Improvising seasonal prediction model for Indian Summer Monsoon Rainfall

A Report

Submitted for

Partial fulfilment of the requirements

For the degree of

BACHELOR OF TECHNOLOGY

in

Civil Engineering

By

Riddhi Milan Bhatt

ID No: 17cl006

Under the supervision of

Dr. Hiteshri Shastri

February 2021

MANUBHAI SHIVABHAI PATEL DEPARTMENT OF CIVIL ENGINEERING

FACULTY OF TECHNOLOGY AND ENGINEERING

CHAROTAR UNIVERSITY OF SCIENCE & TECHNOLOGY

CHANGA – 388421, GUJARAT, INDIA

CERTIFICATE

This is to ratify that the internship entitled project "**Improvising seasonal prediction model for Indian Summer Monsoon Rainfall''** submitted by Bhatt Riddhi Milan **(ID Number: 17CL006)** to the **Charotar University of Science and Technology** for the partial fulfilment of the Degree of Bachelor of Technology in Civil Engineering.

This is to further certify that I have been supervising the Project (CL450) of Bhatt iddhi Milan **(ID Number: 17CL006).** The contents of this report, in full or in parts, have not been submitted to any other Institute or University for award of any degree, diploma or titles.

Faculty Supervisor
Dr. Hiteshri Shastri
Assistant Professor
Date:

Acknowledgement

I wish to convey my heartfelt acknowledgement to Dr. Hiteshri Shastri, for providing a chance to carry out my project on Improvising seasonal prediction model for Indian Summer Monsoon Rainfall.

I am obliged to her, for her invaluable guidance in the helping me understand about new concepts of data sciences.

I am also thankful to her for all the support, and giving me a great help throughout the project work.

TABLE OF FIGURES

FIGURE 1:1 SST CLUSTERS.....9

FIGURE 2:112-DIFFERENT OCEANIC REGION 10

FIGURE 3:1PREDICTED AND OBSERVED ANOMALY OF RAINFALL.....12

FIGURE 4:1 RESULTS AT 900 HPA 15

FIGURE 4:2 RESULTS FOR 950 HPA.....16

FIGURE 4:3 RESULTS OF 1000 HPA.....17

FIGURE 4:4 RESULTS OF 5-REGION CORRELATION 18

FIGURE 4:5 RESULTS OF 9-REGION CORRELATION 19

FIGURE 4:6 RESULTS OF MULTIPLE REGRESSION.....20

FIGURE 4:7 RESULTS OF APPLYING CART ON CENTRAL REGION 21

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION 6

1.1 IMPORTANCE OF ISMR 6

1.2 EFFECTS OF ISMR ..... **ERROR! BOOKMARK NOT DEFINED.**

Agricultural ..... ***Error! Bookmark not defined.***

Economy ..... ***Error! Bookmark not defined.***

Health ..... ***Error! Bookmark not defined.***

Environmental ..... ***Error! Bookmark not defined.***

1.3 PREDICTION OF INDIAN MONSOON 7

Factors Responsible for Inaccurate Monsoon Forecast *7*

Data sets *7*

Flow chart of methodology *8*

RESULTS SHOWING VARIABLITY OF ISMR *8*

CHAPTER 2 METHODS 10

2.1 CART MODEL 10

2.2 PREDICTORS IN CART MODELLING 10

2.3 OCEANIC REGIONS 11

CHAPTER 3 PROCEDURE / METHOD 12

3.1 PREDICTION OF SPATIAL PATTERN OF INDIAN SUMMER RAINFALL USING CART MODEL 12

3.2 RESULTS 13

CHAPTER 4 EXPERIMENTS CARRIED OUT 14

4.1 PREDICTION OF ISMR USING SST, HUMIDITY AND WIND DATA AT DIFFERENT LEVELS 14

Difference between 900 hPa, 950 hPa and 1000 hPa and reason for selection 14

4.2 APPLYING CO-RELATION TO OCEANIC REGIONS 18

4.3 USING MULTIPLE REGRESSION TECHNIQUE 20

4.4 APPLYING CART ONLY ON CENTRAL REGION 21

CHAPTER 5 CONCLUSION 22

REFERENCES 23

Chapter 1 INTRODUCTION

ISMR is a short from for Indian Summer monsoon rainfall, and it is a very crucial phenomenon for Indian subcontinent, as India is agro-based country which means major part of India's GDP is derived from agriculture and monsoon is one of the major source of water availability. A large part of activities evolve around monsoon and so it is important to be able to predict ISMR.

ISMR will last all the way from June to September, we usually notice that several times monsoon is measured as an average/mean which is not correct representation as all the parts of India do not receive uniform amount of rainfall, in-fact it is highly heterogeneous, and hence prediction becomes extremely important.

1.1 Importance of ISMR

Its is extremely important to understand the spatial pattern of distribution of Indian summer monsoon rainfall or in other way we try to determine the amount of variability with which rainfall is distributed because this variability will largely effect India's water transfer between various river basin. Not only does this variation of ISMR effect the agriculture but it also influence socially and economically to the country's population. Moreover, this distribution also results id consequences like global warming phenomenon.

In the learning given by Goswami *al.* (B. N. Goswami, 2006), the tendency of Indian summer monsoon rainfall over Central India was calculated by taking a particular assumption that the distribution of Indian summer monsoon rainfall is homogeneous and we have also tested its validity in certain experiment. However the results are discussed further to know how this assumption effects prediction of ISMR.

1.2 Prediction of Indian monsoon

The research carried out in past have disclose the fact that Indian summer monsoon rainfall is connected to wide-ranging dynamic parameter like SST (sea surface temperature), humidity and wind currents.

Factors Responsible for Inaccurate Monsoon Forecast

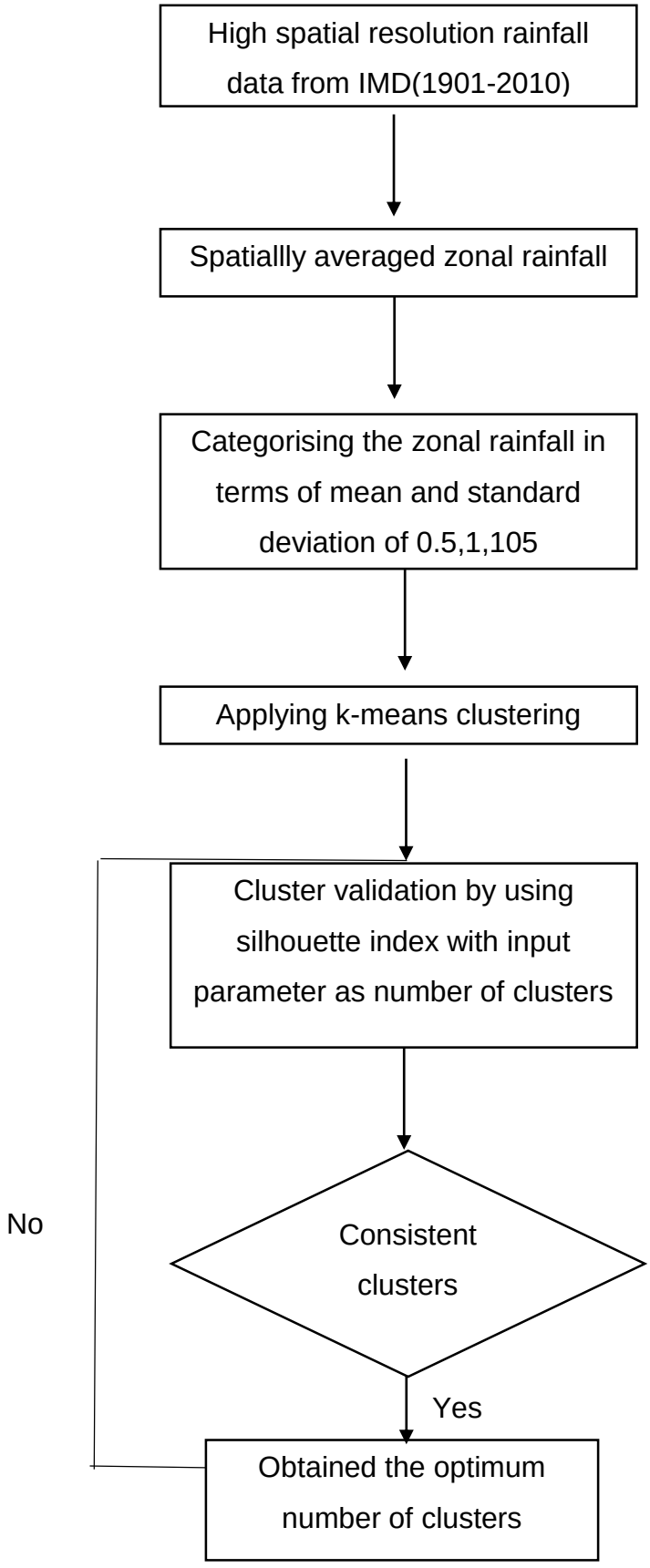
- There is no availability of adequate amount of data at the stations which are monitoring and taking values of all parameters .
- In today's time mostly just automatic weather stations are being used and since them being machines often tend to have some defects in taking the value.
- Sometimes these automatic weather station are not cleaned and calibrated properly will in turn effects data.
- There are various factors like dust, moisture etc which affects data accumulation and hence resulting is inadequate weather forecast.
- Currently the forecast model which are being used in India are developed by scientist from the west and hence those model do not perfectly represent Indian monsoon rainfall.
- And lastly since there are not enough scientist who can develop appropriate model using IMD.

Data sets

The following set of data have been used for the process of forecat of Indian summer monsoon rainfall,

- High amount of resolution that is 0.25° × 0.25° long-range of time period (1901–2015) daily rainfall data in grid pattern over India, is collected from the India Meteorological Department (IMD).
- Statistical monthly sst (sea surface temperature) is taken from extended reconstructed sst (ERSST) version 3b.
- Also monthly zonal and meridional wind data form June to September at 850 HPa, 900 HPa, 950 HPa and 1000 HPa from 1901 to 2010 by ERA-20C.

Flow chart of methodology



RESULTS SHOWING VARIABLITY OF ISMR

The following figure shows rainfall anomaly of Indian sub-continent for the year 1901-2010 distributed in 9 different clusters.

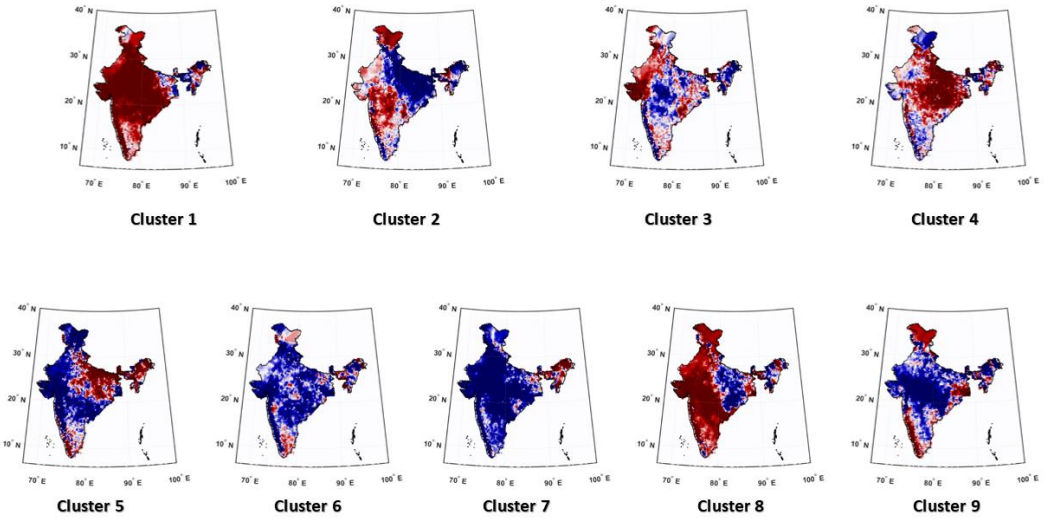


Figure 1:1 sst clusters

Chapter 2 METHODS

There are certain new concept and methods which is adopted in this study, they are as follow:-

2.1 CART MODEL

CART is short from of classification and regression tree. These are basically one type of decision tree which operates on simple algorithm for prediction in modelling. Usually used in machine learning.

Here we are first splitting data into test and training period, test period being 1901-2010 and on this data classification is carried out, while training period being 2011-2015 on which prediction is done. So in this way we try to utilize the concept of binary decision tree method in making our prediction model.

Before understanding how CART is used here, we need a brief idea of predictors

2.2 Predictors in CART modelling

A total of 12 oceanic region have been selected as the input in the CART model and they are called as predictors

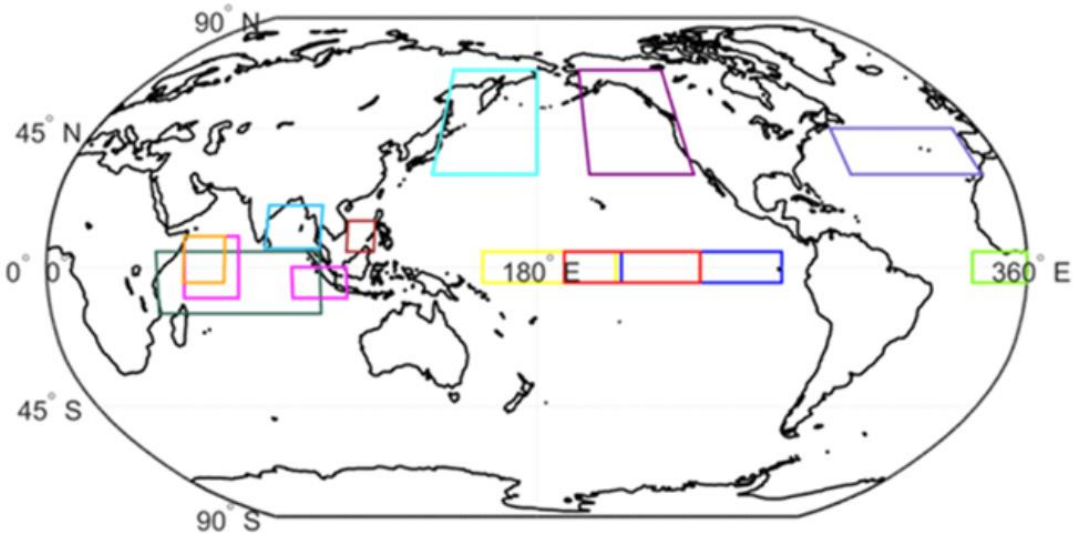


Figure 2:112-different oceanic region

2.3 Oceanic regions

Here, we are selecting 12 different oceanic regions as shown in figure 2:1 and the co-ordinates of oceanic region are as mentioned in table 1

SST Region	SST region Co-ordinates
Atlantic Niño	340°E to 360°E, 5°N to 5°S
Bay of Bengal	80°E to 100°E, 6°N to 20°N
Zonal Mode / Indian Ocean Dipole	50°E to 70°E, 10°S to 10°N, 90°E to 110°E , 10°S to 0°
Niño3	210°E to 270°E, 5°S to 5°N
Niño3.4	190°E to 240°E, 5°S to 5°N
Niño 4	160°E to 210°E, 5°S to 5°N
North Atlantic	300°E to 350°E, 30°N to 45°N
North Pacific 1	140°E to 180°E, 30°N to 65°N
North Pacific 2	200°E to 240°E, 30°N to 65°N
South china sea	110°E to 120°E, 5°N to 15°N
Southern Indian Ocean	40°E to 100°E, 5°S to 15°N
Western Indian Ocean	50°E to 65°E, 5°S to 10°N

Chapter 3 Procedure / Method

3.1 Prediction of spatial pattern of Indian summer rainfall using CART model

Now, we will see how the model of monsoon rainfall is predicted corresponding to period of 2011-2015, firstly we employ categorial mean for march-April-May of SST for each year, over the above mentioned oceanic regions. This is defined as predictor input. For calculating performance of the model anomaly is calculated for the 2011-2015 time period using standard deviation and climatological mean of 1901-2010 period. The computative patterns of annual standard anomalies are collate with CART model predictions.

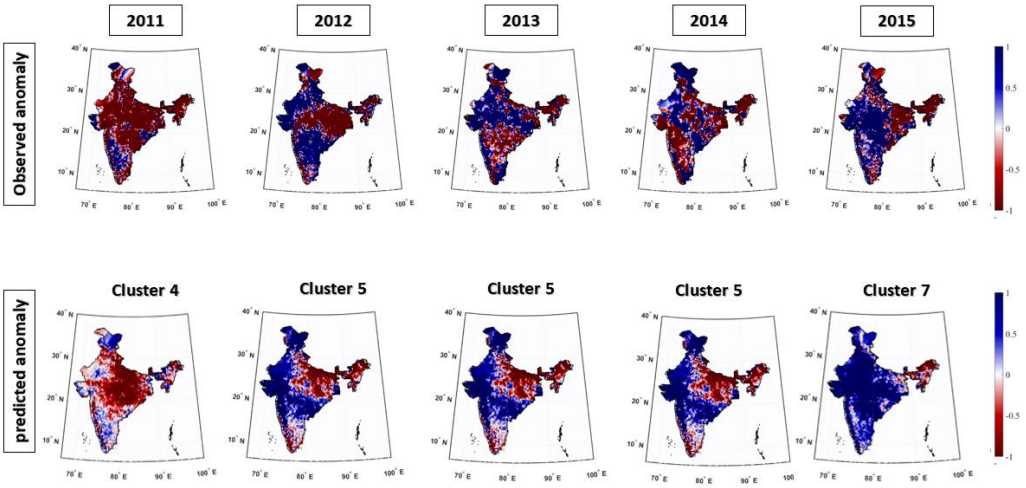


Figure 3:1Predicted and Observed anomaly of rainfall

As we compare predictions obtained from CART model, during 2011 the anomalies show a similar spatial pattern for central India, Observed model shows a negative anomaly over central India and similar pattern is seen in prediction model. Prediction for the year 2012 correspond to cluster 5 which is also quite similar to observed anomaly, western India show similarity in positive rainfall anomaly. However when 2013 comparison is made we don't see

similarity in southern India region, the only good prediction is that of western India as it matches observed 2013 anomaly.

For 2014 again we got cluster 5, here the observed anomaly is very heterogeneous while comparing with predicted model. On the other hand for 2015 we are getting cluster 7 and it shows similar observation as that of predicted model specially in western Ghats.

3.2 Results

To conclude with this method, we can say that to obtain the finest forecast with the present available CART algorithms, It is recommended to use the data with significant number of occurrences as training period as it will ensure that we can see distribution of anomaly. Training period is chosen as 1901 to 2010 that is 110 years. And then next consecutive 5 years are taken for forecast

We can also anticipate that, with the application of statistical forecast, improving of the training periods is an integral motive and depending on this further evaluation is carried out.

Chapter 4 Experiments carried out

In order to get a better forecast for Indian Summer monsoon Rainfall, a series of experiments were carried out. And after reviewing results of all these experiment we can finally conclude that which method bring about relatively closer forecast.

The experiment carried out are as follow:-

4.1 Prediction of ISMR using sst, humidity and wind data at different levels

Previously, CART was applied on cluster indices that are sst (sea surface temperature), wind, humidity and temperature at 850 hPa, while here in this experiment all these predictands are taken at different levels of 900 hPa, 950 hPa and 1000 hPa. Basically 850 hPa denotes the geo-potential height of the data/(Geo-potential height calculates approximate the height **of** a pressure surface above mean sea-level).

Difference between 900 hPa, 950 hPa and 1000 hPa and reason for selection

All these different geo-potential height has different altitude and they are as follow:-

Pressure	Altitude
900 hPa	988 m
950 hPa	1457 m
1000 hPa	1948 m

After carefully studying the above table, a distinct conclusion can be drawn. 1000 hPa is closer to the sea level as compared to 950 hPa and so we assumed that wind, temperature and other parameters near to sea level might show a significant change in our prediction model.

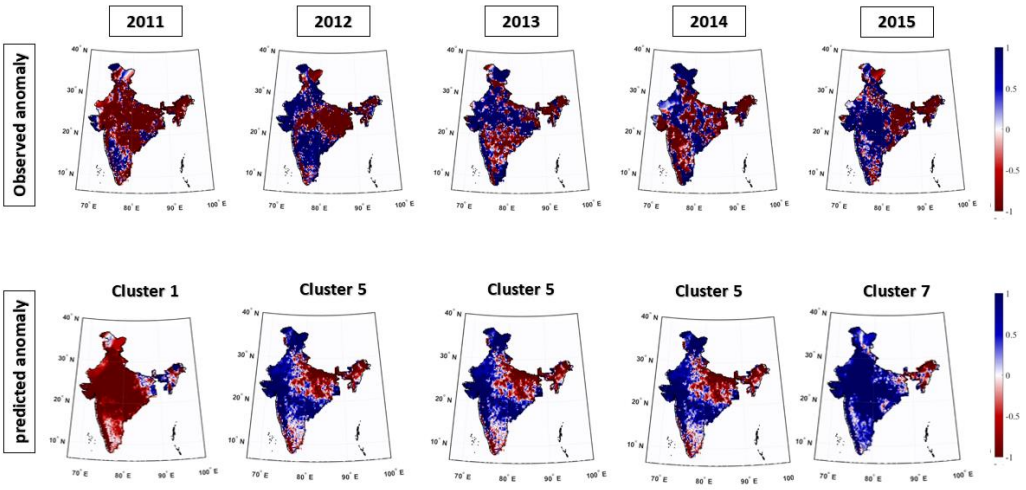


Figure 4:1 Results at 900 hPa

At 900 hPa, we observe all the predictands (sst, wind, humidity etc.) at an altitude of 988 above sea level and compared to figure 3.1 there is only one change which is in prediction of 2011, here the model predicts 2011 to be compared with cluster 1 instead if cluster 4. It is definitely a better prediction for central India but if we see western ghats, observation shows that west part of India has positive anomaly whereas the model predicts it have negative anomaly.

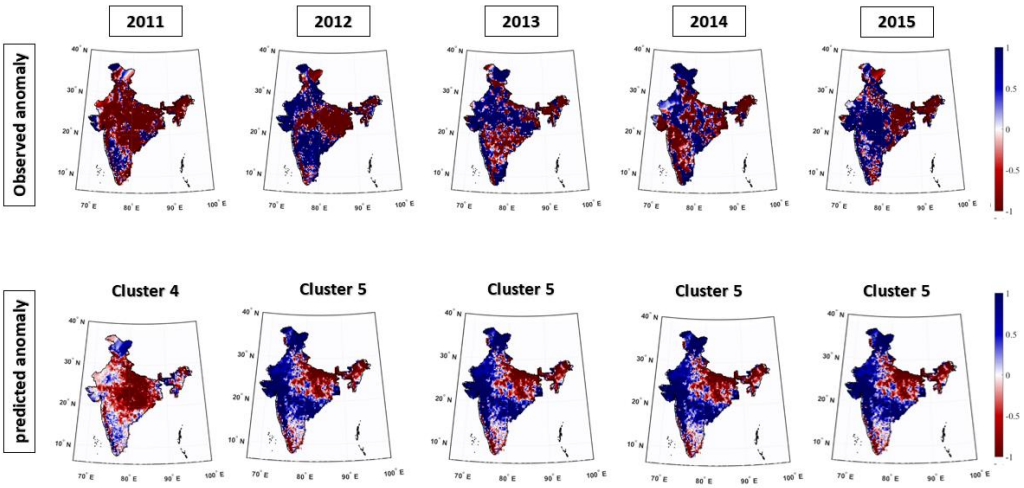


Figure 4:2 Results for 950 hPa

At 950 hPa, we observe all the predictands (sst, wind, humidity etc.) at an altitude of 1457 m above sea level and compared to figure 3.1 there is only one change which is in prediction of 2015, here the model predicts 2011 to be compared with cluster 5 instead if cluster 4. It is definitely a better prediction for central India but if we see southern region, observation shows that western and southern part of India has positive anomaly whereas the model predicts it have negative anomaly in southern region.

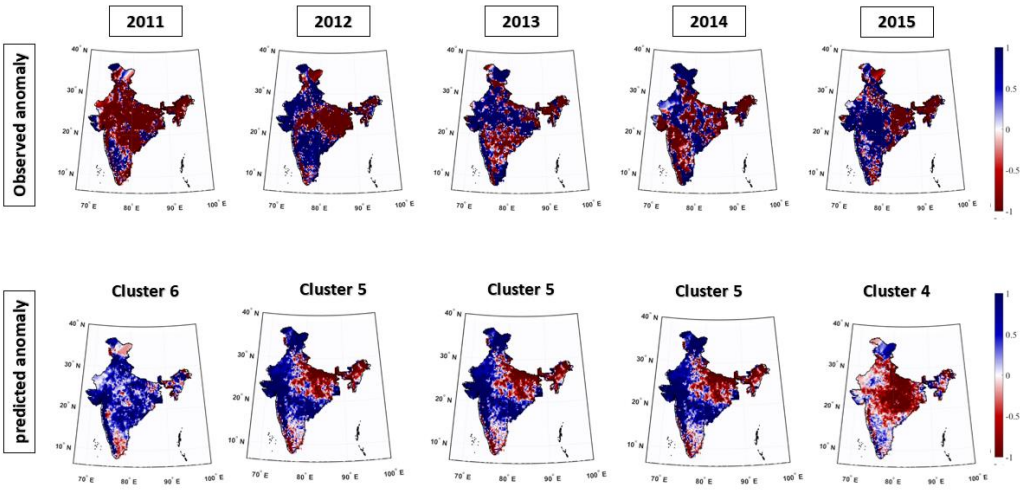


Figure 4:3 Results of 1000 hPa

At 1000 hPa, we observe all the predictands (sst, wind, humidity etc.) at an altitude of 1948 m above sea level and compared to figure 3.1 there is a change which is in prediction of 2011 and 2015, For 2011 observed anomaly shows cluster 6 which is not at all matching our observed model. For 2015 predicted model shows cluster 4 which when compared with observed anomaly does not match, hence taking 1000hPa for prediction is not appropriate.

4.2 Applying co-relation to oceanic regions

In all of the previous experiment we took 12-oceanic regions as input predictors but here we would choose only those predictors with have good co-relation with Indian summer monsoon rainfall

Next experiment was conducted with the aim to identify suitable predictors for the Indian summer rainfall, which can then be further used in statistical forecast models.

Here, the correlation is calculated for each predictor and then based on their value we would select top 5 region that shows maximum correlation. For the figure given below we have selection these given five regions namely nino3, nothpacific2, Atlantic Niño, north pacific, western Indian. Out of these five regions nino3 and nothpacific2 shows a positive value of correlation while the other show negative values.

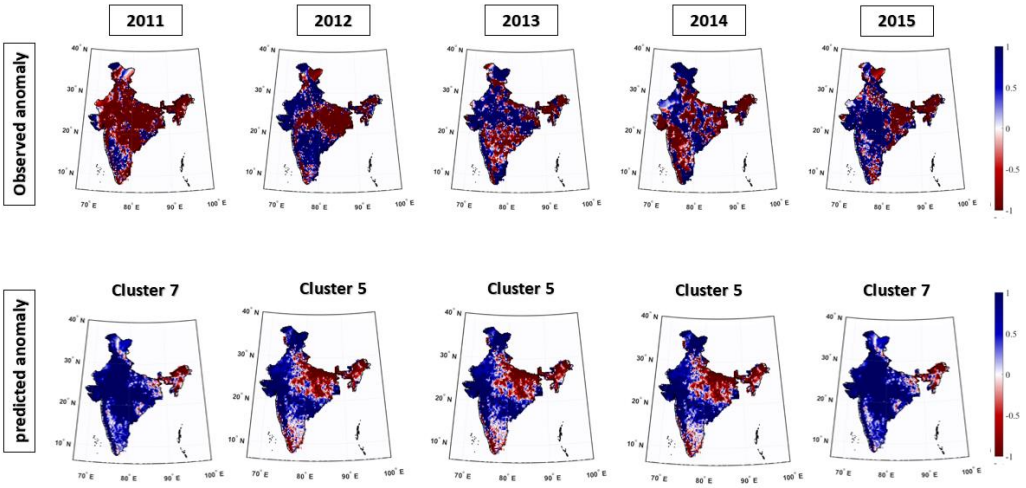


Figure 4:4 Results of 5-region correlation

After applying correlation to oceanic regions and selecting 5-region with maximum correlation, we plot anomaly obtained in test period and compare it with observed anomaly. First most noticeable thing is that for 2011 prediction is exactly opposite to that of observation and hence this prediction model fails to give us the required results

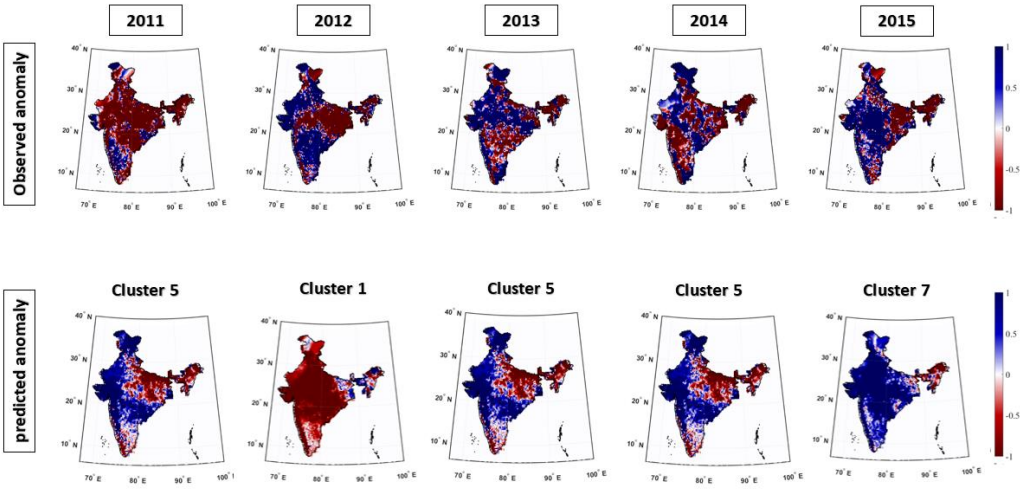


Figure 4:5 Results of 9-region correlation

After applying correlation to oceanic regions and selecting 9-region with maximum correlation, we plot anomaly obtained in test period and compare it with observed anomaly. First most noticeable thing is that for 2011 prediction is exactly opposite to that of observation, also same applies for 2012 and hence this prediction model fails to give us the required results

4.3 Using multiple regression technique

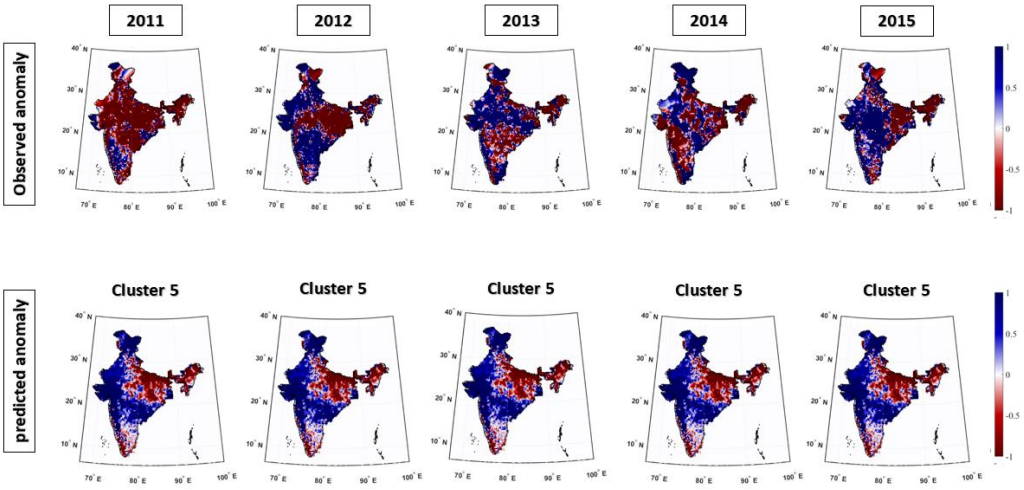


Figure 4:6 Results of multiple regression

For parametric and power regression forecast

- 12 regional and global data parameters
- divided into 4 categories (temperature, pressure, wind, humidity)
- parameters are chosen by physical linkage with monsoon

surprisingly multiple regression technique gives very homogeneous prediction which is not what we are looking for and so this method of prediction fails

4.4 Applying CART only on central region

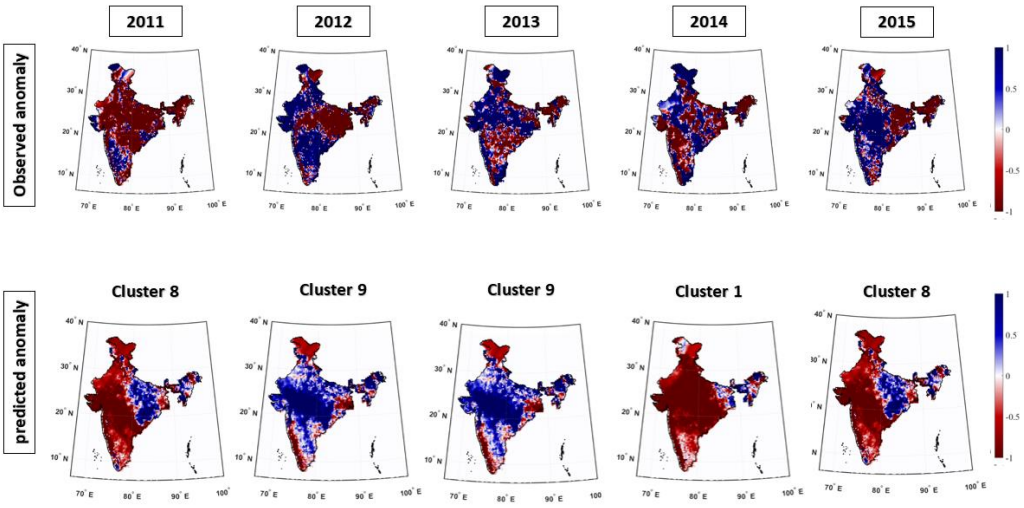


Figure 4:7 Results of applying CART on central region

For the last and final experiment we consider only the central region of India and so we will check if after applying prediction we can required results or not. Observe 2011,2014 and 2015 In all these year we are getting exactly opposite prediction than what is observed and so this method is no appropriate for prediction.

Chapter 5 Conclusion

Here, we are inspecting the spatiotemporal distributions of Indian Summer monsoon rainfall and tried to observe that if it is instantly related with the SST's through different regions of ocean taken in experiment. We are able to obtain 9-different temporal figure and to have a look over the diversity of June-July-August-September monsoon-rainfall within every clusters. These 9-clusters supply us with more data about particular region, which is possible to evaluate by taking the average mean of all zones of All India Monsoon Rainfall (AIMR). We have also discussed the large distance connection of ISMR with different parameters. Finally we can say that as mentioned by researchers that with using long-distance predictor we can get significantly better model of forecast.

Furthermore, in order to get better forecast and anomaly patterns we tried to conduct series of experiments and compared it with anomaly obtained using CART on 12-oceanic region, however we didn't get better results. To conclude we can say that results obtained from 'Prediction of spatial pattern of Indian summer rainfall using CART model' gave us the best prediction model.

References

- (n.d.)Retrievedfrom mausam.gov: https://web.archive.org/web/20110721181131/http://www.https://web.archive.org/web/20110721181131/http://www.mausam.gov.in/WEBIMD/organisation.jspv.in/WEBIMD/organisation.jsp
- *2019*. (n.d.). Retrieved from medanta: https://www.medanta.org/patient-education-blog/monsoon-illnesses-in-india-all-you-need-to-know
- B. N. Goswami, V. V. (2006). Increasing Trend of Extreme Rain Events Over India in a Warming Environment.
- Bell, M. S. (n.d.). *Climate Assessment for 1996*. Retrieved from https://www.cpc.ncep.noaa.gov/products/assessments/assess_96/india.html#:~:text=Indian%20summer%20monsoon,of%20their%20total%20annual%20rainfall.
- *How monsoon impacts the Indian economy*. (2010, june 02). Retrieved from rediff: http://business.rediff.com/slide-show/2010/jun/02/slide-show-1-how-monsoon-impacts-indian-economy.htm
- *Monsoon_of_South_Asia*. (n.d.). Retrieved from wikipedia: https://en.wikipedia.org/wiki/Monsoon_of_South_Asia
- Nathan S. Netanyahu, C. D. (n.d.). An Efficient k-means Clustering Algorithm: Analysis and Implementation.
- Subimal Ghosh, *. V. (2009). Trend analysis of Indian summer monsoon rainfallat different spatial scales. 6.