	30.55051
	stdev (б)	2.926305	2.847682	2.994282	2.777828	2.873227	3.289981
	max	34.21333	34.56129	34.21333	34.56129	33.99333	33.99333
	min	23.12581	23.57742	23.12581	23.57742	23.12581	23.12581
	Pearson	0.961948	0.961948	0.979587	0.979587	0.971032	0.941859
	R^2	0.9253	0.9253	0.9596	0.9596	0.9429	0.9278
	RMSE	0.779591	0.779591	0.635226	0.635226	0.717738	1.120785
	RE_μ	0.588288	0.584847	0.382839	0.381379	0.999793	0.01707
	RE_б	2.760961	2.68678	7.792192	7.228902	0.604609	17.18393
	%age ch in max	1.006782	1.017022	1.006782	1.017022	1.643333	1.643333

	%age ch in min	1.915447	1.952853	1.915447	1.952853	1.915447	1.915447

(e) GUWAHATI							
Statistical parameters		Combinations					
		1	2	3	4	5	6
calib	mean (μ)	30.3679	30.58875	30.32638	30.63026	30.56083	30.50434
	stdev (б)	3.273858	3.217821	3.139057	3.346107	3.259765	3.170529
	max	34.12258	34.28667	34.12258	34.28667	34.28667	34.28667
	min	23.29677	22.65161	23.29677	22.65161	23.29677	23.29677
valid	mean (μ)	30.58875	30.3679	30.63026	30.32638	30.3133	30.34824
	stdev (б)	3.217821	3.273858	3.346107	3.139057	3.216874	3.631796
	max	34.28667	34.12258	34.28667	34.12258	33.49032	33.35161
	min	22.65161	23.29677	22.65161	23.29677	22.65161	22.65161
	Pearson	0.971139	0.971139	0.989603	0.989603	0.988667	0.979604
	R^2	0.9431	0.9431	0.9793	0.9793	0.9775	0.9596
	RMSE	0.783421	0.783421	0.573905	0.573905	0.529643	0.806091
	RE_μ	0.727264	0.722013	1.002034	0.992093	0.809957	0.511713
	RE_б	1.711649	1.741457	6.59593	6.187788	1.315797	14.54857
	%age ch in max	0.480872	0.478571	0.480872	0.478571	2.322606	2.727164
	%age ch in min	2.769316	2.848191	2.769316	2.848191	2.769316	2.769316

2. *Minimum Temperature*

(a) CHERRAPUNJEE							
Statistical parameters		Combinations					
		1	2	3	4	5	6
calib	mean (μ)	14.39	13.97	14.36	14.00	14.49	14.43
	stdev (б)	4.10	4.295763	4.03	4.37	4.07	4.07
	max	19.23	19.16	19.23	19.16	19.23	19.23
	min	6.83	4.69	7.30	4.69	6.83	6.83
valid	mean (μ)	13.97	14.39	14.00	14.36	13.56	12.92
	stdev (б)	4.30	4.103868	4.37	4.03	4.41	4.66
	max	19.16	19.23	19.16	19.23	18.83	18.06
	min	4.69	6.83	4.69	7.30	4.69	4.69
	Pearson	0.991552	0.991552	0.996638	0.996638	0.993733	0.993339
	R^2	0.9832	0.9832	0.9933	0.9933	0.9875	0.9867
	RMSE	0.685588	0.685588	0.567638	0.567638	1.070043	1.65426
	RE_μ	2.876151	2.961324	2.490969	2.554603	6.42035	10.46367
	RE_б	4.67594	4.467063	8.408096	7.755967	8.294104	14.41413

	%age ch in max	0.342282	0.343457	0.342282	0.343457	2.080537	6.052224
	%age ch in min	31.23819	45.42955	35.73322	55.60137	31.23819	31.23819

(b) IMPHAL							
Statistical parameters		Combinations					
		1	2	3	4	5	6
calib	mean (μ)	15.29238	15.59756	15.51196	15.37798	15.37264	15.47589
	stdev (б)	6.219806	6.46847	6.317162	6.376458	6.262164	6.286712
	max	22.49032	23.05806	23.05806	23.03667	23.03667	23.05806
	min	4.012903	5.167742	4.7	4.012903	4.012903	4.012903
valid	mean (μ)	15.59756	15.29238	15.37798	15.51196	15.58963	15.29035
	stdev (б)	6.46847	6.219806	6.376458	6.317162	6.515134	6.66018
	max	23.05806	22.49032	23.03667	23.05806	23.05806	22.7129
	min	5.167742	4.012903	4.012903	4.7	5.167742	5.225806
	Pearson	0.995007	0.995007	0.997811	0.997811	0.996294	0.99486
	R^2	0.99	0.99	0.9956	0.9956	0.9926	0.9897
	RMSE	0.744657	0.744657	0.439653	0.439653	0.614838	0.690563
	RE_μ	1.995689	1.956641	0.863693	0.871217	1.41148	1.198889
	RE_б	3.997939	3.844248	0.938635	0.929907	4.039666	5.940594
	%age ch in max	2.524383	2.462227	0.0928	0.092886	0.092886	1.496922
	%age ch in min	28.77814	22.34707	14.61908	17.12219	28.77814	30.22508

(c) SHILLONG							
Statistical parameters		Combinations					
		1	2	3	4	5	6
calib	mean (μ)	13.08832	12.99384	13.10279	12.97937	13.04765	13.0288
	stdev (б)	4.309118	4.166615	4.094546	4.377285	4.236325	4.173033
	max	18.48065	18.09677	18.48065	18.09677	18.48065	18.48065
	min	5.470968	4.9	5.8	4.9	5.470968	5.470968
valid	mean (μ)	12.99384	13.08832	12.97937	13.10279	13.02794	13.10249
	stdev (б)	4.166615	4.309118	4.377285	4.094546	4.243628	4.57453
	max	18.09677	18.48065	18.09677	18.48065	18.09677	18.09677
	min	4.9	5.470968	4.9	5.8	4.9	4.9
	Pearson	0.996417	0.996417	0.996767	0.996767	0.993492	0.993019
	R^2	0.9928	0.9928	0.9935	0.9935	0.987	0.9861
	RMSE	0.382343	0.382343	0.468891	0.468891	0.471689	0.587501
	RE_μ	0.721842	0.72709	0.941899	0.950855	0.1511	0.565587
	RE_б	3.306998	3.420101	6.905247	6.459222	0.1724	9.621238
	%age ch in max	2.077151	2.121212	2.077151	2.121212	2.077151	2.077151

	%age ch in min	10.43632	11.6524	15.51724	18.36735	10.43632	10.43632

(d) SILCHAR							
Statistical parameters		Combinations					
		1	2	3	4	5	6
calib	mean (μ)	20.54518	20.23946	20.54661	20.23803	20.45714	20.43847
	stdev (б)	4.868057	5.174101	5.01554	5.031178	4.864495	4.982973
	max	26.02903	26.34483	26.34483	25.94194	26.02903	26.34483
	min	11.40645	10.81071	10.81071	11.40645	11.40645	10.81071
valid	mean (μ)	20.23946	20.54518	20.23803	20.54661	20.26269	20.16161
	stdev (б)	5.174101	4.868057	5.031178	5.01554	5.337836	5.24292
	max	26.34483	26.02903	25.94194	26.34483	26.34483	25.57667
	min	10.81071	11.40645	11.40645	10.81071	10.81071	11.73871
	Pearson	0.993527	0.993527	0.99348	0.99348	0.992639	0.993498
	R^2	0.9871	0.9871	0.987	0.987	0.9853	0.9867
	RMSE	0.71081	0.71081	0.642427	0.642427	0.76626	0.6456
	RE_μ	1.488047	1.510524	1.501857	1.524756	0.950515	1.354602
	RE_б	6.286785	5.914927	0.311779	0.31081	9.73054	5.216723
	%age ch in max	1.213243	1.1987	1.529302	1.553053	1.213243	2.915794
	%age ch in min	5.22281	5.510619	5.510619	5.22281	5.22281	8.584034

(e) GUWAHATI							
Statistical parameters		Combinations					
		1	2	3	4	5	6
calib	mean (μ)	20.20911	20.04707	20.21941	20.03677	20.27431	20.26552
	stdev (б)	5.471229	5.527049	5.446615	5.550976	5.381213	5.423731
	max	26.80645	26.64	26.80645	26.50968	26.80645	26.80645
	min	9.877419	9.916129	10.79677	9.877419	9.877419	9.877419
valid	mean (μ)	20.04707	20.20911	20.03677	20.21941	19.83566	19.44096
	stdev (б)	5.527049	5.471229	5.550976	5.446615	5.722894	5.838512
	max	26.64	26.80645	26.50968	26.80645	26.64	25.63548
	min	9.916129	9.877419	9.877419	10.79677	9.916129	9.916129
	Pearson	0.99711	0.99711	0.997033	0.997033	0.99716	0.994387
	R^2	0.9942	0.9942	0.9941	0.9941	0.9943	0.9888
	RMSE	0.444339	0.444339	0.46645	0.46645	0.669112	1.039671
	RE_μ	0.801804	0.808285	0.903284	0.911517	2.163564	4.068763
	RE_б	1.020246	1.009942	1.916067	1.880044	6.349523	7.647534
	%age ch in max	0.620939	0.624818	1.1071	1.119494	0.620939	4.368231
	%age ch in min	0.391901	0.390371	8.515088	9.307642	0.391901	0.391901

3. Annual rainfall

(a) CHERRAPUNJEE							
Statistical parameters		**Combinations**					
		1	2	3	4	5	6
calib	mean (μ)	922.9731	866.2963	942.7998	846.4697	883.4525	930.328
	stdev (б)	1113.632	939.7887	1065.926	991.9985	1061.949	1070.125
	max	4138.8	3340.3	3615.8	4138.8	4138.8	4138.8
	min	0	0	0	0	0	0
valid	mean (μ)	866.2963	922.9731	846.4697	942.7998	916.9992	716.1685
	stdev (б)	939.7887	1113.632	991.9985	1065.926	963.5656	759.935
	max	3340.3	4138.8	4138.8	3615.8	3340.3	1853.4
	min	0	0	0	0	0	0
	Pearson	0.914591	0.914591	0.919798	0.919798	0.839027	0.752489
	R^2	0.8365	0.8365	0.846	0.846	0.704	0.5662
	RMSE	400.527	400.527	367.0508	367.0508	503.3617	639.0651
	RE_μ	6.140678	6.542428	10.21745	11.38022	3.797228	23.01979
	RE_б	15.61048	18.49813	6.935488	7.452344	9.264372	28.98636
	%age ch in max	19.29303	23.90504	14.4643	12.63651	19.29303	55.2189

(b) IMPHAL							
Statistical parameters		**Combinations**					
		1	2	3	4	5	6
calib	mean (μ)	106.6544	110.4515	109.3741	107.7318	101.1252	107.1086
	stdev (б)	86.59804	103.2602	91.82499	98.67161	83.20891	87.51874
	max	256.7	333	300.9	333	256.7	300.9
	min	0	0	0	0	0	0
valid	mean (μ)	110.4515	106.6544	107.7318	109.3741	123.4084	115.7746
	stdev (б)	103.2602	86.59804	98.67161	91.82499	114.6954	129.0744
	max	333	256.7	333	300.9	333	333
	min	0	0	0	0	0	0
	Pearson	0.907715	0.907715	0.89842	0.89842	0.923792	0.833269
	R^2	0.8239	0.8239	0.8072	0.8072	0.8534	0.6943
	RMSE	36.45956	36.45956	35.92699	35.92699	50.82848	75.73198
	RE_μ	3.560124	3.437737	1.501477	1.524365	22.0353	8.090893
	RE_б	19.24079	16.13608	7.456166	6.938797	37.84031	47.48206
	%age ch in max	29.72341	22.91291	10.668	9.63964	29.72341	10.668

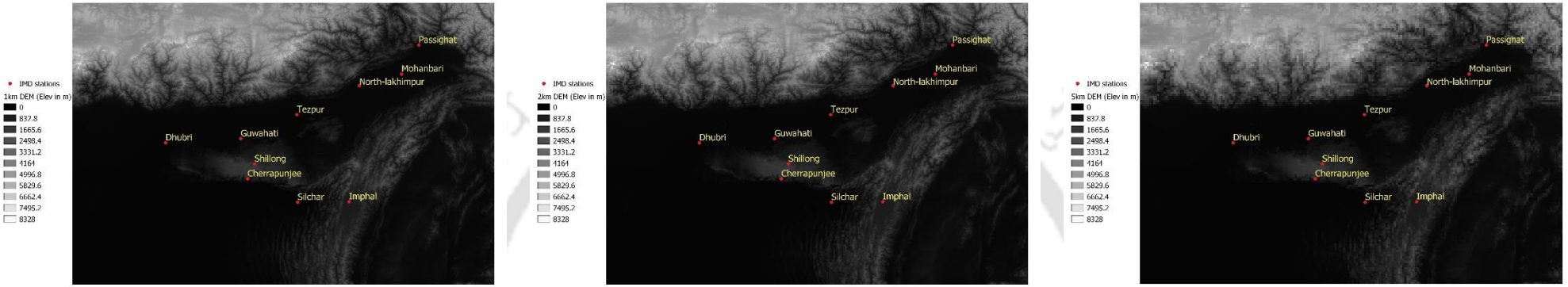
(c) SHILLONG							
Statistical parameters		Combinations					
		1	2	3	4	5	6
calib	mean (μ)	182.0706	173.3731	158.4962	196.9475	178.7009	178.081
	stdev (б)	186.8106	187.2032	155.3582	212.3337	183.2119	181.9337
	max	763.4	749.8	603	763.4	763.4	763.4
	min	0	0	0	0	0	0
valid	mean (μ)	173.3731	182.0706	196.9475	158.4962	175.7638	175.9259
	stdev (б)	187.2032	186.8106	212.3337	155.3582	194.6689	212.4399
	max	749.8	763.4	763.4	603	749.8	749.8
	min	0	0	0	0	0	2.5
	Pearson	0.920455	0.920455	0.937254	0.937254	0.861612	0.847461
	R^2	0.8472	0.8472	0.8784	0.8784	0.7424	0.7182
	RMSE	68.56102	68.56102	81.34169	81.34169	95.3889	109.0526
	RE_μ	4.777032	5.01668	24.26012	19.52366	1.643599	1.210178
	RE_б	0.210176	0.209735	36.67367	26.83302	6.253409	16.76777
	%age ch in max	1.781504	1.813817	26.60033	21.01127	1.781504	1.781504

(d) SILCHAR							
Statistical parameters		Combinations					
		1	2	3	4	5	6
calib	mean (μ)	249.2094	275.2164	278.8462	245.5796	245.0076	265.9517
	stdev (б)	219.3487	254.662	258.8146	213.9165	218.0137	238.0751
	max	838.7	868.5	868.5	665.8	838.7	868.5
	min	0	0	0	0	0	0
valid	mean (μ)	275.2164	249.2094	245.5796	278.8462	296.6235	243.5191
	stdev (б)	254.662	219.3487	213.9165	258.8146	271.0261	236.8094
	max	868.5	838.7	665.8	868.5	868.5	580.9
	min	0	0	0	0	0	0
	Pearson	0.957542	0.957542	0.973564	0.973564	0.906198	0.896935
	R^2	0.9169	0.9169	0.9478	0.9478	0.8212	0.5662
	RMSE	74.87367	74.87367	62.51828	62.51828	121.4935	102.8189
	RE_μ	10.43579	9.449642	11.9301	13.54618	21.06704	8.43482
	RE_б	16.09918	13.86675	17.34759	20.98861	24.31613	0.531673
	%age ch in max	3.553118	3.431203	23.33909	30.44458	3.553118	33.11457

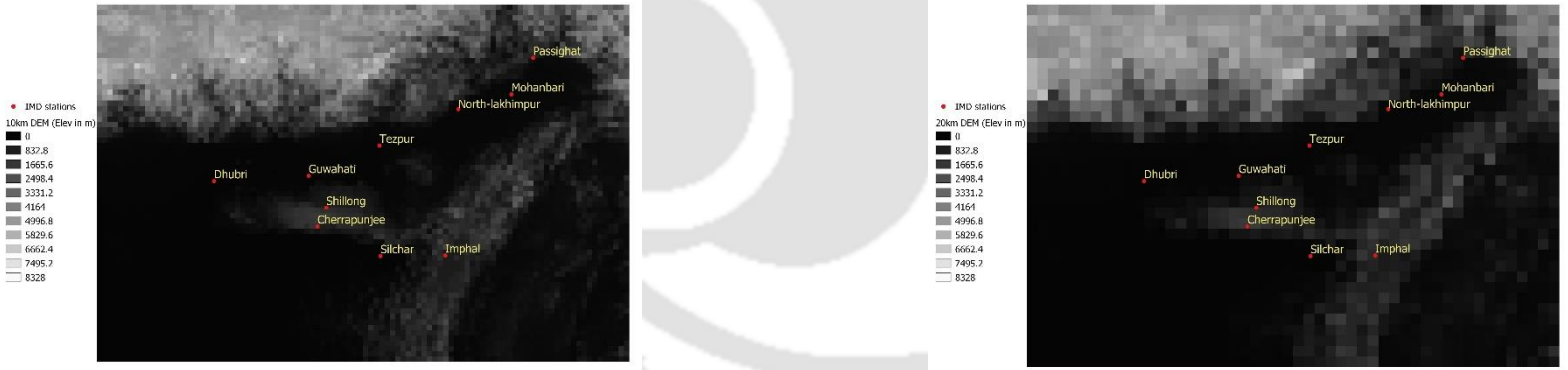
(e) GUWAHATI							
Statistical parameters		Combinations					
		1	2	3	4	5	6
calib	mean (μ)	127.5751	130.0488	137.1269	120.497	126.6343	135.0268
	stdev (б)	113.589	134.4721	130.1782	117.8973	116.9998	127.0304
	max	326.3876	468.6	468.6	388.3	388.3	468.6
	min	0	0	0	0	0	0
valid	mean (μ)	130.0488	127.5751	120.497	137.1269	133.1673	97.73762
	stdev (б)	134.4721	113.589	117.8973	130.1782	138.3959	103.8443
	max	468.6	326.3876	388.3	468.6	468.6	264.2514
	min	0	0	0	0	0	0
	Pearson	0.940255	0.940255	0.821718	0.821718	0.95033	0.8019
	R^2	0.8841	0.8841	0.6752	0.6752	0.9031	0.643
	RMSE	38.28547	38.28547	68.79718	68.79718	39.84875	75.35181
	RE_μ	1.938954	1.902073	12.1274	13.80112	5.15898	27.61614
	RE_б	18.38479	15.52969	9.433896	10.41659	18.28738	18.25242
	%age ch in max	43.57163	30.34836	17.13615	20.67989	20.67989	43.60832

ANNEXURE G

RASTER MAPS OF TERRAIN ANALYSIS

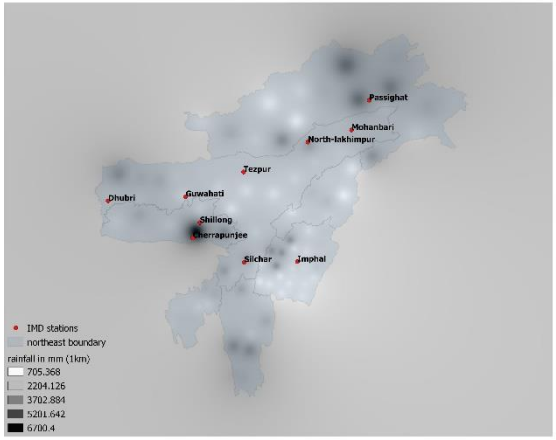


(a) (b) (c)

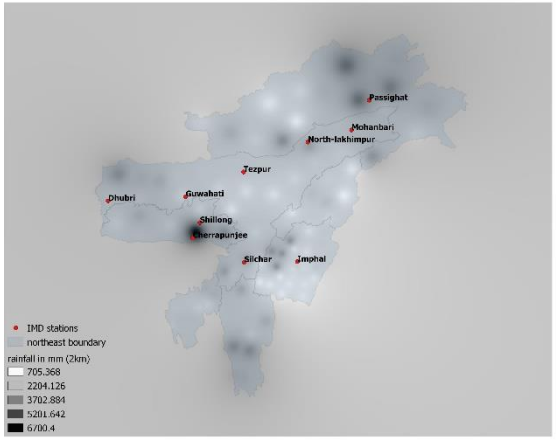


(d) (e)

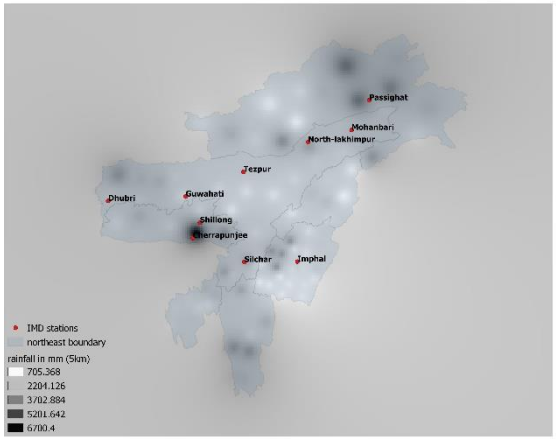
DEM of various resolutions covering the northeast Indian region



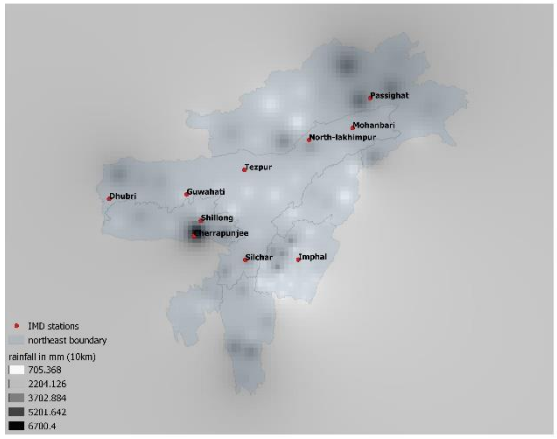
(a)



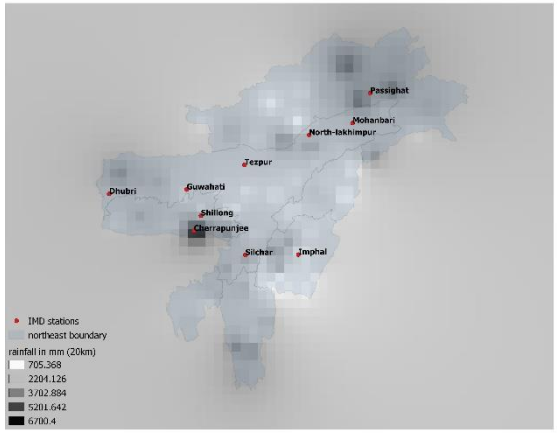
(b)



(c)

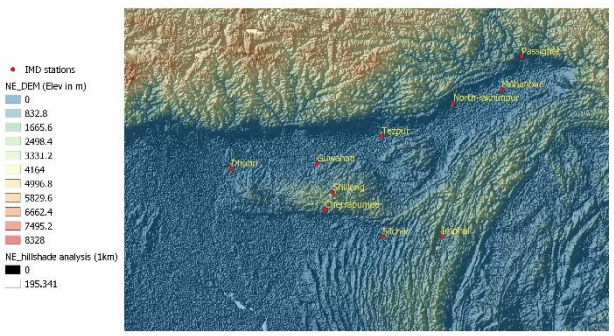


(d)

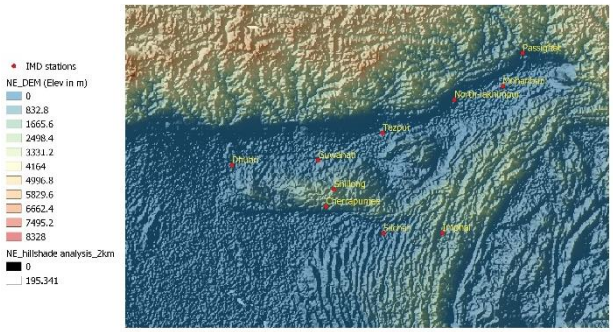


(e)

Raster map of rainfall of various resolutions covering the northeast Indian region

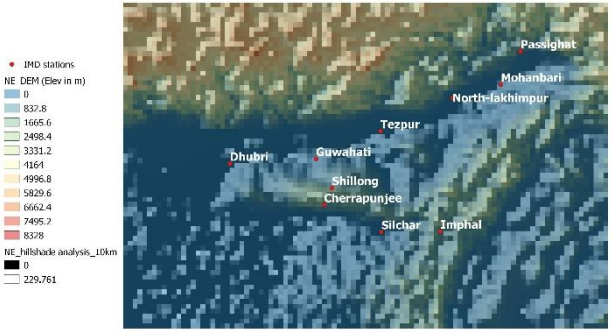


(a)

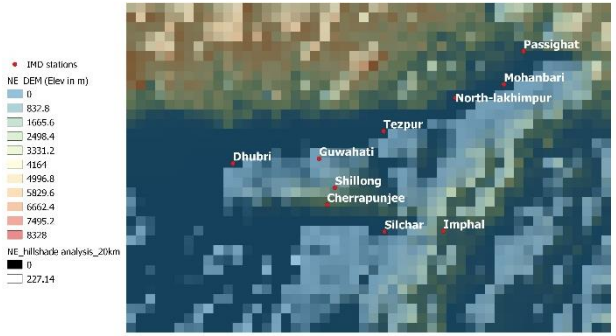


(b)

(c)

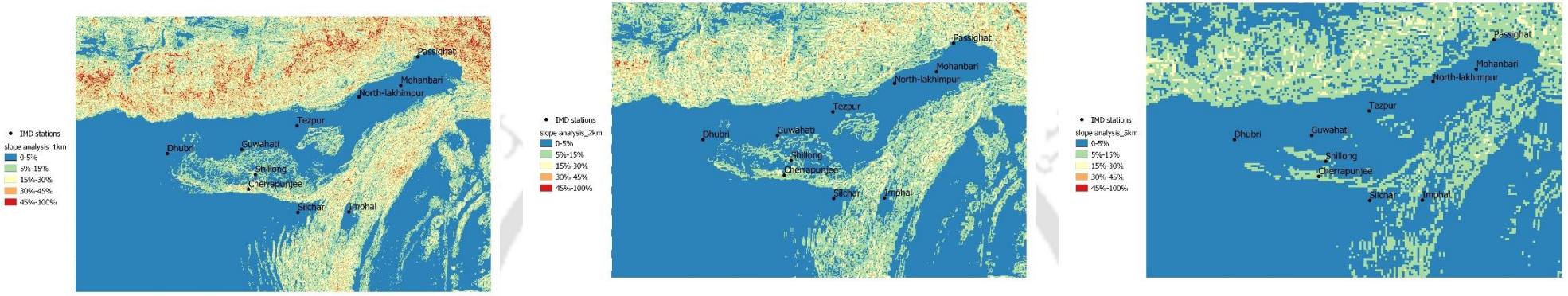


(d)

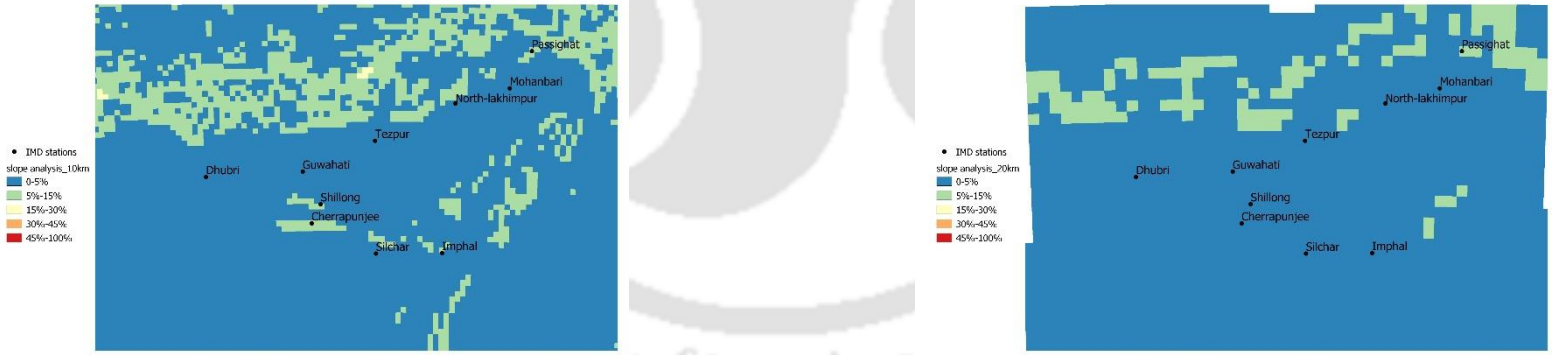


(e)

Raster map of hillshade analysis of various resolutions covering the northeast Indian region

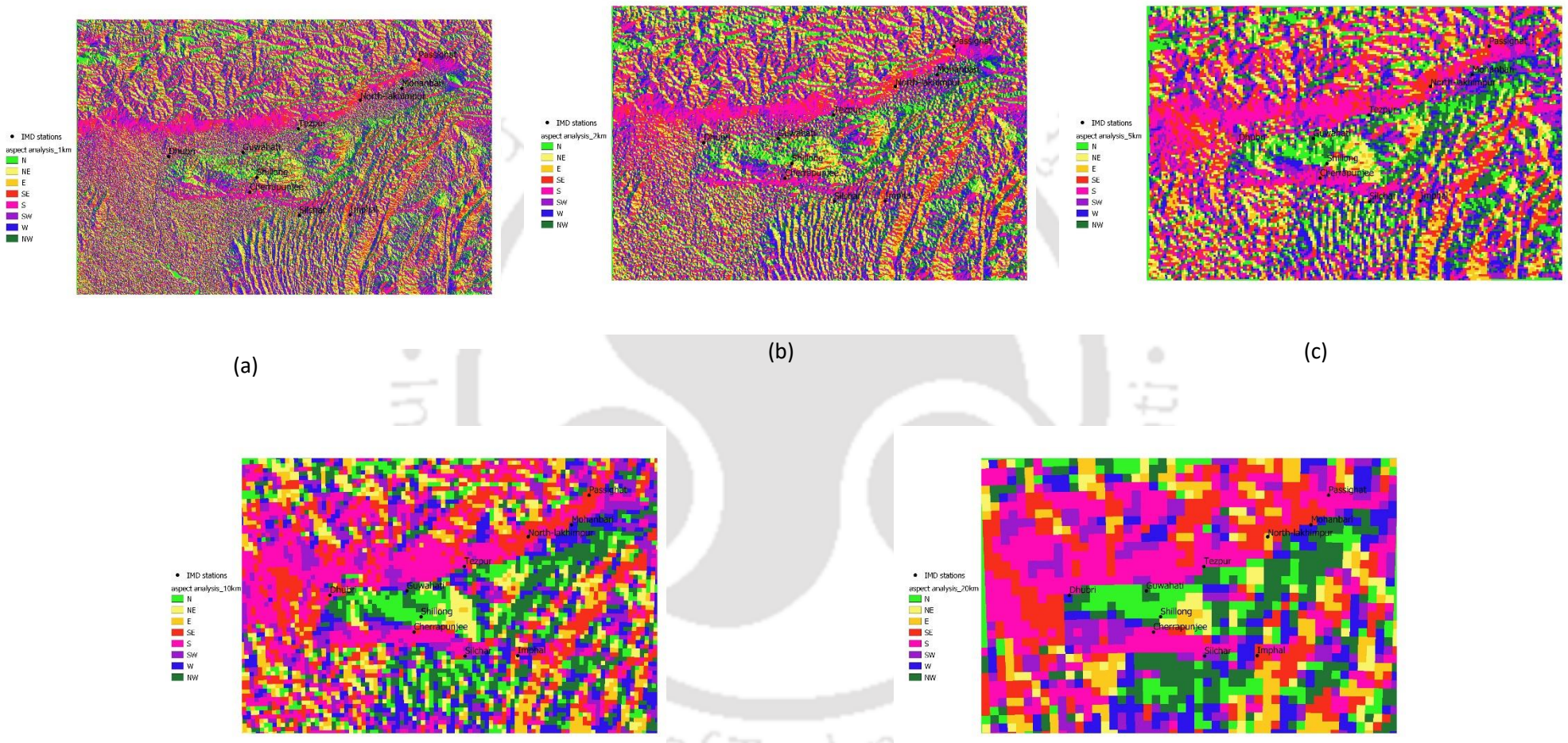


(a) (b) (c)



(d) (e)

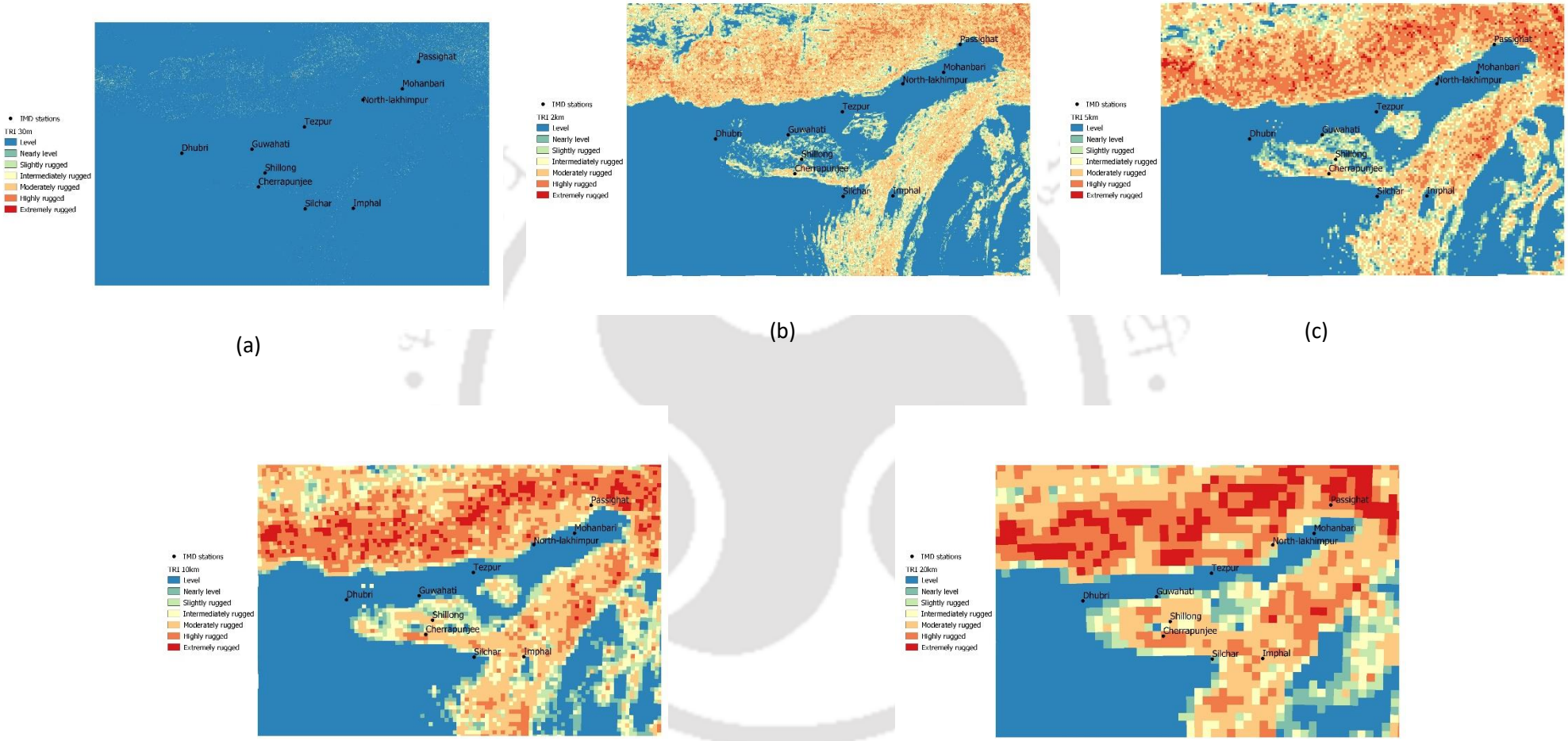
Raster map of slope analysis of various resolutions covering the northeast Indian region



(a) (b) (c)

(d) (e)

Raster map of aspect analysis of various resolutions covering the northeast Indian region



Raster map of TRI analysis of various resolutions covering the northeast Indian region

References

Abida, H., Ellouze, M., 2006. Hydrological Delineation of Homogeneous Regions in Tunisia. Water Resour Manage 20, 961–977. https://doi.org/10.1007/s11269-006-9017-3

Adelekan, I.O., 1998. Spatio-temporal variations in thunderstorm rainfall over Nigeria. International Journal of Climatology 18, 1273–1284. https://doi.org/10.1002/(SICI)1097-0088(199809)18:11<1273::AID-JOC298>3.0.CO;2-4

Agarwal, A., Maheswaran, R., Sehgal, V., Khosa, R., Sivakumar, B., Bernhofer, C., 2016. Hydrologic regionalization using wavelet-based multiscale entropy method. Journal of Hydrology 538, 22–32. https://doi.org/10.1016/j.jhydrol.2016.03.023

Aguilera, H., Murillo, J.M., 2009. The effect of possible climate change on natural groundwater recharge based on a simple model: a study of four karstic aquifers in SE Spain. Environ Geol 57, 963–974. https://doi.org/10.1007/s00254-008-1381-2

Anandhi, A., Srinivas, V.V., Nanjundiah, R.S., Nagesh Kumar, D., 2008. Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. Int. J. Climatol. 28, 401–420. https://doi.org/10.1002/joc.1529

Anghileri, D., Pianosi, F., Soncini-Sessa, R., 2014. Trend detection in seasonal data: from hydrology to water resources. Journal of Hydrology 511, 171–179. https://doi.org/10.1016/j.jhydrol.2014.01.022

Asong, Z.E., Khaliq, M.N., Wheater, H.S., 2015. Regionalization of precipitation characteristics in the Canadian Prairie Provinces using large-scale atmospheric covariates and geophysical attributes. Stoch Environ Res Risk Assess 29, 875–892. https://doi.org/10.1007/s00477-014-0918-z

Bärring, L., 1987. Spatial patterns of daily rainfall in central Kenya: Application of principal component analysis, common factor analysis and spatial correlation. J. Climatol. 7, 267–289. https://doi.org/10.1002/joc.3370070306

Basistha, A., Arya, D.S., Goel, N.K., 2009. Analysis of historical changes in rainfall in the Indian Himalayas. Int. J. Climatol. 29, 555–572. https://doi.org/10.1002/joc.1706

Beable, M.E., McKerchar, A.I., 1982. Regional Flood Estimation in New Zealand (Technical publication No. 20), Water & soil technical publication. National water and Soil conservation organisation, Wellington, New Zealand.

Bedi, H.S., Bindra, M.M.S., 1980. Principal components of monsoon rainfall. Tellus 32, 296–298. https://doi.org/10.3402/tellusa.v32i3.10584

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Springer US, Boston, MA. https://doi.org/10.1007/978-1-4757-0450-1

Bezdek, J.C., 1974. Numerical taxonomy with fuzzy sets. J. Math. Biology 1, 57–71. https://doi.org/10.1007/BF02339490

Bezdek†, J.C., 1973. Cluster Validity with Fuzzy Sets. Journal of Cybernetics 3, 58–73. https://doi.org/10.1080/01969727308546047

Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences 10, 191–203. https://doi.org/10.1016/0098-3004(84)90020-7

Bharath, R., Srinivas, V.V., 2015. Delineation of homogeneous hydrometeorological regions using wavelet-based global fuzzy cluster analysis. Int. J. Climatol. 35, 4707–4727. https://doi.org/10.1002/joc.4318

Bonell, M., Sumner, G., 1992. Atmospheric circulation and daily precipitation in Wales. Theor Appl Climatol 46, 3–25. https://doi.org/10.1007/BF00866443

Burn, D.H., 1997. Catchment similarity for regional flood frequency analysis using seasonality measures. Journal of Hydrology 202, 212–230. https://doi.org/10.1016/S0022-1694(97)00068-1

Buytaert, W., Célleri, R., Timbe, L., 2009. Predicting climate change impacts on water resources in the tropical Andes: Effects of GCM uncertainty: CLIMATE CHANGE IMPACTS ON WATER RESOURCES. Geophys. Res. Lett. 36, n/a-n/a. https://doi.org/10.1029/2008GL037048

Buytaert, W., Celleri, R., Willems, P., Bièvre, B.D., Wyseure, G., 2006. Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes. Journal of Hydrology 329, 413–421. https://doi.org/10.1016/j.jhydrol.2006.02.031

Chatterjee, S., Khan, A., Akbari, H., Wang, Y., 2016. Monotonic trends in spatio-temporal distribution and concentration of monsoon precipitation (1901–2002), West Bengal, India. Atmospheric Research 182, 54–75. https://doi.org/10.1016/j.atmosres.2016.07.010

Chavoshi, S., Azmin Sulaiman, W.N., Saghafian, B., Bin Sulaiman, Md.N., Manaf, L.A., 2013. Regionalization by fuzzy expert system based approach optimized by genetic

algorithm. Journal of Hydrology 486, 271–280. https://doi.org/10.1016/j.jhydrol.2013.01.033

Chen, J., Brissette, F.P., Chaumont, D., Braun, M., 2013. Performance and uncertainty evaluation of empirical downscaling methods in quantifying the climate change impacts on hydrology over two North American river basins. Journal of Hydrology 479, 200–214. https://doi.org/10.1016/j.jhydrol.2012.11.062

Chen, J., Xia, J., Zhao, C., Zhang, S., Fu, G., Ning, L., 2014. The mechanism and scenarios of how mean annual runoff varies with climate change in Asian monsoon areas. Journal of Hydrology 517, 595–606. https://doi.org/10.1016/j.jhydrol.2014.05.075

Chen, J., Zhao, S., Wang, H., 2011. Risk Analysis of Flood Disaster Based on Fuzzy Clustering Method. Energy Procedia 5, 1915–1919. https://doi.org/10.1016/j.egypro.2011.03.329

Chew, H.H., Heiler, T.D., 1974. Magnitude and Frequency of floods in Peninsular Malaysia, Hydrological procedure ; no. 4. Kuala Lumpur : Ministry of Agriculture and Fisheries, 1974., Malaysia.

Climate Change 2007: Synthesis Report, IPCC

Cox, D.R., Stuart, A., 1955. SOME QUICK SIGN TESTS FOR TREND IN LOCATION AND DISPERSION. Biometrika 42, 80–95. https://doi.org/10.1093/biomet/42.1-2.80

Darand, M., Mansouri Daneshvar, M.R., 2014. Regionalization of Precipitation Regimes in Iran Using Principal Component Analysis and Hierarchical Clustering Analysis. Environ. Process. 1, 517–532. https://doi.org/10.1007/s40710-014-0039-1

Davis, J.C., 2002. Statistics and data analysis in geology, 3rd ed. ed. Wiley, New York, NY.

Deka, S.K., Sarma, A.K., 2011. Impact of climate change on precipitation characteristics of Brahmaputra basin. IIT Guwahati, Guwahati, https://www.iitg.ac.in/bpccp/report/report1.pdf

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, 145–163. https://doi.org/10.1016/j.jhydrol.2004.10.012

Dikbas, F., Firat, M., Koc, A.C., Gungor, M., 2012. Classification of precipitation series using fuzzy cluster method. Int. J. Climatol. 32, 1596–1603. https://doi.org/10.1002/joc.2350

Dinpashoh, Y., Fakheri-Fard, A., Moghaddam, M., Jahanbakhsh, S., Mirnia, M., 2004. Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. Journal of Hydrology 297, 109–123. https://doi.org/10.1016/j.jhydrol.2004.04.009

Dunn, D.M., Landwehr, J.M., 1980. Analyzing Clustering Effects across Time. Journal of the American Statistical Association 75, 8–15. https://doi.org/10.1080/01621459.1980.10477414

Dunn, J.C., 1973. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics 3, 32–57. https://doi.org/10.1080/01969727308546046

Dutta, P., Sarma, A.K., 2020. Hydrological modeling as a tool for water resources management of the data-scarce Brahmaputra basin. Journal of Water and Climate Change jwc2020186. https://doi.org/10.2166/wcc.2020.186

Eliasson, I., 2000. The use of climate knowledge in urban planning. Landscape and Urban Planning 48, 31–44. https://doi.org/10.1016/S0169-2046(00)00034-7

Emissions Scenarios (2000), IPCC, *Cambridge University Press*, UK. pp. 570

Estrela, T., Pérez-Martin, M.A., Vargas, E., 2012. Impacts of climate change on water resources in Spain. Hydrological Sciences Journal 57, 1154–1167. https://doi.org/10.1080/02626667.2012.702213

Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J., Taylor, K.E., 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci. Model Dev. 9, 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Faccioli, M., Torres, C.M., Font, A.R., 2014. Decision making under inherent uncertainty: does preference analysis play a role in the design of wetland adaptation to climate change? (DEA Working Papers No. 66). Universitat de les Illes Balears, Departament d'Economía Aplicada.

Farsadnia, F., Rostami Kamrood, M., Moghaddam Nia, A., Modarres, R., Bray, M.T., Han, D., Sadatinejad, J., 2014. Identification of homogeneous regions for regionalization of watersheds by two-level self-organizing feature maps. Journal of Hydrology 509, 387–397. https://doi.org/10.1016/j.jhydrol.2013.11.050

Fatichi, S., Rimkus, S., Burlando, P., Bordoy, R., Molnar, P., 2015. High-resolution distributed analysis of climate and anthropogenic changes on the hydrology of an Alpine catchment. Journal of Hydrology 525, 362–382. https://doi.org/10.1016/j.jhydrol.2015.03.036

Fill, H.D., Stedinger, J.R., 1995. Homogeneity tests based upon Gumbel distribution and a critical appraisal of Dalrymple's test. Journal of Hydrology 166, 81–105. https://doi.org/10.1016/0022-1694(94)02599-7

Flato, G.M., 2011. Earth system models: an overview: Earth system models. WIREs Clim Change 2, 783–800. https://doi.org/10.1002/wcc.148

Forsythe, N., Fowler, H.J., Blenkinsop, S., Burton, A., Kilsby, C.G., Archer, D.R., Harpham, C., Hashmi, M.Z., 2014. Application of a stochastic weather generator to assess climate change impacts in a semi-arid climate: The Upper Indus Basin. Journal of Hydrology 517, 1019–1034. https://doi.org/10.1016/j.jhydrol.2014.06.031

Fukuyama, Y., Sugeno, M., 1989. A new method of choosing the number of clusters for the fuzzy c-means method, in: Proc. 5th Fuzzy System Symposium. pp. 247–250.

Gadgil, S., Yadumani, Joshi, N.V., 1993. Coherent rainfall zones of the Indian region. Int. J. Climatol. 13, 547–566. https://doi.org/10.1002/joc.3370130506

Gao, J., Sheshukov, A.Y., Yen, H., Douglas-Mankin, K.R., White, M.J., Arnold, J.G., 2019. Uncertainty of hydrologic processes caused by bias-corrected CMIP5 climate change projections with alternative historical data sources. Journal of Hydrology 568, 551–561. https://doi.org/10.1016/j.jhydrol.2018.10.041

Gardelle, J., Arnaud, Y., Berthier, E., 2011. Contrasted evolution of glacial lakes along the Hindu Kush Himalaya mountain range between 1990 and 2009. Global and Planetary Change 75, 47–55. https://doi.org/10.1016/j.gloplacha.2010.10.003

Ghose, S., Mujumdar, P.P., 2006. Future Rainfall Scenario over Orissa with GCM Projections by Statistical Downscaling. Current Science 90, 396–404.

Goktepe, A.B., Altun, S., Sezer, A., 2005. Soil clustering by fuzzy c-means algorithm. Advances in Engineering Software 36, 691–698. https://doi.org/10.1016/j.advengsoft.2005.01.008

Gosain, A.K., Rao, S., Basuray, D., 2006. Climate change impact assessment on hydrology of Indian river basins. Current Science 90, 346–353.

Goyal, M.K., Sharma, A., 2016. A fuzzy c-means approach regionalization for analysis of meteorological drought homogeneous regions in western India. Nat Hazards 84, 1831–1847. https://doi.org/10.1007/s11069-016-2520-9

Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R., 1979. Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. Water Resour. Res. 15, 1049–1054. https://doi.org/10.1029/WR015i005p01049

Guttman, N.B., 1993. The Use of L-Moments in the Determination of Regional Precipitation Climates. Journal of Climate 6, 2309–2325. https://doi.org/10.1175/1520-0442(1993)006<2309:TUOLMI>2.0.CO;2

Halkidi, M., Batistakis, Y., Vazirgiannis, M., 2001. On clustering validation techniques. Journal of Intelligent Information Systems 17, 107–145. https://doi.org/10.1023/A:1012801612483

Hall, M.J., Minns, A.W., 1999. The classification of hydrologically homogeneous regions. Hydrological Sciences Journal 44, 693–704. https://doi.org/10.1080/02626669909492268

Hamed, K.H., Ramachandra Rao, A., 1998. A modified Mann-Kendall trend test for autocorrelated data. Journal of Hydrology 204, 182–196. https://doi.org/10.1016/S0022-1694(97)00125-X

Hessami, M., Gachon, P., Ouarda, T.B.M.J., St-Hilaire, A., 2008. Automated regression-based statistical downscaling tool. Environmental Modelling & Software 23, 813–834. https://doi.org/10.1016/j.envsoft.2007.10.004

Heyen, H., Zorita, E., Von Storch, H., 1996. Statistical downscaling of monthly mean North Atlantic air-pressure to sea level anomalies in the Baltic Sea. Tellus A: Dynamic Meteorology and Oceanography 48, 312–323. https://doi.org/10.3402/tellusa.v48i2.12062

Hosking, J.R.M., 1990. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. Journal of the Royal Statistical Society: Series B (Methodological) 52, 105–124. https://doi.org/10.1111/j.2517-6161.1990.tb01775.x

Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis: An Approach Based on L-Moments, 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9780511529443

Hosking, J.R.M., Wallis, J.R., 1995. Correction to "Some Statistics Useful in Regional Frequency Analysis." Water Resour. Res. 31, 251–251. https://doi.org/10.1029/94WR02510

Hosking, J.R.M., Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. Water Resour. Res. 29, 271–281. https://doi.org/10.1029/92WR01980

Hoyer, R., Chang, H., 2014. Assessment of freshwater ecosystem services in the Tualatin and Yamhill basins under climate change and urbanization. Applied Geography 53, 402–416. https://doi.org/10.1016/j.apgeog.2014.06.023

Huang, J., Zhang, J., Zhang, Z., Xu, C., Wang, B., Yao, J., 2011. Estimation of future precipitation change in the Yangtze River basin by using statistical downscaling method. Stoch Environ Res Risk Assess 25, 781–792. https://doi.org/10.1007/s00477-010-0441-9

Huan-Yu LIN, Jer-Ming HU, Tze-Ying CHEN, Chang-Fu HSIEH, Guangyu WANG, Tongli WANG, 2018. A dynamic downscaling approach to generate scale-free regional climate data in Taiwan. TAIWANIA 63. https://doi.org/10.6165/tai.2018.63.251

Immerzeel, W.W., van Beek, L.P.H., Bierkens, M.F.P., 2010. Climate Change Will Affect the Asian Water Towers. Science 328, 1382–1385. https://doi.org/10.1126/science.1183188

IPCC, 2014: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the FifthAssessment Report of the Intergovernmental Panel on Climate Change*[Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1132 pp

IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp

Irwin, S., Srivastav, R.K., Simonovic, S.P., Burn, D.H., 2017. Delineation of precipitation regions using location and atmospheric variables in two Canadian climate regions: the role of attribute selection. Hydrological Sciences Journal 62, 191–204. https://doi.org/10.1080/02626667.2016.1183776

Iyengar, R.N., Basak, P., 1994. Regionalization of Indian monsoon rainfall and long-term variability signals. Int. J. Climatol. 14, 1095–1114. https://doi.org/10.1002/joc.3370141003

Jena, P., Azad, S., Rajeevan, M., 2015. Statistical Selection of the Optimum Models in the CMIP5 Dataset for Climate Change Projections of Indian Monsoon Rainfall. Climate 3, 858–875. https://doi.org/10.3390/cli3040858

Joseph, J., Ghosh, S., Pathak, A., Sahai, A.K., 2018. Hydrologic impacts of climate change: Comparisons between hydrological parameter uncertainty and climate model uncertainty. Journal of Hydrology 566, 1–22. https://doi.org/10.1016/j.jhydrol.2018.08.080

Jung, I.-W., Moradkhani, H., Chang, H., 2012. Uncertainty assessment of climate change impacts for hydrologically distinct river basins. Journal of Hydrology 466–467, 73–87. https://doi.org/10.1016/j.jhydrol.2012.08.002

Kang, H., An, K.-H., Park, C.-K., Solis, A.L.S., Stitthichivapak, K., 2007. Multimodel output statistical downscaling prediction of precipitation in the Philippines and Thailand: PRECIPITATION IN THE PHILIPPINES AND THAILAND. Geophys. Res. Lett. 34. https://doi.org/10.1029/2007GL030730

Kannan, S., Ghosh, S., 2011. Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output. Stoch Environ Res Risk Assess 25, 457–474. https://doi.org/10.1007/s00477-010-0415-y

Karamouz, M., Hosseinpour, A., Nazif, S., 2011. Improvement of Urban Drainage System Performance under Climate Change Impact: Case Study. J. Hydrol. Eng. 16, 395–412. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000317

Kendall, M.G., 1975. Rank Correlation methods. Griffin, London.

Komalasari, K.E., Pawitan, H., Faqih, A., 2017. Descriptive Statistics and Cluster Analysis for Extreme Rainfall in Java Island. IOP Conference Series: Earth and Environmental Science 58, 012039. https://doi.org/10.1088/1755-1315/58/1/012039

Kulkarni, A., 2017. Homogeneous clusters over India using probability density function of daily rainfall. Theor Appl Climatol 129, 633–643. https://doi.org/10.1007/s00704-016-1808-8

Kulkarni, A., Kripalani, R.H., Singh, S.V., 1992. Classification of summer monsoon rainfall patterns over India. Int. J. Climatol. 12, 269–280. https://doi.org/10.1002/joc.3370120304

Kumar, P., Wiltshire, A., Mathison, C., Asharaf, S., Ahrens, B., Lucas-Picher, P., Christensen, J.H., Gobiet, A., Saeed, F., Hagemann, S., Jacob, D., 2013. Downscaled climate change projections with uncertainty assessment over India using a high resolution multi-model approach. Science of The Total Environment 468–469, S18–S30. https://doi.org/10.1016/j.scitotenv.2013.01.051

Kumar, R.R., Sarma, A.K., 2010. Rain Water Harvesting for Negotiating Impact of Variation in Climatic Parameter on Agriculture, in: Pre Conf. Proc. of Indo-Italian Workshop on Impact of Climate Change and Anthropogenic Activities on Soil and Water Resources. IIT Roorkee.

Kumari, M., Singh, C.K., Basistha, A., 2017. Clustering Data and Incorporating Topographical Variables for Improving Spatial Interpolation of Rainfall in Mountainous Region. Water Resour Manage 31, 425–442. https://doi.org/10.1007/s11269-016-1534-0

Kundzewicz, Z.W., Krysanova, V., Benestad, R.E., Hov, Ø., Piniewski, M., Otto, I.M., 2018. Uncertainty in climate change impacts on water resources. Environmental Science & Policy 79, 1–8. https://doi.org/10.1016/j.envsci.2017.10.008

Kwon, S.H., 1998. Cluster validity index for fuzzy clustering. Electron. Lett. 34, 2176. https://doi.org/10.1049/el:19981523

LANZANTE, JOHN.R., 1996. RESISTANT, ROBUST AND NON-PARAMETRIC TECHNIQUES FOR THE ANALYSIS OF CLIMATE DATA: THEORY AND EXAMPLES, INCLUDING APPLICATIONS TO HISTORICAL RADIOSONDE STATION DATA. International Journal of Climatology 16, 1197–1226. https://doi.org/10.1002/(SICI)1097-0088(199611)16:11<1197::AID-JOC89>3.0.CO;2-L

Lespinas, F., Ludwig, W., Heussner, S., 2014. Hydrological and climatic uncertainties associated with modeling the impact of climate change on water resources of small Mediterranean coastal rivers. Journal of Hydrology 511, 403–422. https://doi.org/10.1016/j.jhydrol.2014.01.033

Longobardi, A., Villani, P., 2009. Trend analysis of annual and seasonal rainfall time series in the Mediterranean area. Int. J. Climatol. n/a-n/a. https://doi.org/10.1002/joc.2001

Mahmood, R., Babel, M.S., 2013. Evaluation of SDSM developed by annual and monthly sub-models for downscaling temperature and precipitation in the Jhelum basin, Pakistan and India. Theor Appl Climatol 113, 27–44. https://doi.org/10.1007/s00704-012-0765-0

Mann, H.B., 1945. Nonparametric Tests Against Trend. Econometrica 13, 245. https://doi.org/10.2307/1907187

Mannan, A., Chaudhary, S., Dhanya, C.T., Swamy, A.K., 2018. Regionalization of rainfall characteristics in India incorporating climatic variables and using self-organizing maps. ISH Journal of Hydraulic Engineering 24, 147–156. https://doi.org/10.1080/09715010.2017.1400409

Mehrotra, D., Mehrotra, R., 1995. Climate change and hydrology with emphasis on the Indian subcontinent. Hydrological Sciences Journal 40, 231–242. https://doi.org/10.1080/02626669509491406

Mehrotra, R., Sharma, A., 2006. A nonparametric stochastic downscaling framework for daily rainfall at multiple locations. J. Geophys. Res. 111, D15101. https://doi.org/10.1029/2005JD006637

Mehrotra, R., Sharma, A., 2005. A nonparametric nonhomogeneous hidden Markov model for downscaling of multisite daily rainfall occurrences. J. Geophys. Res. 110, D16108. https://doi.org/10.1029/2004JD005677

Milne, A.E., Glendining, M.J., Bellamy, P., Misselbrook, T., Gilhespy, S., Rivas Casado, M., Hulin, A., van Oijen, M., Whitmore, A.P., 2014. Analysis of uncertainties in the estimates of nitrous oxide and methane emissions in the UK's greenhouse gas inventory for agriculture. Atmospheric Environment 82, 94–105. https://doi.org/10.1016/j.atmosenv.2013.10.012

Misra, V., Dirmeyer, P.A., Kirtman, B.P., 2003. Dynamic Downscaling of Seasonal Simulations over South America. Journal of Climate 16, 103–117. https://doi.org/10.1175/1520-0442(2003)016<0103:DDOSSO>2.0.CO;2

Mok, P.Y., Huang, H.Q., Kwok, Y.L., Au, J.S., 2012. A robust adaptive clustering analysis method for automatic identification of clusters. Pattern Recognition 45, 3017–3033. https://doi.org/10.1016/j.patcog.2012.02.003

Moursi, H., Kim, D., Kaluarachchi, J.J., 2017. A probabilistic assessment of agricultural water scarcity in a semi-arid and snowmelt-dominated river basin under climate change. Agricultural Water Management 193, 142–152. https://doi.org/10.1016/j.agwat.2017.08.010

Mujumdar, P.P., 2008. Implications of climate change for sustainable water resources management in India. Physics and Chemistry of the Earth, Parts A/B/C 33, 354–358. https://doi.org/10.1016/j.pce.2008.02.014

Mujumdar, P.P., Ghosh, S., 2008. Modeling GCM and scenario uncertainty using a possibilistic approach: Application to the Mahanadi River, India: GCM AND SCENARIO UNCERTAINTY. Water Resour. Res. 44. https://doi.org/10.1029/2007WR006137

Najafi, M.R., Moradkhani, H., 2015. Multi-model ensemble analysis of runoff extremes for climate change impact assessments. Journal of Hydrology 525, 352–361. https://doi.org/10.1016/j.jhydrol.2015.03.045

Najafi, M. R., Moradkhani, H., Jung, I.W., 2011. Assessing the uncertainties of hydrologic model selection in climate change impact studies. Hydrol. Process. 25, 2814–2826. https://doi.org/10.1002/hyp.8043

Najafi, Mohammad Reza, Moradkhani, H., Wherry, S.A., 2011. Statistical Downscaling of Precipitation Using Machine Learning with Optimal Predictor Selection. J. Hydrol. Eng. 16, 650–664. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000355

NERC, 1975. Flood Studies Report, five volumes. Natural Environmental Research Council (NERC), Department of the Environment, London.

Neupane, R.P., White, J.D., Alexander, S.E., 2015. Projected hydrologic changes in monsoon-dominated Himalaya Mountain basins with changing climate and deforestation. Journal of Hydrology 525, 216–230. https://doi.org/10.1016/j.jhydrol.2015.03.048

Nkhonjera, G.K., 2017. Understanding the impact of climate change on the dwindling water resources of South Africa, focusing mainly on Olifants River basin: A review. Environmental Science & Policy 71, 19–29. https://doi.org/10.1016/j.envsci.2017.02.004

Nkhonjera, G.K., Dinka, M.O., 2017. Significance of direct and indirect impacts of climate change on groundwater resources in the Olifants River basin: A review. Global and Planetary Change 158, 72–82. https://doi.org/10.1016/j.gloplacha.2017.09.011

Oguntunde, P.G., Abiodun, B.J., Olukunle, O.J., Olufayo, A.A., 2012. Trends and variability in pan evaporation and other climatic variables at Ibadan, Nigeria, 1973-2008. Met. Apps 19, 464–472. https://doi.org/10.1002/met.281

Owen, S.M., MacKenzie, A.R., Bunce, R.G.H., Stewart, H.E., Donovan, R.G., Stark, G., Hewitt, C.N., 2006. Urban land classification and its uncertainties using principal component and cluster analyses: A case study for the UK West Midlands. Landscape and Urban Planning 78, 311–321. https://doi.org/10.1016/j.landurbplan.2005.11.002

Pal, N.R., Bezdek, J.C., 1995. On cluster validity for the fuzzy c-means model. IEEE Trans. Fuzzy Syst. 3, 370–379. https://doi.org/10.1109/91.413225

Pelczer, I., Ramos, J., Domínguez, R., González, F., 2007. ESTABLISHMENT OF REGIONAL HOMOGENEOUS ZONES IN A WATERSHED USING CLUSTERING ALGORITHMS, in: Harmonizing the Demands of Art and Nature in Hydraulics. Presented at the International Association for Hydraulic Research (IAHR), CORILA, Venice, Italy, p. 791.

Plain, M.B., Minasny, B., McBratney, A.B., Vervoort, R.W., 2008. Spatially explicit seasonal forecasting using fuzzy spatiotemporal clustering of long-term daily rainfall and temperature data. Hydrol. Earth Syst. Sci. Discuss. 5, 1159–1189. https://doi.org/10.5194/hessd-5-1159-2008

Rahman, M., Jahan, S., Kamal, M.M., 2010. RESPONSE OF CLIMATE CHANGE ON THE MORPHOLOGICAL BEHAVIOR OF THE MAJOR RIVER SYSTEM OF BANGLADESH.

Rao, A.R., Srinivas, V.V., 2006a. Regionalization of watersheds by fuzzy cluster analysis. Journal of Hydrology 318, 57–79. https://doi.org/10.1016/j.jhydrol.2005.06.004

Rao, A.R., Srinivas, V.V., 2006b. Regionalization of watersheds by hybrid-cluster analysis. Journal of Hydrology 318, 37–56. https://doi.org/10.1016/j.jhydrol.2005.06.003

Remesan, R., Holman, I.P., 2015. Effect of baseline meteorological data selection on hydrological modelling of climate change scenarios. Journal of Hydrology 528, 631–642. https://doi.org/10.1016/j.jhydrol.2015.06.026

Riley, S.J., DeGloria, S.D., Elliot, R., 1999. A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity. Int. J. Sci. 5, 23–27.

Rossi, R.E., Mulla, D.J., Journel, A.G., Franz, E.H., 1992. Geostatistical Tools for Modeling and Interpreting Ecological Spatial Dependence. Ecological Monographs 62, 277–314. https://doi.org/10.2307/2937096

Roubens, M., 1982. Fuzzy clustering algorithms and their cluster validity. European Journal of Operational Research 10, 294–301. https://doi.org/10.1016/0377-2217(82)90228-4

Runting, R.K., Bryan, B.A., Dee, L.E., Maseyk, F.J.F., Mandle, L., Hamel, P., Wilson, K.A., Yetka, K., Possingham, H.P., Rhodes, J.R., 2017. Incorporating climate change into ecosystem service assessments and decisions: a review. Glob Change Biol 23, 28–41. https://doi.org/10.1111/gcb.13457

Sadri, S., Burn, D.H., 2011. A Fuzzy C-Means approach for regionalization using a bivariate homogeneity and discordancy approach. Journal of Hydrology 401, 231–239. https://doi.org/10.1016/j.jhydrol.2011.02.027

Saikranthi, K., Rao, T.N., Rajeevan, M., Bhaskara Rao, S.V., 2013. Identification and Validation of Homogeneous Rainfall Zones in India Using Correlation Analysis. Journal of Hydrometeorology 14, 304–317. https://doi.org/10.1175/JHM-D-12-071.1

Salvi, K., S., K., Ghosh, S., 2013. High-resolution multisite daily rainfall projections in India with statistical downscaling for climate change impacts assessment: DOWNSCALING OF INDIAN RAINFALL. J. Geophys. Res. Atmos. 118, 3557–3578. https://doi.org/10.1002/jgrd.50280

Sarma, A.K., Sarma, P.K., Bezbaruah, A., Vinnarasi, R., 2012. Climate Data Collection from Tea Gardens and other Sources of Northeast India and Their Analysis for Climate Change Study (A network project taken by B.P. Chaliha Chair for water resources). Civil Engineering Department, IIT Guwahati, Guwahati.

Sarma, B., Sarma, A.K., Singh, V.P., 2013. Optimal Ecological Management Practices (EMPs) for Minimizing the Impact of Climate Change and Watershed Degradation Due to

Urbanization. Water Resour Manage 27, 4069–4082. https://doi.org/10.1007/s11269-013-0396-y

Satyanarayana, P., Srinivas, V.V., 2011. Regionalization of precipitation in data sparse areas using large scale atmospheric variables – A fuzzy clustering approach. Journal of Hydrology 405, 462–473. https://doi.org/10.1016/j.jhydrol.2011.05.044

Satyanarayana, P., Srinivas, V.V., 2008. Regional frequency analysis of precipitation using large-scale atmospheric variables. J. Geophys. Res. 113, D24110. https://doi.org/10.1029/2008JD010412

Serrat-Capdevila, A., Valdés, J.B., Pérez, J.G., Baird, K., Mata, L.J., Maddock, T., 2007. Modeling climate change impacts – and uncertainty – on the hydrology of a riparian system: The San Pedro Basin (Arizona/Sonora). Journal of Hydrology 347, 48–66. https://doi.org/10.1016/j.jhydrol.2007.08.028

Shamir, E., Megdal, S.B., Carrillo, C., Castro, C.L., Chang, H.-I., Chief, K., Corkhill, F.E., Eden, S., Georgakakos, K.P., Nelson, K.M., Prietto, J., 2015. Climate change and water resources management in the Upper Santa Cruz River, Arizona. Journal of Hydrology 521, 18–33. https://doi.org/10.1016/j.jhydrol.2014.11.062

Sharma, A., Goyal, M.K., 2020. Assessment of the changes in precipitation and temperature in Teesta River basin in Indian Himalayan Region under climate change. Atmospheric Research 231, 104670. https://doi.org/10.1016/j.atmosres.2019.104670

Sharma, R.H., Shakya, N.M., 2006. Hydrological changes and its impact on water resources of Bagmati watershed, Nepal. Journal of Hydrology 327, 315–322. https://doi.org/10.1016/j.jhydrol.2005.11.051

Shen, M., Chen, J., Zhuan, M., Chen, H., Xu, C.-Y., Xiong, L., 2018. Estimating uncertainty and its temporal variation related to global climate models in quantifying climate change impacts on hydrology. Journal of Hydrology 556, 10–24. https://doi.org/10.1016/j.jhydrol.2017.11.004

Shivam, G., Goyal, M.K., Sarma, A.K., 2017b. Index-Based Study of Future Precipitation Changes over Subansiri River Catchment under Changing Climate. J ENV INFORM. https://doi.org/10.3808/jei.201700376

Shivam, Goyal, M.K., Sarma, A.K., 2017a. Analysis of the change in temperature trends in Subansiri River basin for RCP scenarios using CMIP5 datasets. Theor Appl Climatol 129, 1175–1187. https://doi.org/10.1007/s00704-016-1842-6

Singh, K., Singh, S., 1996. Space-time variation and regionalization of seasonal and monthly summer monsoon rainfall of the sub-Himalayan region and Gangetic plains of India. Clim. Res. 6, 251–262. https://doi.org/10.3354/cr006251

Singh, V., Goyal, M.K., 2016. Analysis and trends of precipitation lapse rate and extreme indices over north Sikkim eastern Himalayas under CMIP5ESM-2M RCPs experiments. Atmospheric Research 167, 34–60. https://doi.org/10.1016/j.atmosres.2015.07.005

Singh, V., Sharma, A., Goyal, M.K., 2019. Projection of hydro-climatological changes over eastern Himalayan catchment by the evaluation of RegCM4 RCM and CMIP5 GCM models. Hydrology Research 50, 117–137. https://doi.org/10.2166/nh.2017.193

Soraisam, B., Karumuri, A., D.S., P., 2018. Uncertainties in observations and climate projections for the North East India. Global and Planetary Change 160, 96–108. https://doi.org/10.1016/j.gloplacha.2017.11.010

Srinivas, V.V., Tripathi, S., Rao, A.R., Govindaraju, R.S., 2008. Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. Journal of Hydrology 348, 148–166. https://doi.org/10.1016/j.jhydrol.2007.09.046

Srinivasa Raju, K., Nagesh Kumar, D., 2007. Classification of Indian meteorological stations using cluster and fuzzy cluster analysis, and Kohonen artificial neural networks. Hydrology Research 38, 303–314. https://doi.org/10.2166/nh.2007.013

Sumner, G., Bonell, M., 1988. Variation in the spatial organisation of daily rainfall during the north Queensland wet seasons, 1979?82. Theor Appl Climatol 39, 59–72. https://doi.org/10.1007/BF00866390

Thomas, D.M., Benson, M.A., 1970. Generalization of streamflow characteristics from drainage-basin characteristics, GEOLOGICAL SURVEY WATER-SUPPLY PAPER 1975. Washington. https://doi.org/10.3133/wsp1975

Tisseuil, C., Vrac, M., Lek, S., Wade, A.J., 2010. Statistical downscaling of river flows. Journal of Hydrology 385, 279–291. https://doi.org/10.1016/j.jhydrol.2010.02.030

Van Dingenen, R., Dentener, F.J., Raes, F., Krol, M.C., Emberson, L., Cofala, J., 2009. The global impact of ozone on agricultural crop yields under current and future air quality legislation. Atmospheric Environment 43, 604–618. https://doi.org/10.1016/j.atmosenv.2008.10.033

Venkatesh, B., Jose, M.K., 2007. Identification of homogeneous rainfall regimes in parts of Western Ghats region of Karnataka. J Earth Syst Sci 116, 321–329. https://doi.org/10.1007/s12040-007-0029-z

Viglione, A., Laio, F., Claps, P., 2007. A comparison of homogeneity tests for regional frequency analysis. Water Resour. Res. 43. https://doi.org/10.1029/2006WR005095

Vinnarasi, R., 2012. Impact of climate change on rainfal and stream flow of Dhansiri basin (M.Tech Thesis). IIT Guwahati, Guwahati.

Vinnarasi, R., Sarma, A.K., 2011. Impact of climate change on a southern tributary of Brahmaputra basin: a study report, IIT Guwahati, Guwahati, https://www.iitg.ac.in/bpccp/report/report2.pdf

Vinnarasi, R., Sarma, A.K., 2012. Statistical Downscaling of GCM for Predicting Seasonal Rainfall with Short Duration Historical Data, International Conference on Advances in Mechanical, Manufacturing and Building Sciences (ICAMB) – 2012, SMBS, VIT University, Vellore, India

Von Storch, H., Navarra, A., 2010. Analysis of climate variability: applications of statistical techniques ; proceedings of an autumn school organized by the Commission of the European Community on Elba from October 30 to November 6, 1993, 2nd, updated a. extended ed ed. Presented at the Autumn School, Springer Berlin, Berlin.

Wagner, P.D., Reichenau, T.G., Kumar, S., Schneider, K., 2015. Development of a new downscaling method for hydrologic assessment of climate change impacts in data scarce regions and its application in the Western Ghats, India. Reg Environ Change 15, 435–447. https://doi.org/10.1007/s10113-013-0481-z

Wang, D., Hagen, S.C., Alizad, K., 2013. Climate change impact and uncertainty analysis of extreme rainfall events in the Apalachicola River basin, Florida. Journal of Hydrology 480, 125–135. https://doi.org/10.1016/j.jhydrol.2012.12.015

Wang, Q., Fan, X., Qin, Z., Wang, M., 2012. Change trends of temperature and precipitation in the Loess Plateau Region of China, 1961–2010. Global and Planetary Change 92–93, 138–147. https://doi.org/10.1016/j.gloplacha.2012.05.010

Wang, Z., Zeng, Z., Lai, C., Lin, W., Wu, X., Chen, X., 2017. A regional frequency analysis of precipitation extremes in Mainland China with fuzzy c-means and L-moments approaches. Int. J. Climatol 37, 429–444. https://doi.org/10.1002/joc.5013

Watanabe, S., Kanae, S., Seto, S., Yeh, P.J.-F., Hirabayashi, Y., Oki, T., 2012. Intercomparison of bias-correction methods for monthly temperature and precipitation simulated by multiple climate models: BIAS-CORRECTION-METHODS INTERCOMPARISON. J. Geophys. Res. 117, n/a-n/a. https://doi.org/10.1029/2012JD018192

Weiss, J., Bernardara, P., Benoit, M., 2014. Formation of homogeneous regions for regional frequency analysis of extreme significant wave heights. J. Geophys. Res. Oceans 119, 2906–2922. https://doi.org/10.1002/2013JC009668

Wigley, T.M.L., barnett, T.P., 1990. Detection of the Greenhouse Effect in the Observations, in: Climate Change: The IPCC Assessment. Cambridge University Press, Cambridge, pp. 239–55.

Wilby, R.L., 2005. Uncertainty in water resource model parameters used for climate change impact assessment. Hydrol. Process. 19, 3201–3219. https://doi.org/10.1002/hyp.5819

Wilby, R.L., Dawson, C.W., 2007. SDSM 4.2 — A decision support tool for the assessment of regional climate change impacts. SDSM User Manual.

Wilby, R.L., Dawson, C.W., Barrow, E.M., 2002. sdsm — a decision support tool for the assessment of regional climate change impacts. Environmental Modelling & Software 17, 145–157. https://doi.org/10.1016/S1364-8152(01)00060-3

Wilby, Robert L., Hassan, H., Hanaki, K., 1998. Statistical downscaling of hydrometeorological variables using general circulation model output. Journal of Hydrology 205, 1–19. https://doi.org/10.1016/S0022-1694(97)00130-3

Wilby, R. L., Wigley, T.M.L., Conway, D., Jones, P.D., Hewitson, B.C., Main, J., Wilks, D.S., 1998. Statistical downscaling of general circulation model output: A comparison of methods. Water Resour. Res. 34, 2995–3008. https://doi.org/10.1029/98WR02577

Woldemeskel, F.M., Sharma, A., Sivakumar, B., Mehrotra, R., 2016. Quantification of precipitation and temperature uncertainties simulated by CMIP3 and CMIP5 models: Comparison of CMIP3 and CMIP5 Uncertainty. J. Geophys. Res. Atmos. 121, 3–17. https://doi.org/10.1002/2015JD023719

Woldemeskel, F.M., Sharma, A., Sivakumar, B., Mehrotra, R., 2012. An error estimation method for precipitation and temperature projections for future climates: QUANTIFYING GCMS OUTPUT UNCERTAINTY. J. Geophys. Res. 117, n/a-n/a. https://doi.org/10.1029/2012JD018062

Wotling, G., Bouvier, C., Danloux, J., Fritsch, J.-M., 2000. Regionalization of extreme precipitation distribution using the principal components of the topographical environment. Journal of Hydrology 233, 86–101. https://doi.org/10.1016/S0022-1694(00)00232-8

Xie, X.L., Beni, G., 1991. A validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Machine Intell. 13, 841–847. https://doi.org/10.1109/34.85677

Xu, J., Grumbine, R.E., Shrestha, A., Eriksson, M., Yang, X., Wang, Y., Wilkes, A., 2009. The Melting Himalayas: Cascading Effects of Climate Change on Water, Biodiversity, and Livelihoods. Conservation Biology 23, 520–530. https://doi.org/10.1111/j.1523-1739.2009.01237.x

Yang, X.L., Xu, L.R., Liu, K.K., Li, C.H., J.Hu, Xia, X.H., 2012. Trends in Temperature and Precipitation in the Zhangweinan River Basin during the last 53 Years. Procedia Environmental Sciences 13, 1966–1974. https://doi.org/10.1016/j.proenv.2012.01.190

Yanming, Z., Jun, W., Xinhua, W., 2012. Study on the Change Trend of Precipitation and Temperature in Kunming City Based on Mann-Kendall Analysis, in: Zhang, T. (Ed.), Future Computer, Communication, Control and Automation, Advances in Intelligent and Soft Computing. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 505–513. https://doi.org/10.1007/978-3-642-25538-0_71

You, Q., Wang, D., Jiang, Z., Kang, S., 2017. Diurnal temperature range in CMIP5 models and observations on the Tibetan Plateau: Diurnal Temperature Range on the Tibetan Plateau. Q.J.R. Meteorol. Soc 143, 1978–1989. https://doi.org/10.1002/qj.3057

Yount, W.R., 1988. Research design and statistical analysis for Christian ministry. Fort Worth, Tex.: W.R. Yount.

Yue, S., Pilon, P., Cavadias, G., 2002. Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. Journal of Hydrology 259, 254–271. https://doi.org/10.1016/S0022-1694(01)00594-7

Zaifoglu, H., Akintug, B., Yanmaz, A.M., 2018. Regional Frequency Analysis of Precipitation Using Time Series Clustering Approaches. J. Hydrol. Eng. 23, 05018007. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001659

Zhang, X.-C., 2005. Spatial downscaling of global climate model output for site-specific assessment of crop production and soil erosion. Agricultural and Forest Meteorology 135, 215–229. https://doi.org/10.1016/j.agrformet.2005.11.016

Zhuang, X.W., Li, Y.P., Nie, S., Fan, Y.R., Huang, G.H., 2018. Analyzing climate change impacts on water resources under uncertainty using an integrated simulation-optimization approach. Journal of Hydrology 556, 523–538. https://doi.org/10.1016/j.jhydrol.2017.11.016

Zorita, E., von Storch, H., 1999. The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods. Journal of Climate 12, 2474–2489. https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2

Zuo, Xu, Chen, Wang, 2019. Downscaling Precipitation in the Data-Scarce Inland River Basin of Northwest China Based on Earth System Data Products. Atmosphere 10, 613. https://doi.org/10.3390/atmos10100613

LIST OF PUBLICATIONS

JOURNALS

1. Hazarika J. and Sarma A. K. "A novel approach to improve downscaling model efficiency by considering orographic influence and uncertainty assessment." *Global and Planetary Change*. Elsevier. (***communicated***)
2. Hazarika J. and Sarma A. K. "Importance of Regional Rainfall Data in Homogeneous Clustering of Data Sparse Areas: A Study in the Upper Brahmaputra Valley Region." *Theoretical and Applied Climatology*. Springer. (***under review***)
3. Sarma, A.K. and Hazarika, J. (2014), "GCM based fuzzy clustering to identify homogeneous climatic regions of North-east India", *WASET International Journal of Environmental, Ecological, Geological and Geophysical Engineering*, **8**(12), 732–739. doi.org/10.5281/zenodo.1096994

CONFERENCES

1. Hazarika J. & Sarma A. K, "Challenges of Data Scarcity in Statistical Downscaling of Rainfall Using Large Scale GCM Models", *International Conference on Sustainable Water Resources Management (SWARM 2020)*, Guwahati, Assam, India. (***accepted***)
2. Hazarika J. & Sarma A. K, "Identification of homogeneous rainfall regions in Northeast India using Fuzzy clustering technique for climate change impacted water resources planning and management", *Inter-Institutional Symposium on Meteorological Research for Improved Water Resources Management*, 3rd March 2019, Cotton University, Assam, India.
3. Hazarika J. & Sarma A. K, "Assessment of Climate Change Impact on Temperature & Rainfall of Guwahati", *22nd International Conference on Hydraulics, Water Resources and Coastal Engineering*, 21-23 December 2017, L. D. College of Engineering, Ahmedabad, India.
4. Hazarika J. & Sarma A. K., "A Study on Impact of Climate Change in Temperature and Precipitation Characteristics of Cherrapunjee, Meghalaya", presented in *the HYDRO-2016 INTERNATIONAL CONFERENCE*, 21st International Conference on Hydraulics, Water Resources and Coastal Engineering, 8-10 December 2016, CWPRS Pune, India.

5. Hazarika J. & Sarma A. K., "Identification of homogeneous climatic regions of North-east India to study impact of climate change in the region-A comparative study", presented in *Regional Forum on Climate Change (RFCC)*, 1-3 July 2015, Asian Institute of Technology, Thailand.