PREDICTION OF RAINFALL USING DATA MINING TECHNIQUES

BY

HIMEL DAS
ID: 183-25-698

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Masters of Science in Computer Science and Engineering

Supervised By

Shah Md Tanvir Siddiquee
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2019

APPROVAL

This Thesis titled "**PREDICTION OF RAINFALL USING DATA MINING TECHNIQUES**", submitted by Himel Das (ID: 183-25-698) to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 06 December.

BOARD OF EXAMINERS

(Dr. Syed Akter Hossain) — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
DaffodilInternationalUniversity

Dr. Md. Ismail Jabiullah — **Internal Examiner**
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

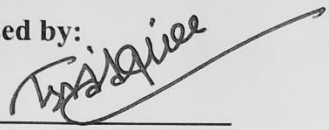
Dr. Sheak Rashed Haider Noori — **Internal Examiner**
Associate Professor& Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Mohammad Shorif Uddin — **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

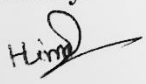
I hereby declare that, this thesis has been done by me under the supervision of **Shah Md Tanvir Siddiquee, Assistant Professor, Department of CSE** Daffodil International University. I also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Shah Md Tanvir Siddiquee
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Submitted by:



Himel Das
ID: 183-25-698
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes me possible to complete thesis successfully.

We really grateful and wish our profound our indebtedness to Supervisor **Shah Md Tanvir Siddiquee, Assistant Professor** Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of Computer Science to carry out this thesis. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Md. Zahid Hasan, Assistant Professor, Department of CSE, for his kind help to finish our thesis and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

The research for this paper focus on finding inter-relations between various climate ratio and predict rainfall therefore.

In addition, since precipitation is the detectable explanation for flood, our evaluation can help immensely in anticipating flood and masterminding a fitting risk the official's structure.

Being a riverine country, Flood happens in Bangladesh pretty much every other year.

Anticipating flood precisely can help us in working up our economy. Our assessment shows how the climate parameters (SOFI EI Nino) are obligated for critical precipitation in Bangladesh.

In spite of the fact that numerous different inquiries about on foreseeing precipitation have been directed utilizing other climate variables, the southern swaying record and the EI nino 3.4 shows.

For working up a relationship among precipitation, SOI and El Nino, we have applied Data Mining strategy.

The particular information calculations that we have executed in our paper are K- grouping. Decision tree and Regression model.

The yields of these calculations give us a clear connection among precipitation and the information parameters.

Executing our strategy on the dataset of precipitation for the recent years, our evaluated precipitation is nearly equivalent to the real ones of those years.

In this way, in organizing a potential gauge model for Bangladesh, our work can accept a basic activity as a result of its high effectiveness.

Foe working up a relationship among precipitation, SOI and Nino, we have applied Data mining strategy.

TABLE OF CONTENTS

CHAPTER-1

INTRODUCTION

Bangladesh is a riverine country. It falls under the dynamic tempest locale of the world. Among the absolute territory of 1,47,570 sq. kms of this deltaic land, the waterway region alone holds 9399 sq. kms (DOE 2002). Having an agro based economy, practically 40% of the all-out populace of this nation are identified with horticulture expertly. Subsequently, precipitation has a huge impact in the economy of Bangladesh. The inconsistent substantial precipitation events may impact situations, cultivating, sustenance security, urban leakage. Water openness, water quality and prosperity and business of people the country. On the rainstorm of 2007, the nation got extraordinary substantial precipitation together with the beginning of flooding by the Himalayan streams which prompted extreme flood in practically 50% of Bangladesh. Generous precipitation at fourteenth September of 2004 and 28th July of 2009 achieved outrageous urban flooding and waste stop up in the capital city of Dhaka. Landside casualty likewise was seen during 2008 and 2010 because of overwhelming precipitation occasions over the nation. Ongoing thinks about additionally recommend that the recurrence and extent of substantial precipitation occasions have just been expanded under the a dangerous atmospheric deviation situation including the high height regions of Bangladesh. It is hence, basic to have an effective precipitation estimation framework for this catastrophic event inclined nation.

The El Nino and SOI (Southern Oscillation Index) are characterizations of comparative procedures identified with SST (Sea Surface Temperature) and Pressure peculiarities for ocean surface and surface degree of the environment and together it is known as ENSO (El-Nino and Southern Oscillation). ENSO is a prevailing method of interannual atmosphere changeability that creates from air-ocean collaborations in the tropical Pacific, however impacts climate designs comprehensively. Since Bangladesh is such a nation which is somewhat a long way from the region where these climatic components are prevailing, discovering co relations among precipitation and the elements has consistently been a test. In spite of the fact that the vast majority of the parameters are feebly identified with precipitation in our nation, ENSO shows that it has a critical task to carry out behind the sporadic precipitation occurrences in this nation.

The examination directed in our paper chiefly utilizes the two above notice parameters as information and discover the evaluated precipitation. Since these two components happen sometime before the precipitation, the expectation can be made early and helpfully. We have applied our calculations of Decision tree, K-implies grouping and relapse model to assess the precipitation of two stations, Dhaka and Khulna. We have chosen these stations dependent on their connection factor with ENSO. Running our calculation on the years 1980-2016, our yield shows that the precipitation evaluated by means of our proposed technique coordinates indistinguishably with the genuine precipitation information. We have additionally thought about the exactness of these three calculations with respect to precipitation estimation.

1.2 Objectives

The main reason behind this thesis to predict rainfall using data mining techniques. The objective of this thesis was to make the farmers informed about the rainfall possibility long before it takes place. As a result, they will be able to face the result and the damages will be reduced. Mainly this thesis will contribute to the agricultural sector of Bangladesh. Each year several areas in Bangladesh are drowned underwater due to heavy rainfall. As a result, farm land are severely damaged and many cattle farmers suffer from this. Several housings are also collapsed in storms and floods affecting daily lives. Most of the sufferers are the people who are living in the river side. The sufferers are not only the people of those areas, but also the rest of the country as such situations lead to food crisis. In addition, this thesis has been done using different algorithms such as ‘Linear Regression’ and ‘K-means clustering’ which can be helpful to others who will work with similar type of problems using data mining and these algorithms. Like that algorithms will enable the farmers know when any tropical disaster will take place. Not only that, the farmers can ready a month or a week early before they can see their hard work being drowned away by cyclones.

We went to Bangladesh Meteorological Department (BMD) and talked to several experts as regards rainfall. They told us about the main areas where flood affects the most. So, we were mainly focused on the stations of those areas.

In our future work, which will be mentioned later in the paper, we are going to introduce an app for that purpose so that they can work without depending on the weather station in person.

- General objective:

i) The general objective of the research is to examine the relationship between climatic indices and product precipitation.

- Specially The research seeks:

i) Two know how the climatic parameters are liable for significant precipitation.
ii) To know the relationship among precipitation.
iii) To investigate the temperature, precipitation and wind speed of various month.
iv) To investigate how rainfall, effect the economy.

1.3 Motivation:

Ours is a developing country. One of the key factors behind the economic growth of our country is agriculture. Since agriculture directly depends on rainfall, it is crucial to anticipate it earlier and take necessary measures so that it does not spread any adverse effects. With an intention of helping our agriculture, we have done our study in this area and came up with this research work consequently. Another significant reason for this research to be conducted is flood. For the past four decades, flood has been a significant thorn on the way of our economic advancement. Almost every year, several parts of this country have to stay underwater for a long time which damages the crops and washes away all the hard work of the farmers. Its prevention can only be done by anticipating it earlier and taking measures accordingly.

1.4 Problem statement:

In our exploration, about thousand information occasions containing qualities like precipitation, temperature, gaseous tension, year and so on have been utilized and broke down with the assistance of various calculations for getting an appropriate forecast. Gathering every one of the information was very troublesome as we needed to go to the Bangladesh Meteorological Department (BMD) for precipitation information. We went to the middle a few times and attempted to converse with certain specialists and there were a few times they couldn't set aside a few minutes. After that we gathered other information from various solid locales. Getting increasingly exact information was significant for us. Discovering the site where we could get the real information was troublesome. In addition, none of us was so master in programming. Subsequently we needed to experience intense occasions to comprehend the codes well and work dependent on that. Giving contribution to a

precise way and comprehending what is the normal yield and what we are getting was likewise a significant errand for us. We needed to experience numerous preliminaries and blunders in such manner during working with those calculations. Broad measure of research was done to comprehend the fundamental of the calculations. Besides, figuring out how to work with huge information in exceed expectations was somewhat new experience.

CHAPTER 2

LITERATURE REVIEW

2.1 Background Study

As per AMS (2014), all strong or fluid watery particles that start in the environment and fall into the earth surface are precipitation. Precipitation, hail, day off, or sheets are the principle type of precipitation. Precipitation can be considered as the measure of fluid precipitation that arrive at earth surface and gathered by downpour check. Precipitation is the most prevailing component of the atmosphere in Bangladesh. The nation has a tropical storm atmosphere with high measure of precipitation. Unmistakable occasional example exists in yearly precipitation cycle which is more conspicuous than temperature cycle.

Determining overwhelming precipitation occasions has been a difficult activity for Bangladesh because of its powerless relationship with the central point behind south Asian storm. There are a few techniques accessible to estimate precipitation occasions.

Rahman et al. (2013) utilized basic relapse model to gauge summer rainstorm precipitation over Bangladesh.

Numerous territorial and occasion-based investigation has been done as such far utilizing NWP models to improve the high serious precipitation gauge. And in Bangladesh, these studies have been done more seldom. Till date, only a few systems have been designed for rainfall estimation. Among them, data mining has been a major tool since most of these systems have been implemented using one or another data mining model.

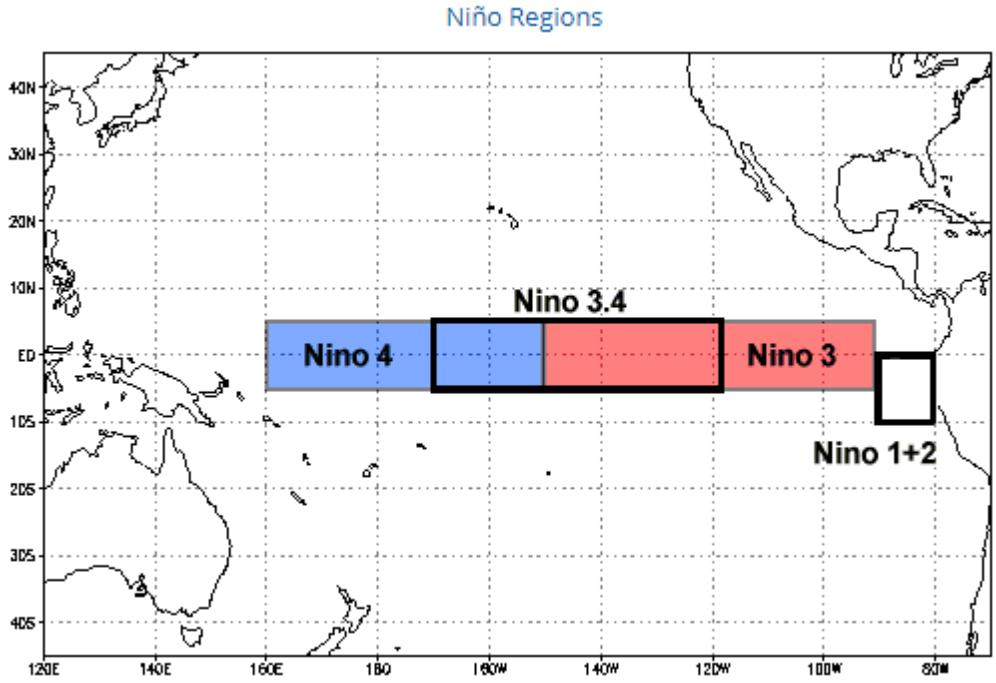
2.2 Methodology

ENSO

The episodic inconstancy in sea surface temperature, Known as EI Nino and the pneumatic force of the superimposing climate that is, Southern Oscillation all through the tropical Pacific Ocean joins to frame ENSO.[1].

El Nino arrives at its full quality toward the finish of consistently as a rule of every year usually. It is an event in the tropical Pacific Ocean discounted as five successive multi month running normal of ocean surface temperature (SST) oddities in the Nino 3.4 sector which is underneath or more the

edge of +0.5 degree celcius to - 0.5 degree celcius. Maritime Nino Index (ONI) is the standard estimation [2].

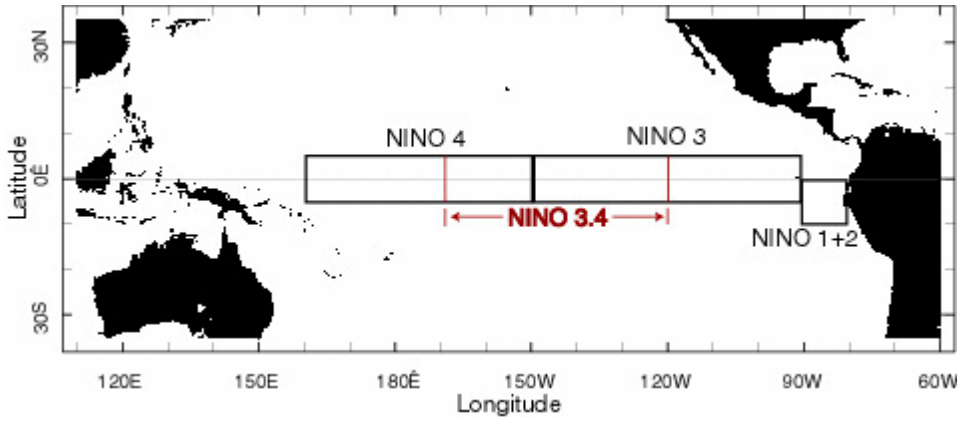


The Southern oscillation is to a greater extent a double pinnacle variety in ocean level barometric weight between two stations- Darwin and Tahiti. It is listed in the Southern Oscillation Index (DOI) which is essentially the change between lower pressure over Darwin and higher weight over Tahiti. The weights advance a gather together of air from east to west that brings out warm surface water towards the west makes it downpour in Australia and the western Pacific. Heavy precipitation can cause floods when pressure difference lowers in parts of western Pacific. Southern Oscillation is emphatically unpredicted with EI Nino conditions [1]

The two ingredient of ENSO which are sea rear temperature and atmospheric pressure are sharply related although its specific initiating causes are not fully understood. During the El Nino case, the eastern winds slow down by the central Pacific. This thusly eases back the sea ebb and flow that draws surface water away from the western shoreline of South America and diminishes the upsurge of chilly, supplement rich water from the more profound sea, squashing out the thermocline and enabling warm surface water to make in the eastern piece of the bowl. There is a relation among the Southern Oscillation and rainfall which is also clear in the amount of deep energy waves that leaves the atmosphere. A great deal of long radiation waves exists under the clear skies which are

free into the atmosphere from the rear and release into space. Under the fuzzy skies, some energy is denied from fleeing. These activities are measured using satellites.

To monitor the ENSO condition, we first need to focus on the sea surface temperature (SST) in 4 geographical areas of the equatorial pacific. SST exception are amounted to be greater than O.5 degree Celsius in the Nino 3.4 region and are a sign of ENSO warm state condition which is El Nino condition. Whereas, exception less than or equal to -0.5 degree Celsius which are soothing and are called La Nina condition. If the Oceanic Nino Index (ONI) views hot or cold phase condition in less than or equal to five sequential values, it occurs an El Nino or La Nina event [1].



The reason why we have chosen Southern Oscillation Index instead of other climate exponent like Arctic Oscillation (AO), Antarctic Oscillation (AAO), North Atlantic oscillation (NAO) etc. is because these indices have a correlation coefficient of zero when conducting simulation.

2.3 Related work

The paper [4] describes a model based on K-means clustering technique coupled with a supervised data classification technique, namely Classification and Regression Tree is used for generation of rainfall states from large scale atmospheric variables in a river basin. The K-means clustering is used to derive the daily rainfall state from the historical daily multi-site rainfall data. The optimum number of clusters in the observed rainfall data is obtained after application of various cluster validity measures to the clustered data. The model is then trained to establish relationship between the daily rainfall state of the river basin and the standardized, dimensionally-reduced National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis climatic data set. The paper [1] describes empirical method technique belongs to the regression approach which try to make a short-term forecast of rainfalls over specified region in

our state. The paper analyses the three months rainfall data of particular region for five years. Multiple linear regressions are used to predict the rainfall using the previous year's data from the corresponding time period.

CHAPTER 3
DATA COLLECTION AND PREPROCESSING

3.1 Data collection

Data collection is a principle essential piece of implementing mining strategies, for this test, a thermometer, barometer, hygrometer, anemometer and data recording systems was used. Data recording framework gives climate information to exceed expectations in forbidden structure. "Data record dependent on an advanced processor which is utilized by the built-in sensor or an outside instrument and sensors related with position information of the hour of the electronic contraption can normally records data of 24 hours.

This was the main and most important benefits of data recorder. It was used to collect weather data from local stations at Chattogram to a devoted lab PC, then copy the transferred weather data to an Excel spreadsheet and recorded on daily basis along with monthly basis to identify data.

3.2 Data pre-preparing

The Data pre-processing is the subsequent stage of information mining after gathering of information. Difficulties in temperature, precipitation and wind velocity information; learning revelation procedure is confronting poor information quality. Along these lines, the information is pre-prepared to evacuate commotion and undesirable information. Pre-treatment means thinking the expulsion of several undesirable factors from the information, while the information preprocessing incorporates various means:

3.3 Data scouring: It's where commotion and unessential data is expelled. Data decontamination methodology are executed to round out default qualities and to kill clamor in perceiving exceptions and to address data anomalies.

3.4 Data reconciliation

It's perceived as the Data transformation; in this period, the appropriate type of data is changed over for the system of data mining by decrease of data and development of attributes. Device can naturally gather and records information of 24 hours.

3.5 Disclosure of Knowledge

For the verifying of data, the different Data mining systems are executed that are Outlier-Analysis, Clustering and Classification in WEKA (Data Mining Software).

That is also known as the data association. This is selected form of data into a suitable data mining stage. Save the data files in comma-separated by value (CSV) format of file and data set was standardized to reduce the data scaling.

CHAPTER 4
ANALYSIS AND RESULT

4.1 Analysis of Result

The qualities under Analysis, for example, temperature, precipitation and wind speed were examined relying upon the mining results.

4.2 Proposed approach

In this research, different data mining techniques are applied. Firstly, the K-means clustering algorithm was applied on the given data set which was then altered into appropriate form from unstructured data format after the stage of preprocessing. Where 70% of data was taken as training data and remaining 30% was testing data. The model of training methodology was shown in figure 3.

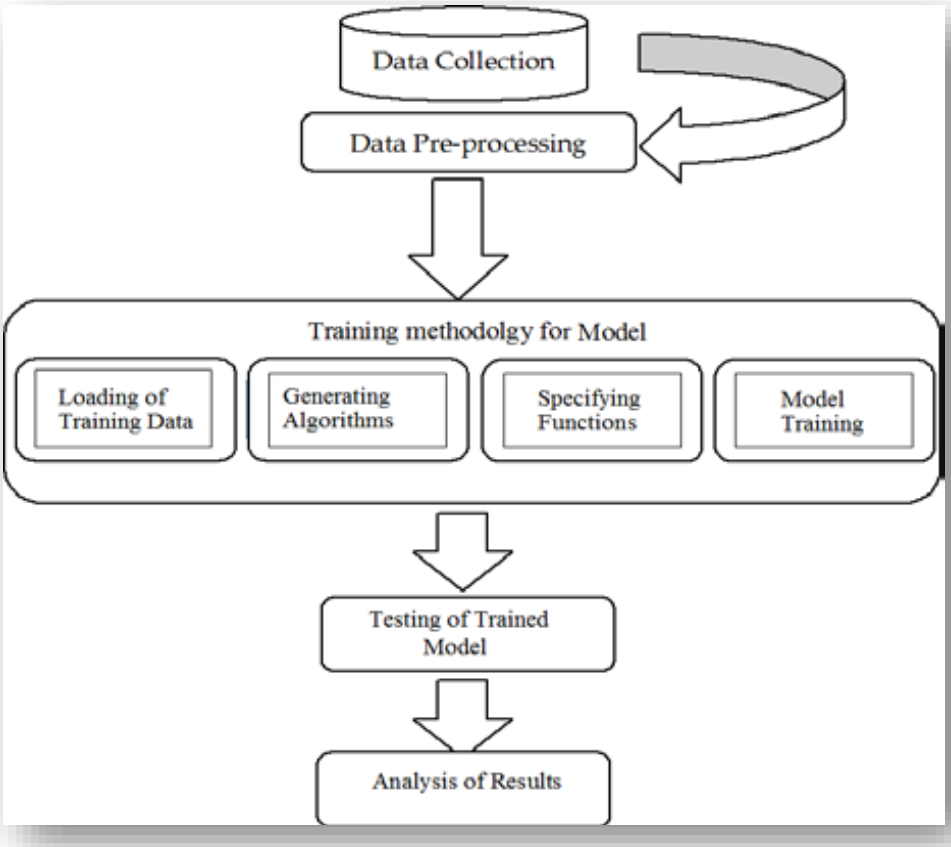


Figure 3: Training methodology of the model

4.3 Results and Discussion

There was used the Weka environment that has 4 applications which are Explorer, Knowledge Flow, Experimenter and Simple CLI. The collected data-set was changed in a file of extension ". arff" and loaded in the environment of Weka. First, the attribute reduction was used for data preprocessing then the simple k-means clustering and j48 algorithm were used.

4.4 K-means clustering:

Dhaka June:

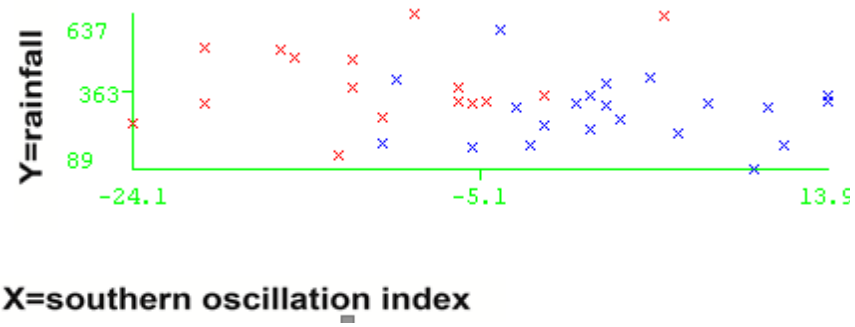


Figure 1

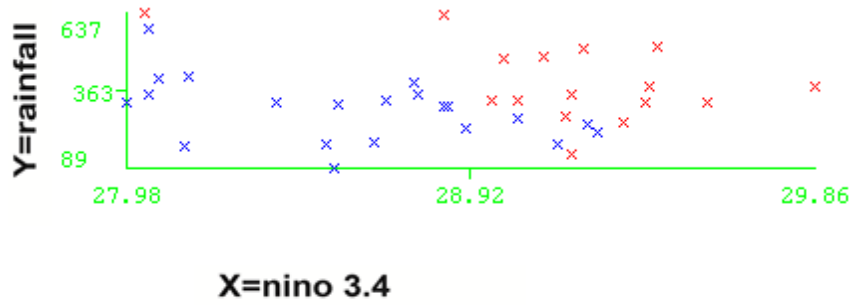


Figure 2

Khulna June:

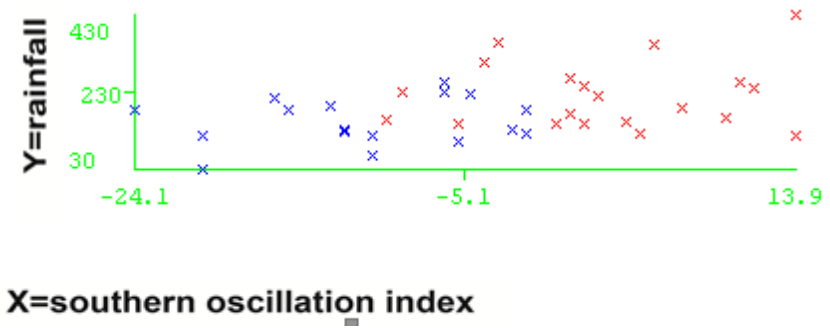


Figure 3

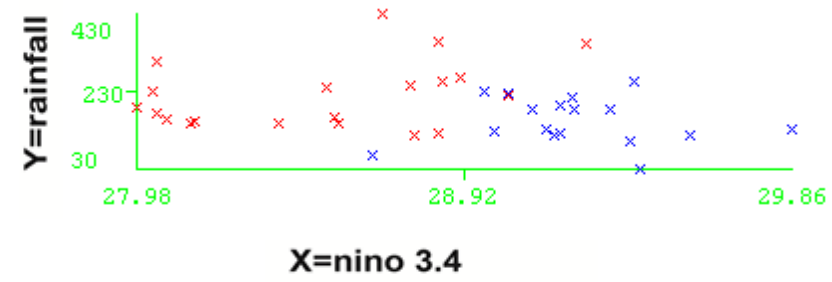


Figure 4

Bogra June:

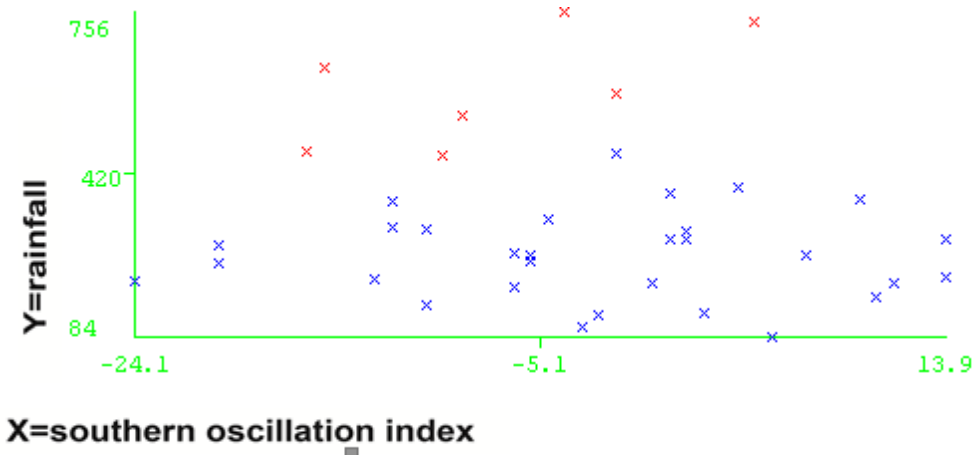


Figure 5

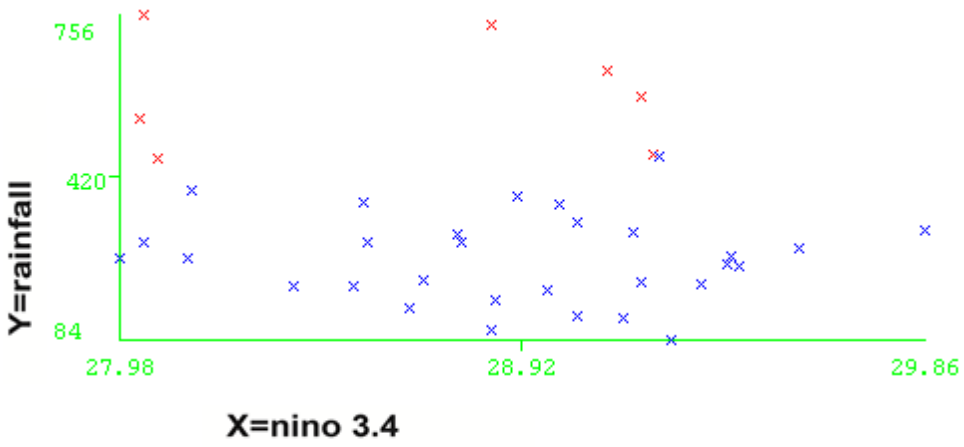


Figure 6

Barisal June

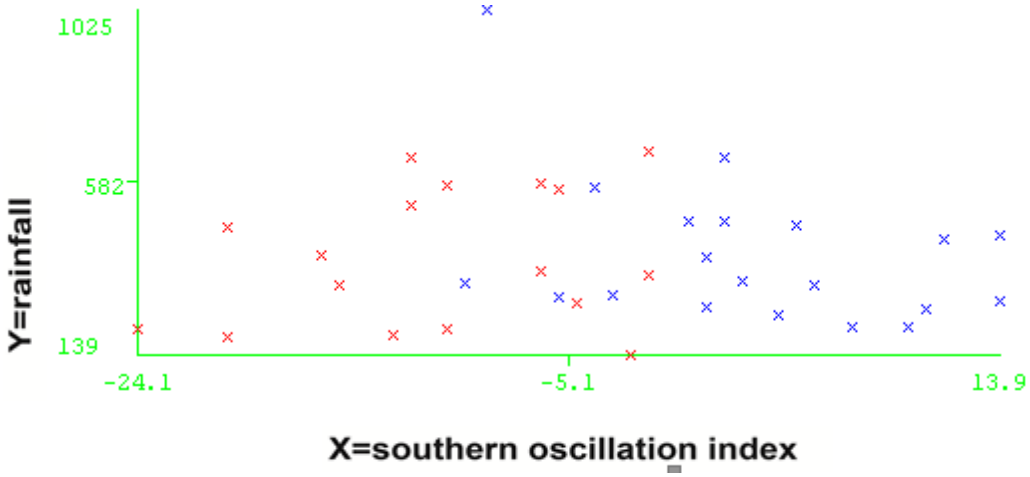


Figure 7

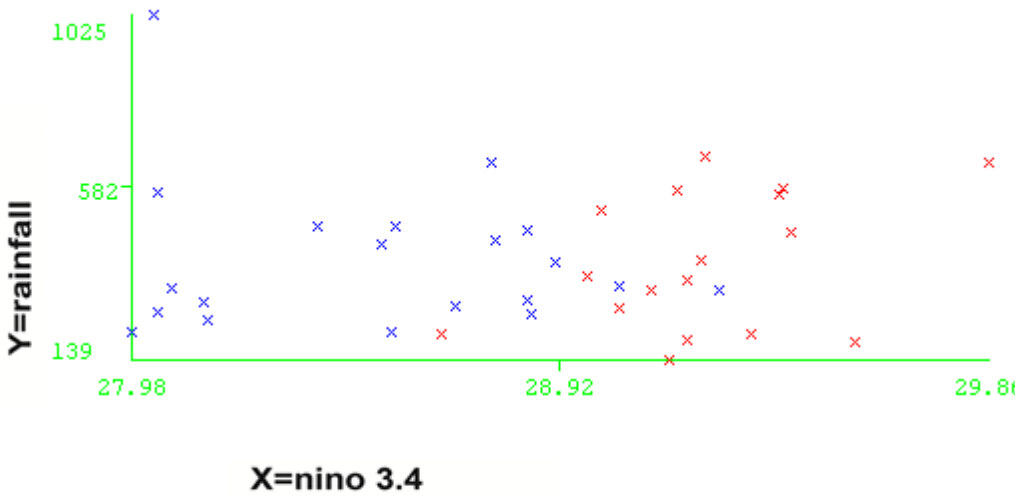


Figure 8

Bhola June:

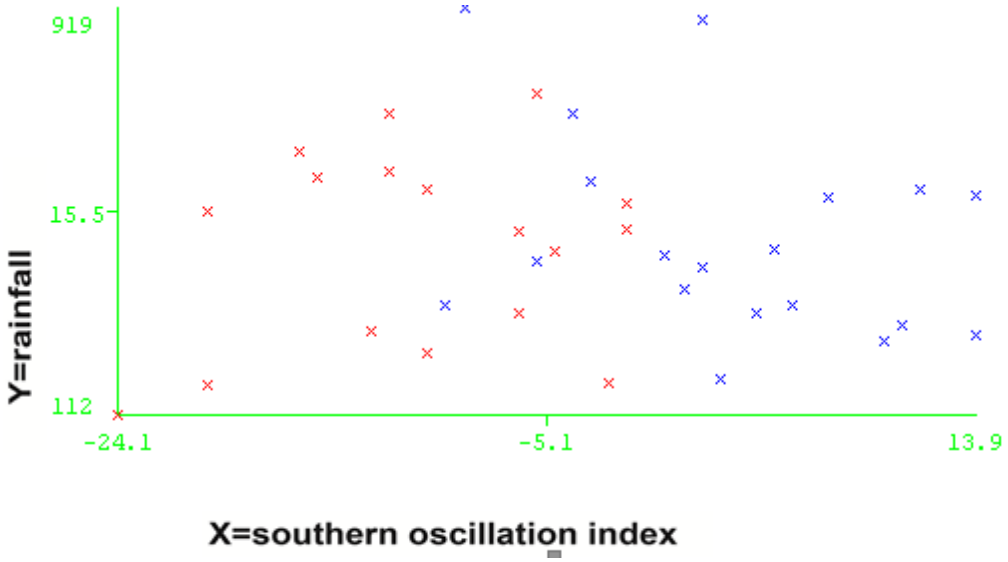


Figure 9

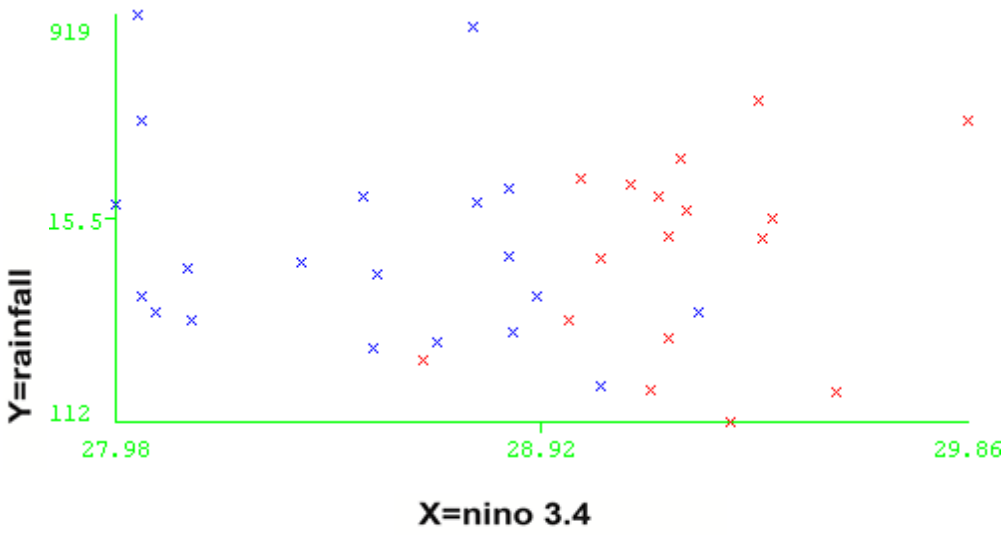


Figure 10

Cox's Bazar June:

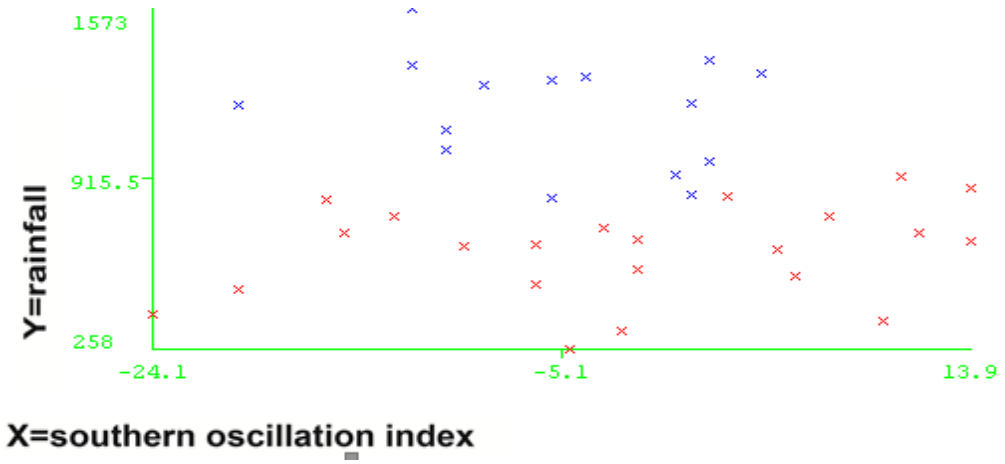


Figure 11

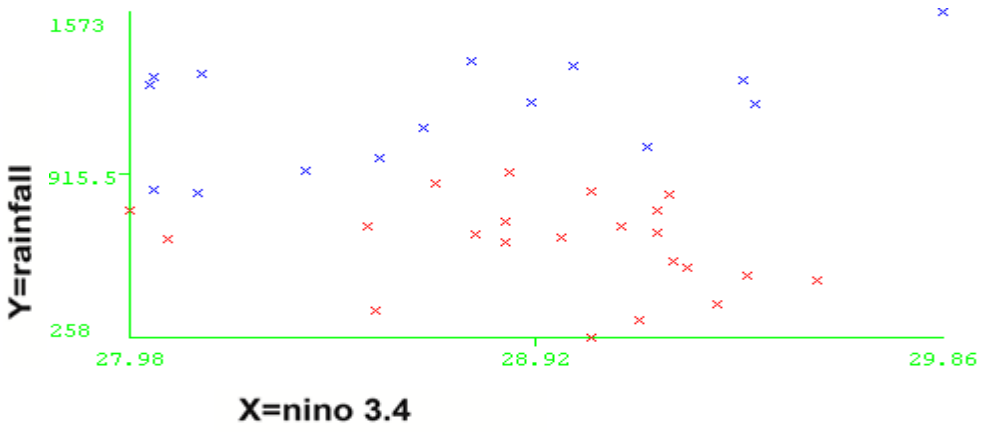


Figure 12

Dhaka July

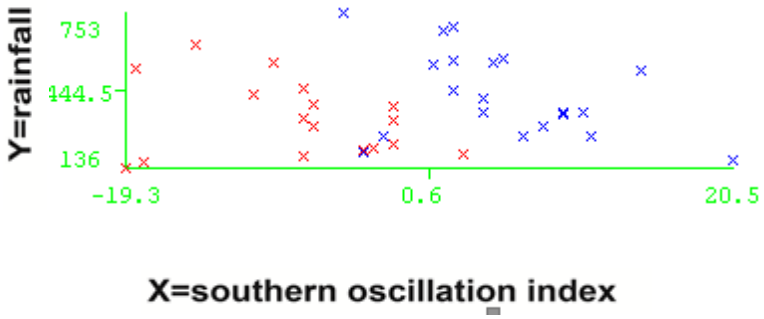


Figure 13

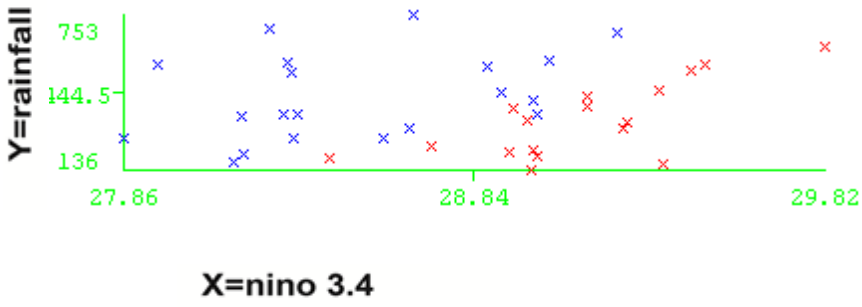


Figure 14

Khulna July:

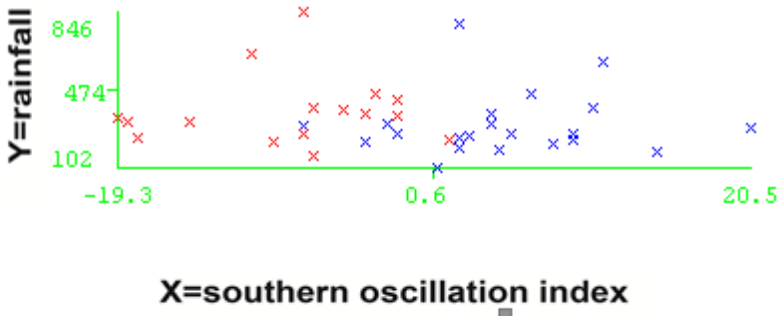


Figure 15

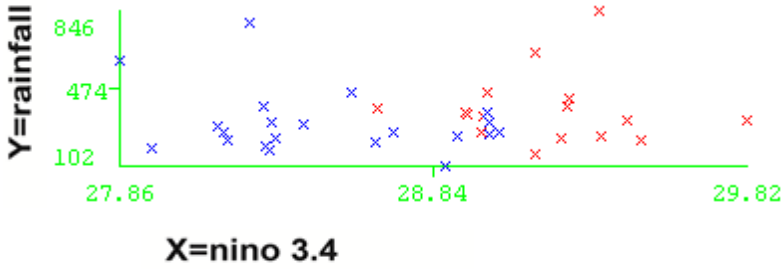


Figure 16

Bogra July:

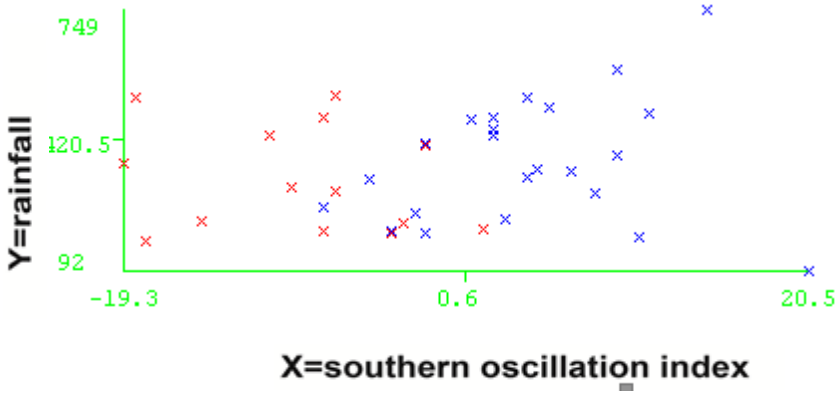


Figure 17

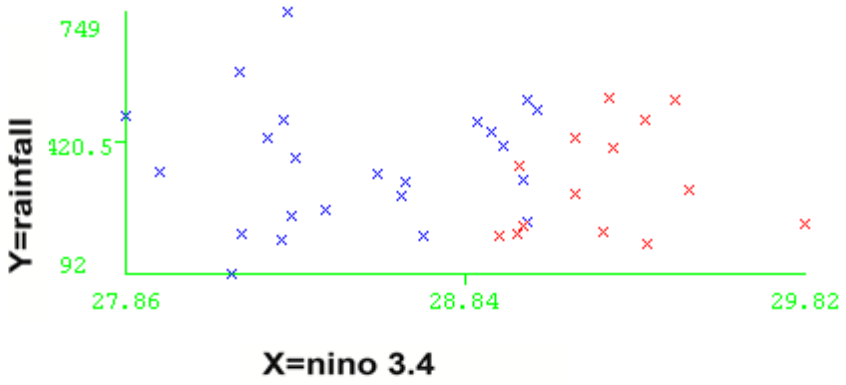


Figure 18

Barisal July:

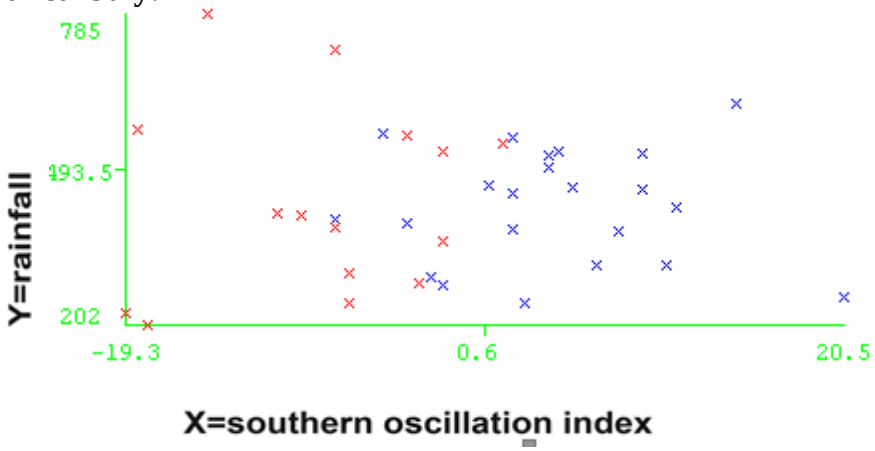


Figure 19

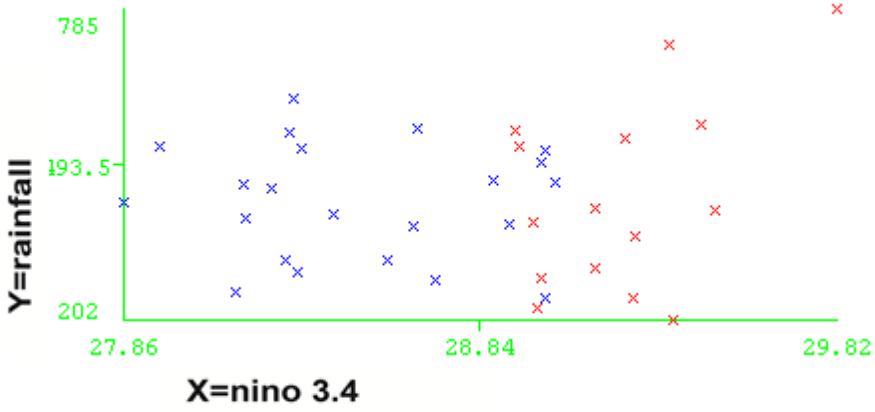


Figure 20

Bhola July:

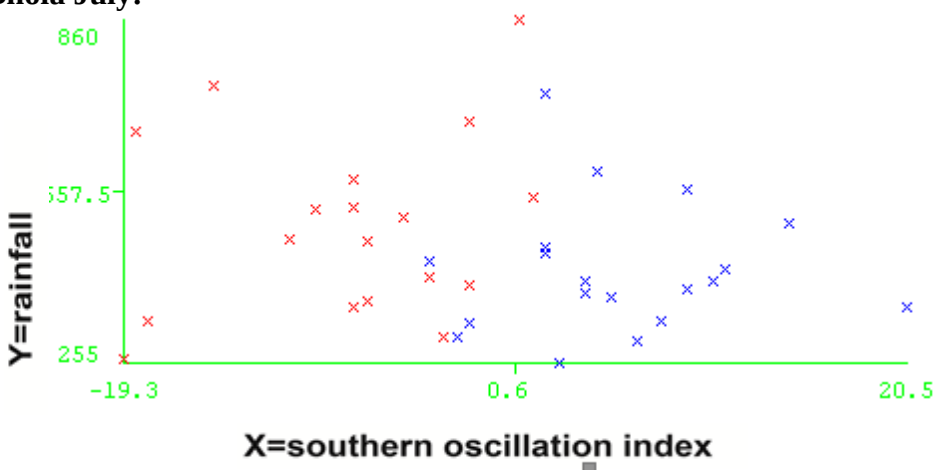


Figure 21

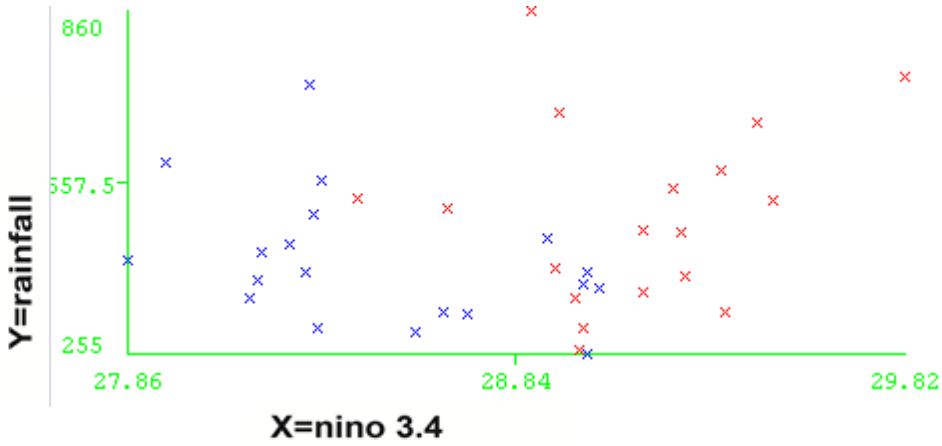


Figure 22

Cox's Bazar July:

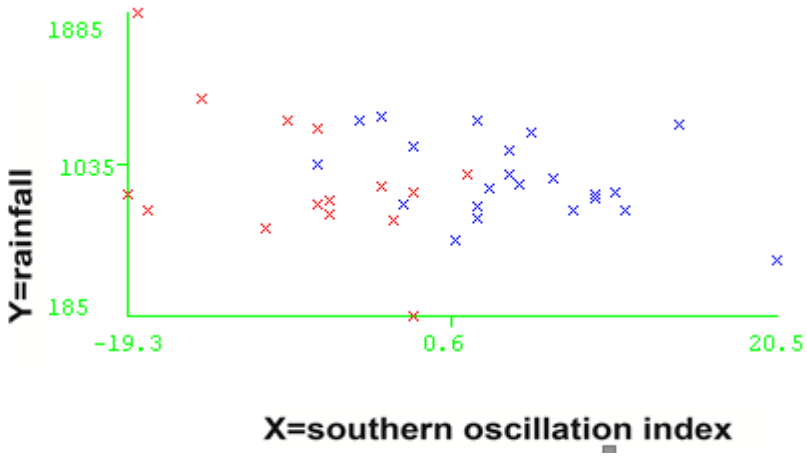


Figure 23

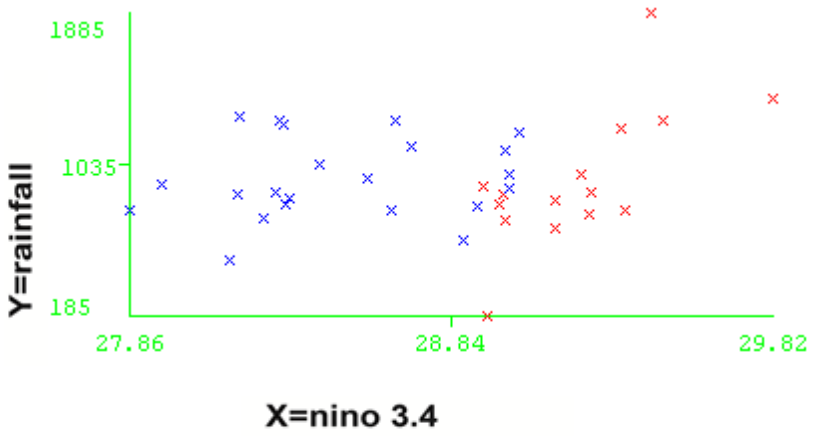


Figure 24

Dhaka august:

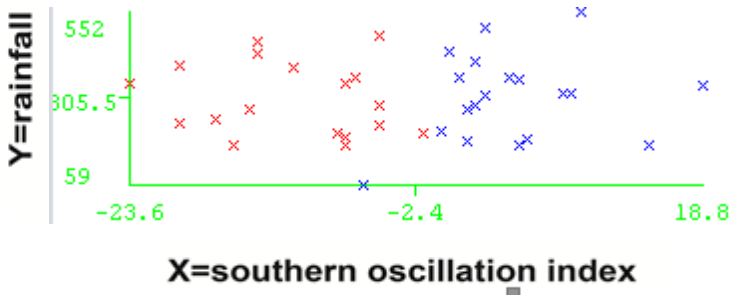


Figure 25

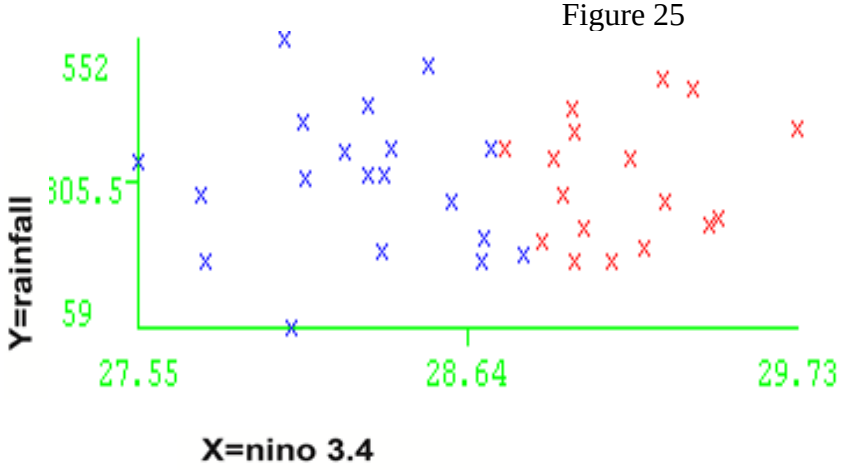


Figure 26

Khulna August:

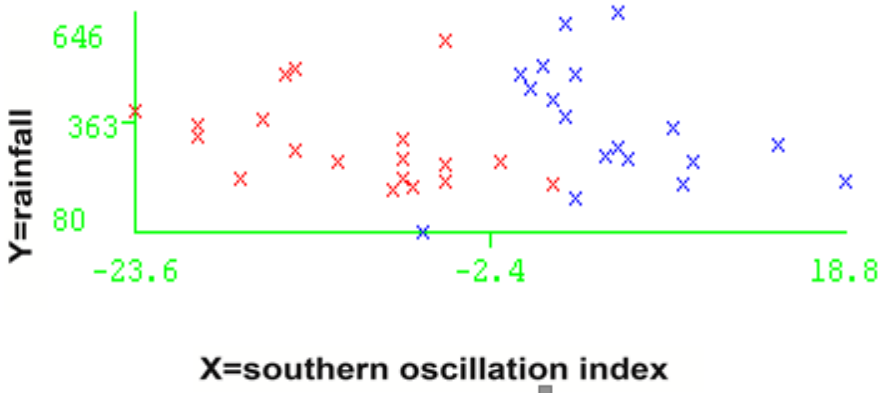


Figure 27

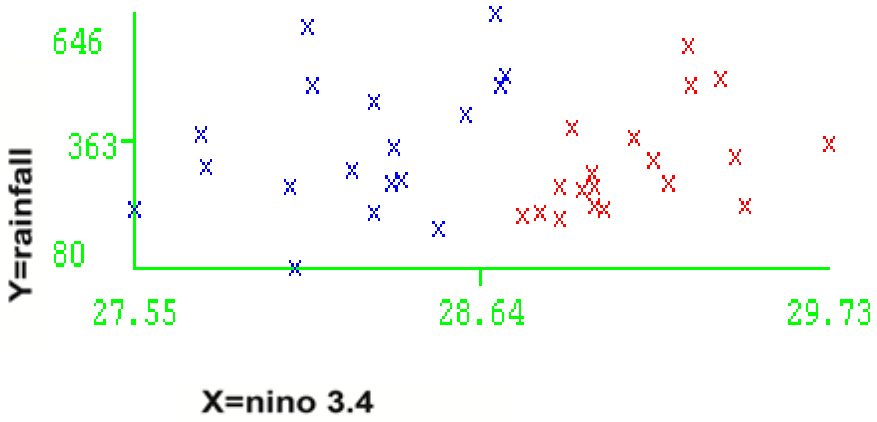


Figure 28

Bogra August:

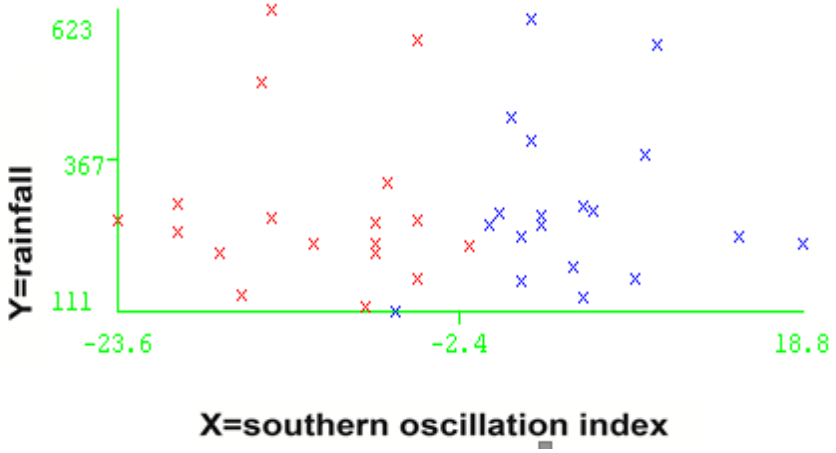


Figure 29

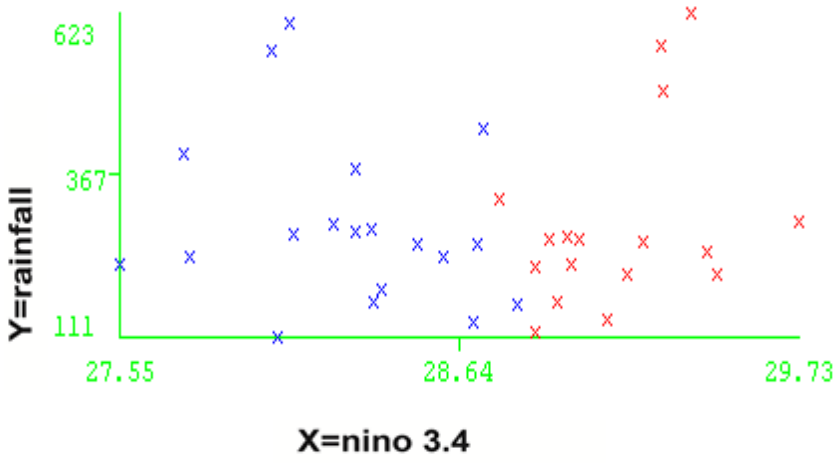


Figure 30

Barisal August:

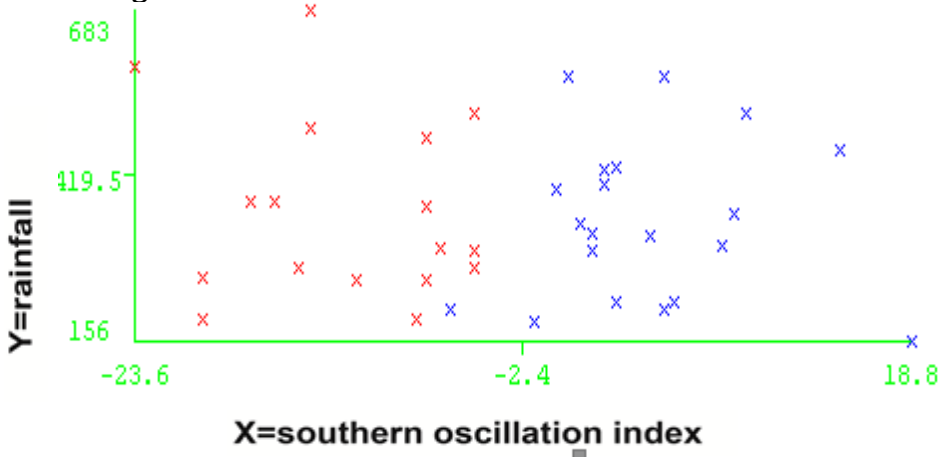


Figure 31

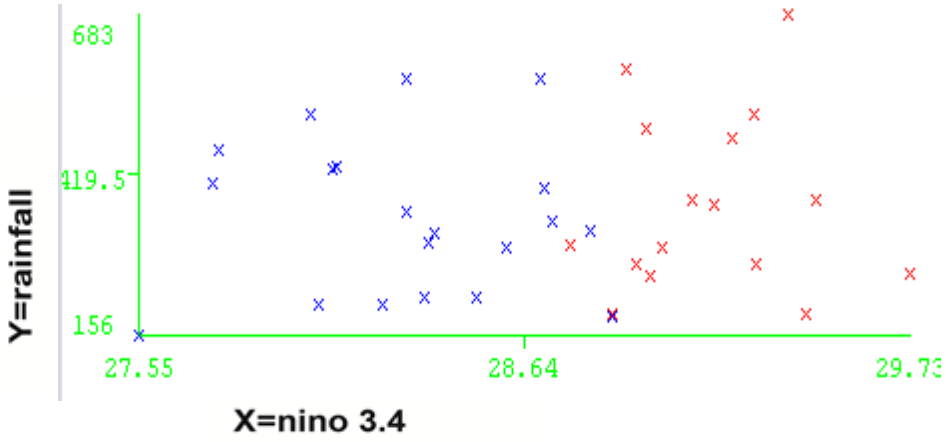


Figure 32

Bhola August:

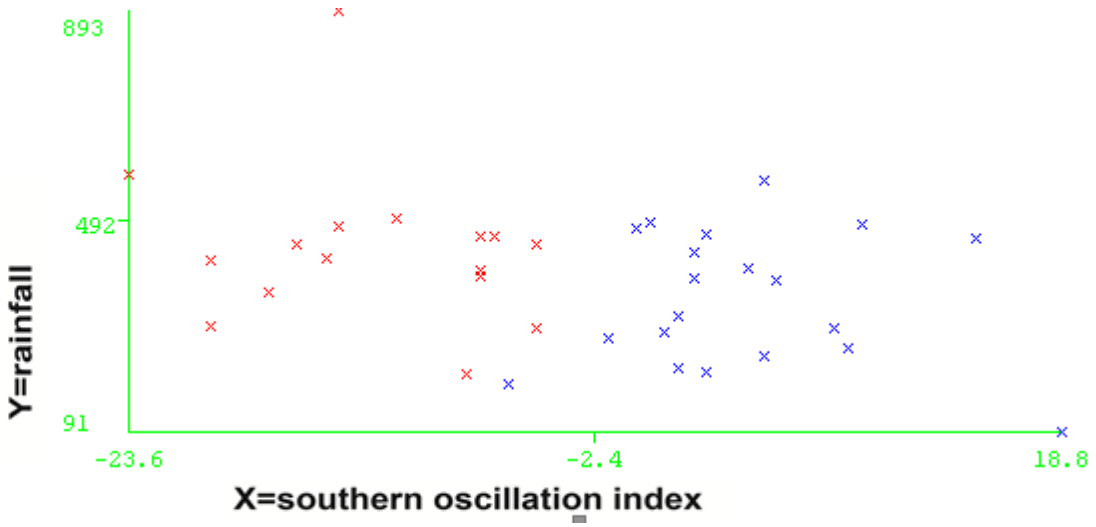


Figure 33

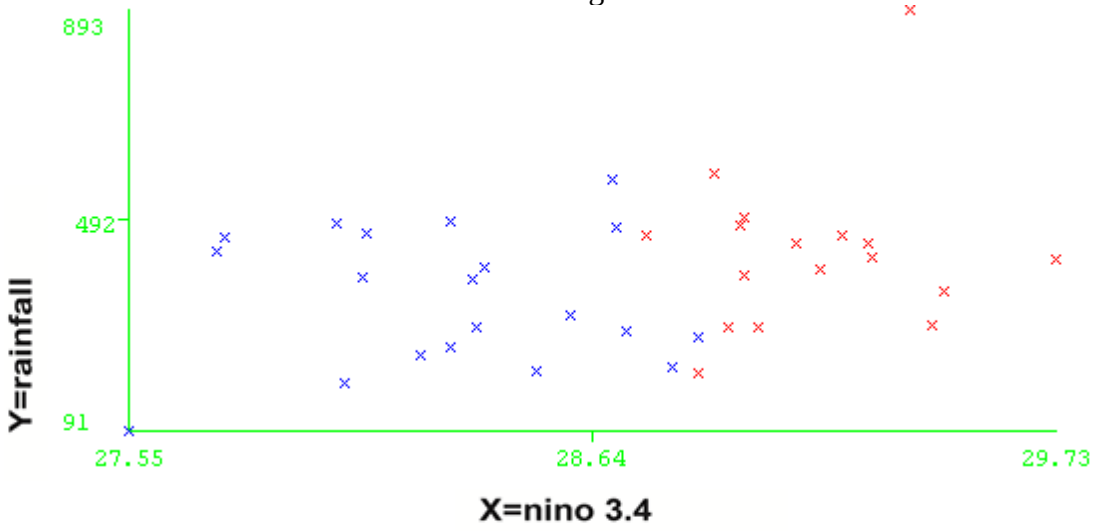


Figure 34

Cox's Bazar August:

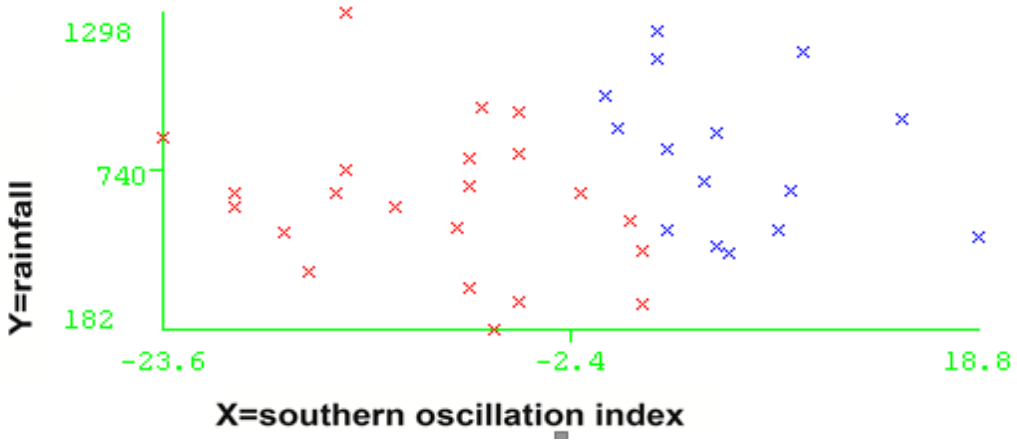


Figure 35

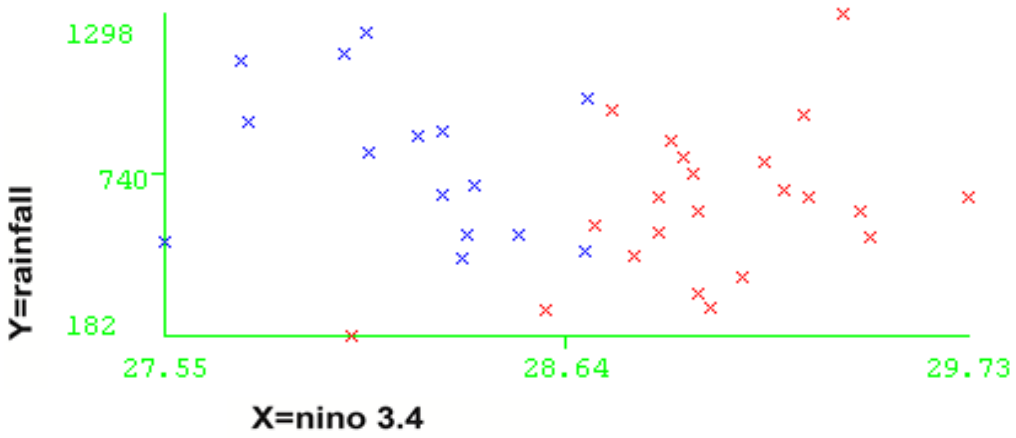


Figure 36

Dhaka September:

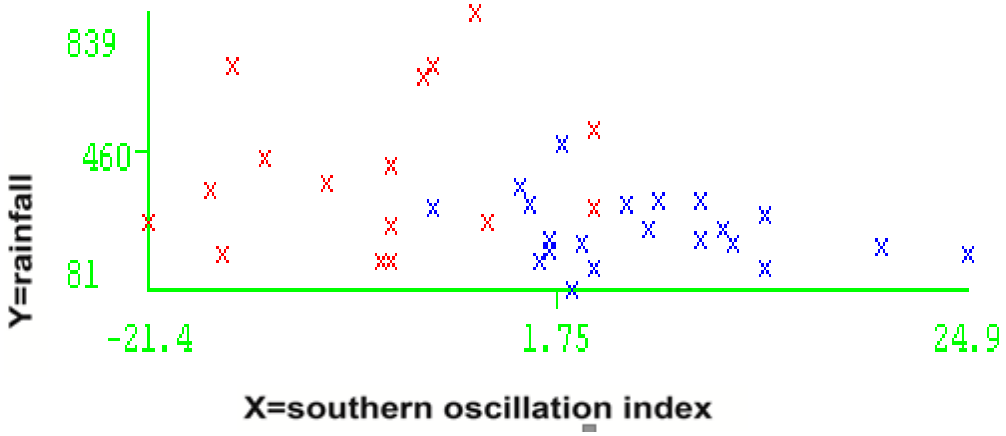


Figure 37

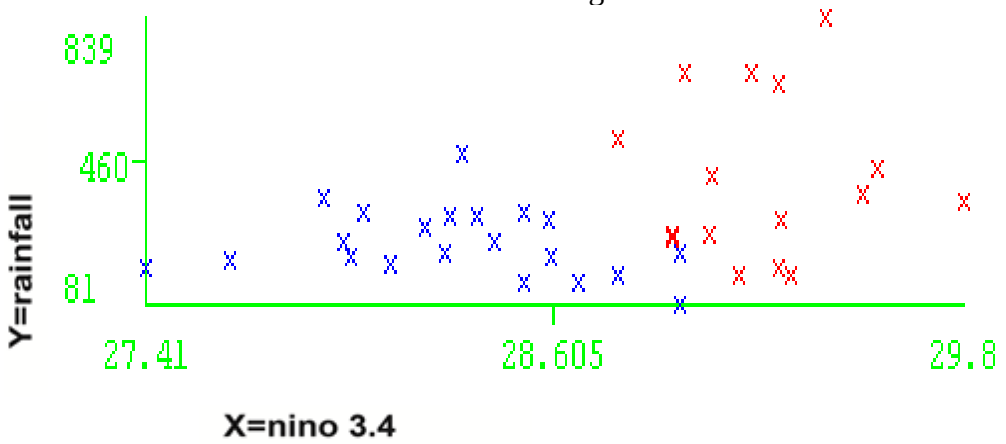


Figure 38

Khulna September:

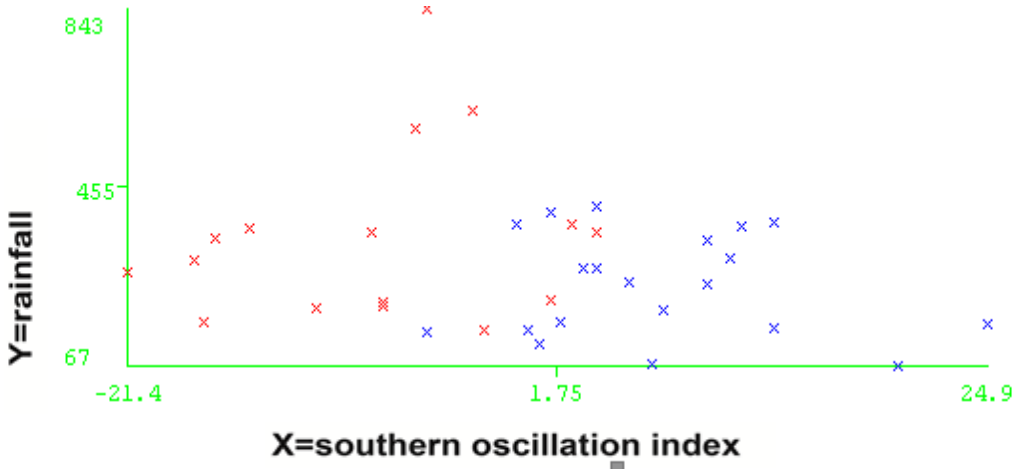


Figure 39

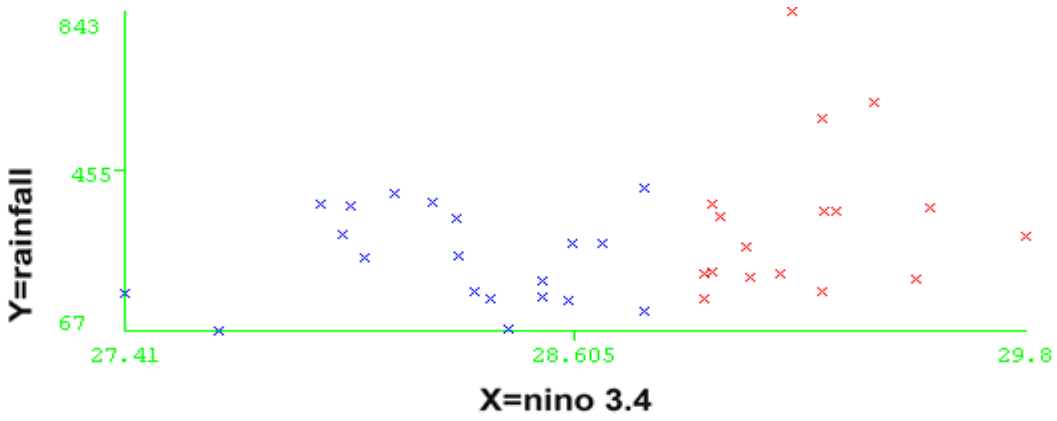


Figure 40

Bogra September:

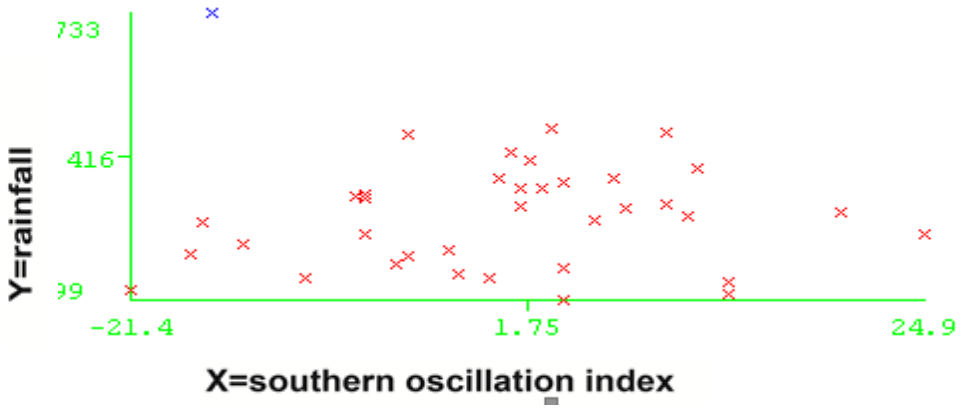


Figure 41

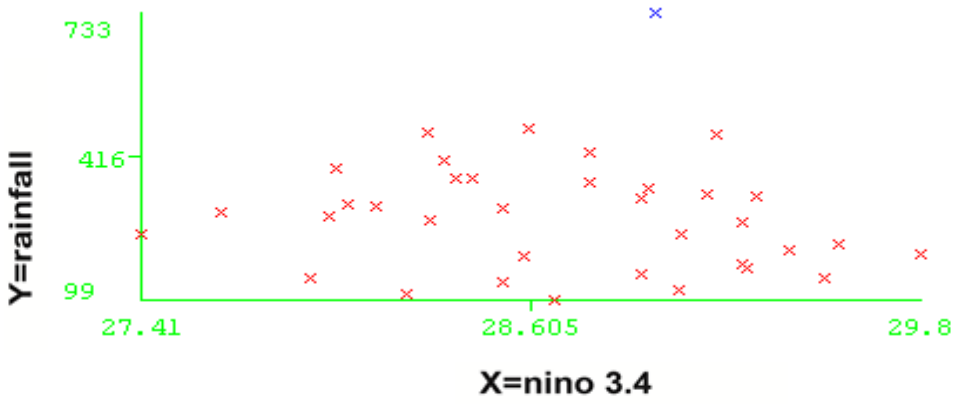


Figure 42

Barisal September:

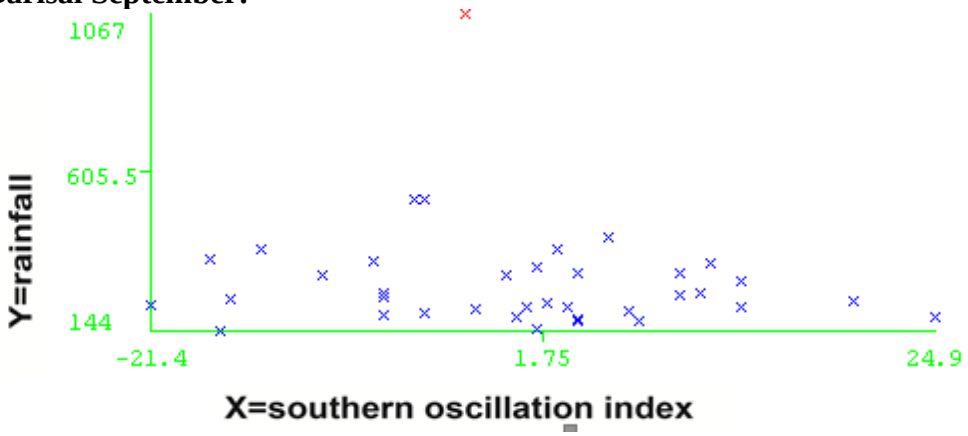


Figure 43

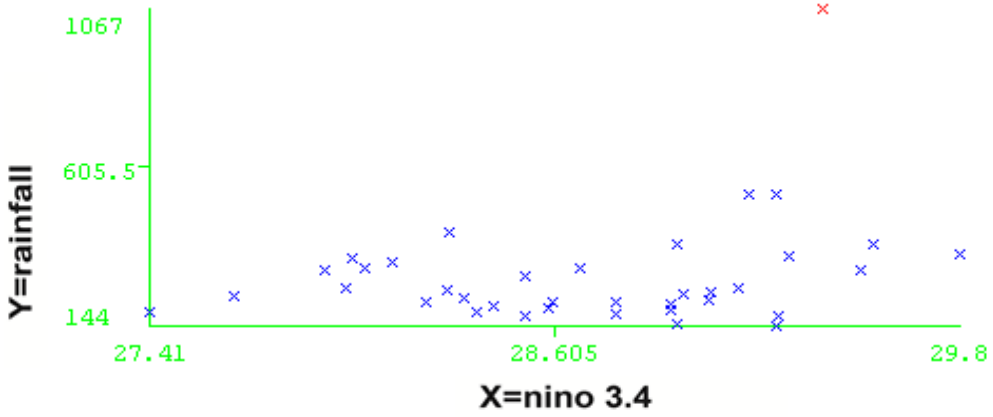


Figure 44

Cox's bazar September:

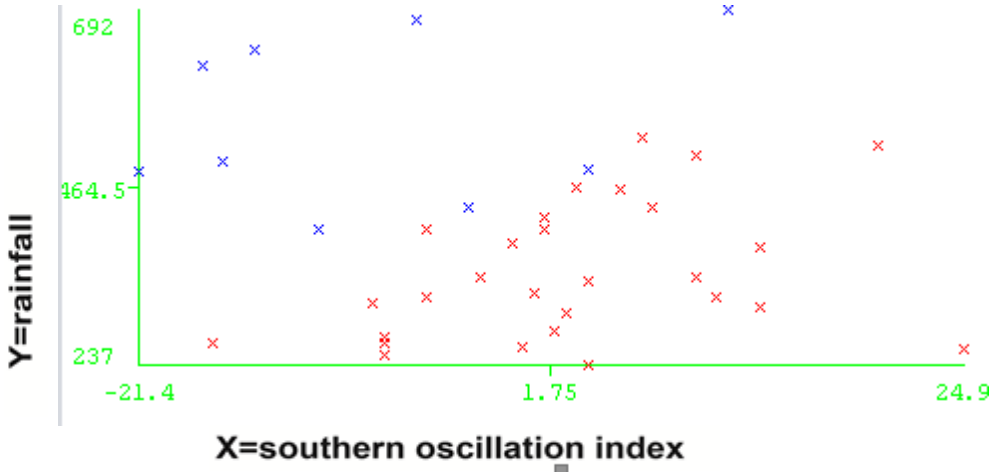


Figure 45

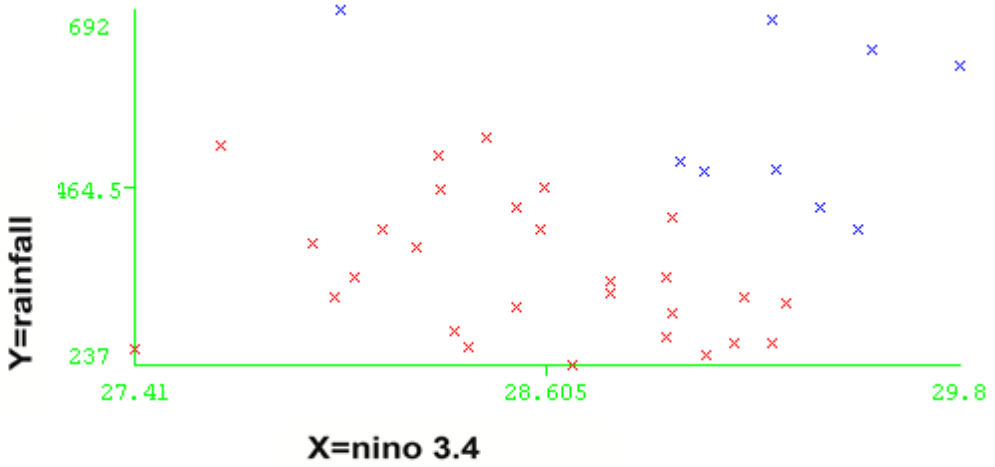


Figure 46

Dhaka October:

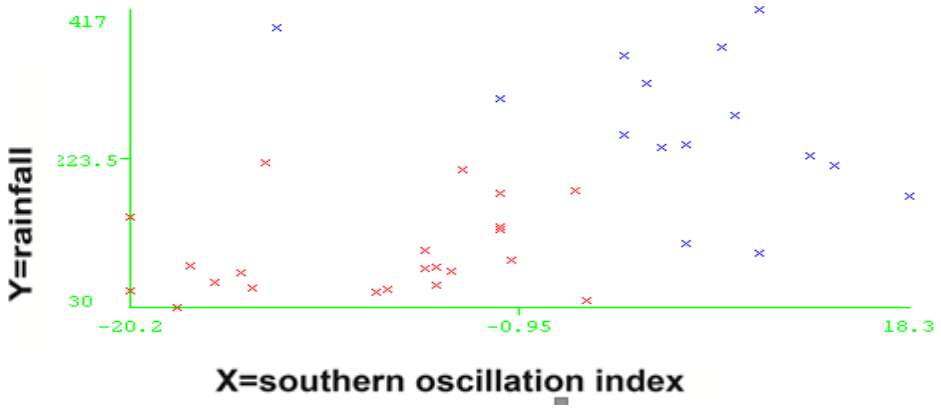


Figure 47

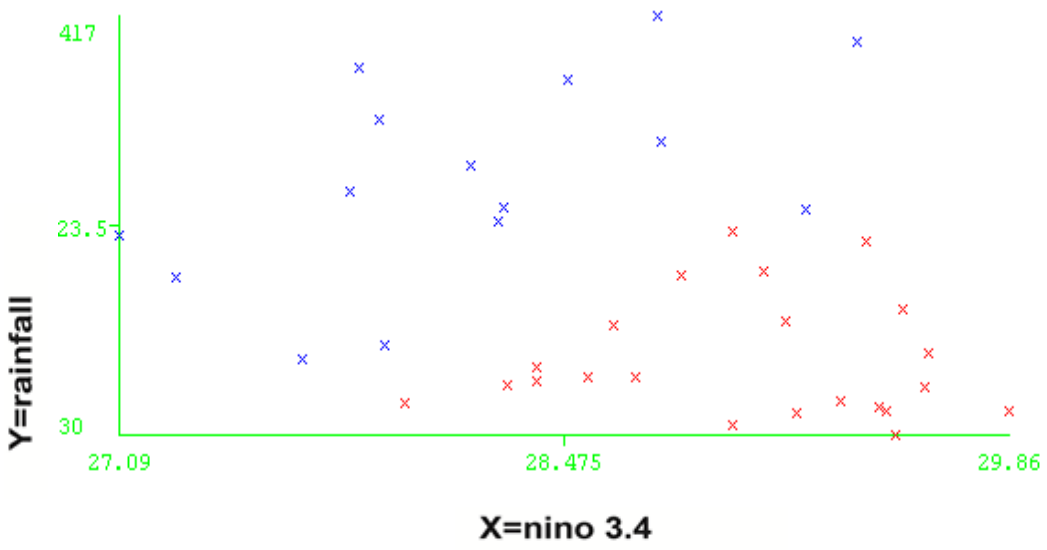


Figure 48

Khulna October:

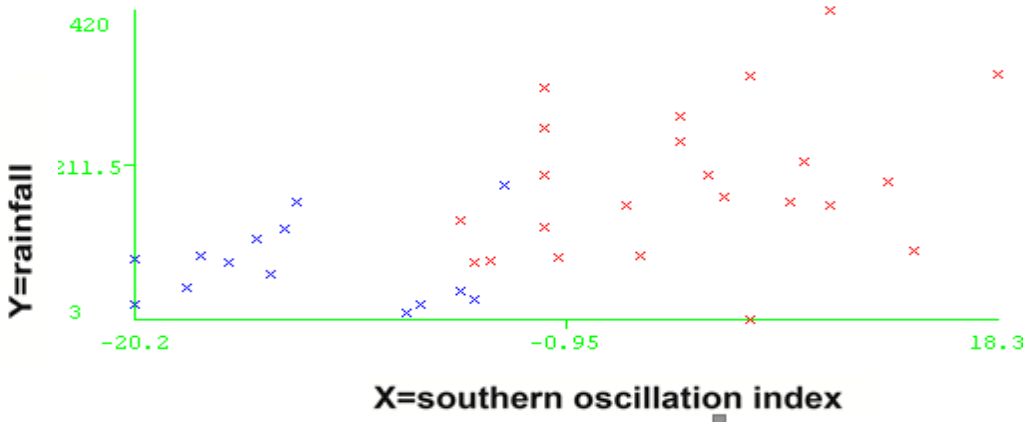


Figure 49

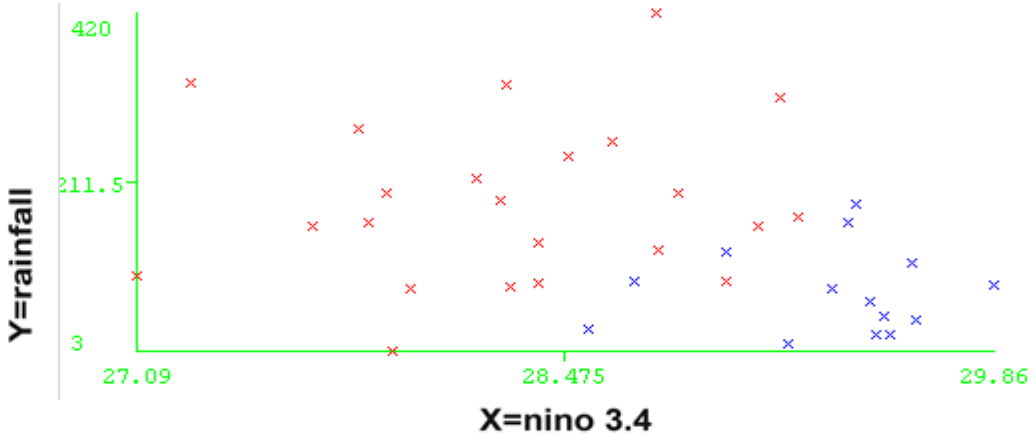


Figure 50

Bogra October:

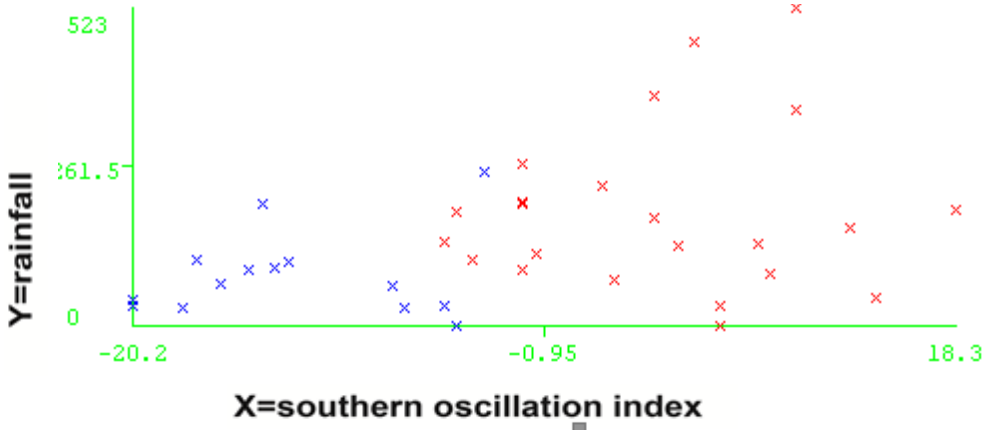


Figure 51

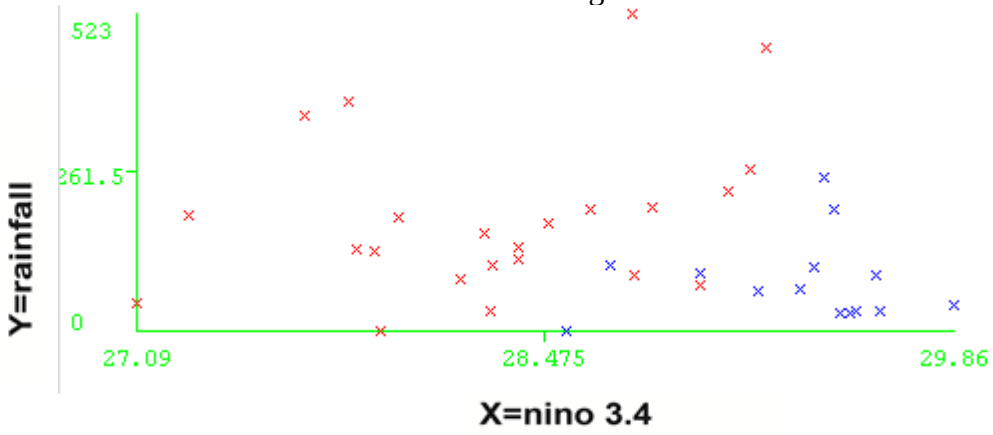


Figure 52

Barisal October:

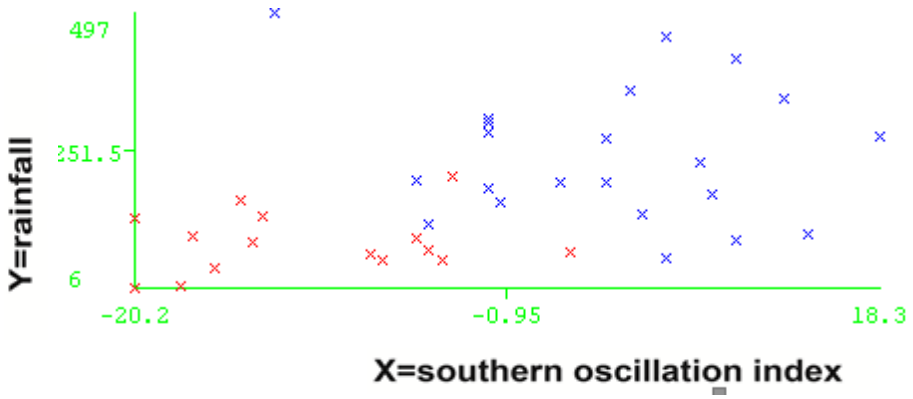


Figure 53

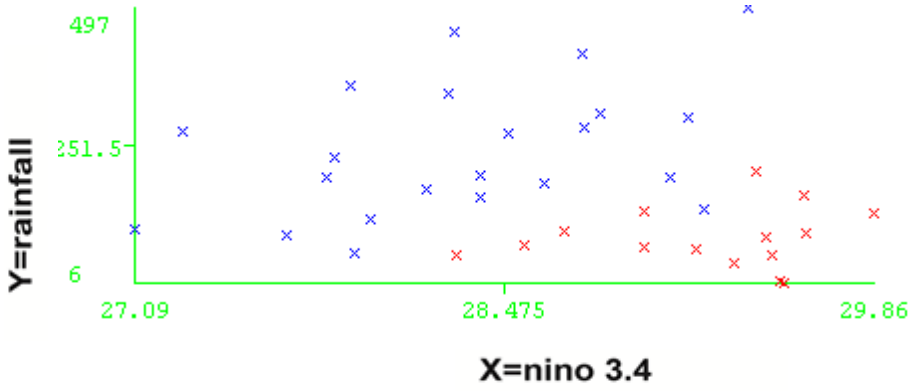


Figure 54

Bhola October:

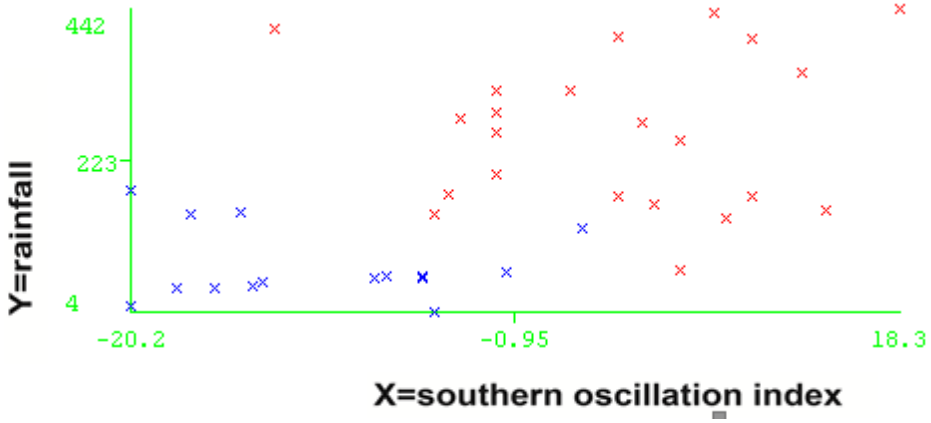


Figure 55

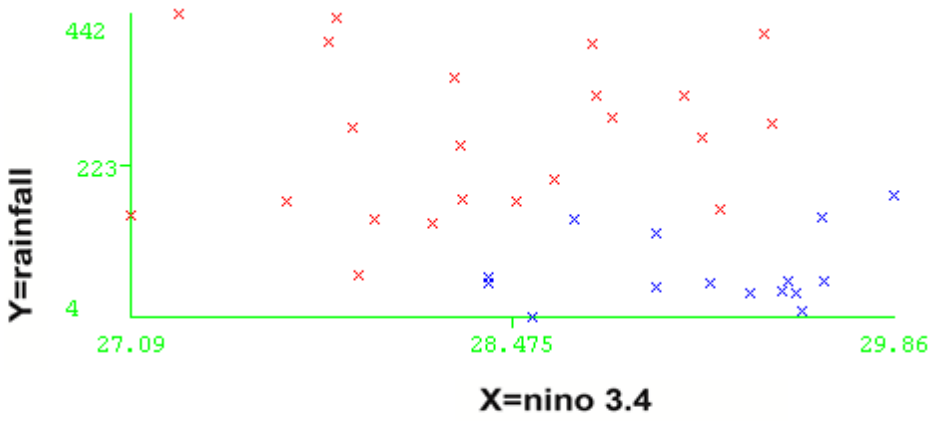


Figure 56

Cox's Bazar October:

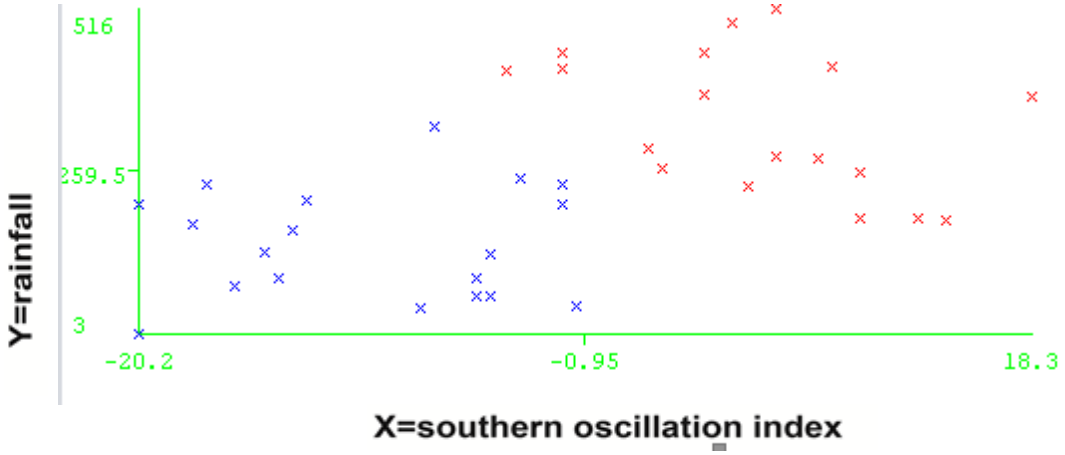


Figure 57

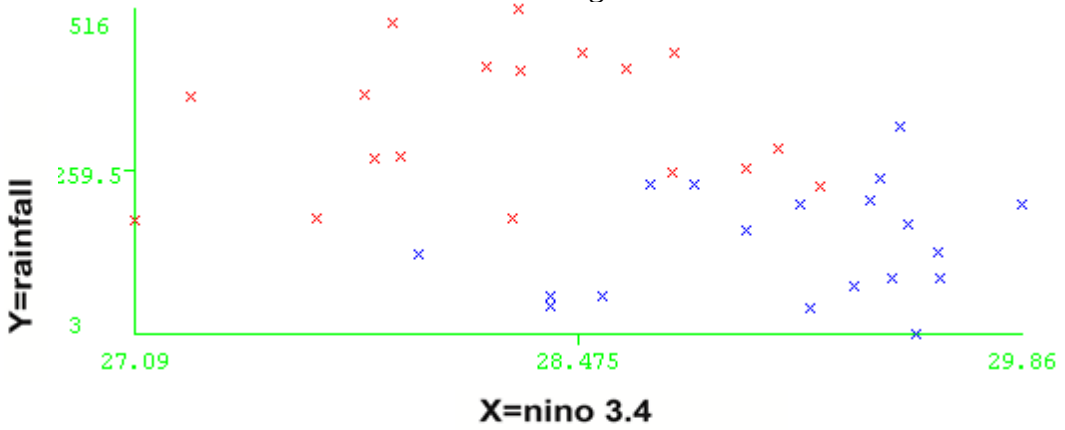


Figure 58

District	Cluster 0	Cluster 1	Within cluster sum of squared error
Dhaka June	21(57%)	16(43%)	4.657
Khulna June	17(46%)	19(51%)	3.596
Bogra June	30(81%)	7(19%)	5.358
Barisal June	20(54%)	17(46%)	3.834
Bhola June	20(54%)	17(46%)	4.434
Cox's bazar June	15(41%)	22(59%)	5.378
Dhaka July	20(54%)	17(46%)	4.485
Khulna July	21(57%)	16(43%)	3.538
Bogra July	23(62%)	14(38%)	3.681
Barisal July	22(59%)	15(41%)	3.919
Bhola July	19(51%)	18(49%)	3.841
Cox's bazar July	22(59%)	15(41%)	2.996
Dhaka August	19(51%)	18(49%)	
Khulna August	18(49%)	19(51%)	
Bogra August	19(51%)	18(49%)	3.921
Barisal August	20(54%)	17(46%)	3.713
Bhola August	20(54%)	17(46%)	2.481
Cox's bazar August	15(41%)	22(59%)	3.633
Dhaka September	21(57%)	16(43%)	
Khulna September	20(54%)	17(46%)	2.841
Bogra September	1(3%)	36(97%)	4.760
Barisal September	36(97%)	1(3%)	4.192
Bhola September	22(59%)	15(41%)	2.60
Cox's bazar September	9(24%)	28(76%)	4.432
Dhaka October	15(41%)	22(59%)	4.039
Khulna October	14(38%)	23(62%)	3.654
Bogra October	22(59%)	15(41%)	4.107
Barisal October	14(38%)	23(62%)	3.833
Bhola October	15(41%)	22(59%)	4.472
Cox's bazar October	19(51%)	18(49%)	3.687

Regression Model:

District name and month	Regression equation:
Dhaka June	Rainfall=-4.5487*soi-7-0.7357*nino+2358.2466
Khulna June	Rainfall=3.321*soi+2.6107*nino+123.3996
Bogra June	Rainfall= -4.1185*soi-103.2292 *nino3.4 +3280.1286
Barisal June	Rainfall= -1.5697*soi -10.7427*nino3.4 +696.8635
Bhola June	Rainfall=-1.8505 *soi-46.7179 *nino3.4 +1790.9236
Cox's bazar June	Rainfall=-5.8381 *soi-141.3786 *nino3.4 +4911.0679
Dhaka July	Rainfall=3.5175*soi+97.5969*nino-2426.1785
Khulna July	Rainfall=-2.4698*soi-19.478*nino+889.8517
Bogra July	Rainfall= 3.1997*soi +21.5401*nino3.4 -269.5193
Barisal July	Rainfall= 2.3871*soi +68.0195*nino3.4 -1531.5143
Bhola July	Rainfall= -1.37*soi +19.7782*nino3.4 -144.787
Cox's bazar July	Rainfall= -8.2272*soi +31.8686*nino3.4 +22.5709
Dhaka August	Rainfall=1.1063*soi+14.95967*nino-122.8869
Khulna August	Rainfall=1.8781*soi+51.3832*nino-1139.462
Bogra August	Rainfall=0.8008 *soi +12.7681*nino3.4 -78.4846
Barisal August	Rainfall=-1.9426 *soi -16.3274*nino3.4

	+816.4407
Bhola August	Rainfall=-4.8127 *soi+2.4962 *nino3.4 +295.3855
Cox’s bazar August	Rainfall= -5.1124*soi -151.6419*nino3.4 +5028.4244
Dhaka September	Rainfall=-2.3086*soi+70.688*nino-1716.5867
Khulna September	Rainfall=2.7614*soi+124.6217*nino-3294.9295
Bogra September	Rainfall= -3.122*soi -69.5424*nino3.4 +2285.9868
Barisal September	Rainfall= 4.9259*soi +158.3372*nino3.4 -4259.0519
Bhola September	Rainfall= 1.647*soi +86.6134*nino3.4 -2187.8701
Cox’s bazar September	Rainfall= 0.1369*soi +31.2987*nino3.4 -496.184
Dhaka October	Rainfall=6.8388*soi+32.7084*nino3.4-758.3915
Khulna October	Rainfall=7.3131 *soi+37.0893*nino3.4-904.718
Bogra October	Rainfall= 7.4559*soi +51.9861*nino3.4 -1299.0449
Barisal October	Rainfall= 8.2601*soi+70.3078 *nino 3.4-1855.2078
Bhola October	Rainfall=8.0967 *soi +28.9663*nino3.4 -630.0227
Cox’s bazar October	Rainfall= 6.3717*soi -0.1629*nino3.4 +255.5053

We see in every clustering graph there is some point that are very far away from its own cluster (cluster 0 and cluster 1). In this case the prediction model does not work properly as it is not consistent. For Khulna we see most consistent result. For Dhaka, Cox's Bazar, Barisal, Bogra in every month we could see some point that are scattered from its cluster. we did k means clustering to show exactly in which point our prediction model show inconsistent result.

CHAPTER 5
CONCLUSION AND FUTURE WORK

5.1 Conclusion:

During the simulation of the algorithms, in the graph some points were not forming cluster rather these points were away from the clusters and the straight-line distance (Euclidean Distance) was greater. I faced this in some month's data. As a result, the result was giving quite different value than the actual rainfall amount in that month. Otherwise the simulation gave us a closer value to the actual data. I hope this thesis will help people, especially the farmers who depend on the weather to grow their crops. They will come to know about the possibility of rainfall long before it will take place and a huge amount of disasters due to the heavy rainfall and flood will be shorten in future.

5.2 Future work:

The following chapter will draw a conclusion about our thesis project. The ending will start with future work – things that we could have done provided we had enough time, knowledge and resources. To make our project state-of-the art, we definitely need to add more into the project.

1. Currently we have focused on 34 stations across Bangladesh. We would like to increase the range of our capabilities. Such initiation will enable farmers to do efficient farming across the country. Increasing the range will also give rise to the number of data in the system.

2. For that purpose, we could use MOWCATL algorithm. We have used this algorithm on flood predicting to some extent. But to fully integrate this in our project, we would need more research. We were interested in patterns that represent unorthodoxies from the normal seasonal differences. Time series data in continuous domains is fundamentally inaccurate as we cannot avoid inaccuracy of such measuring devices, clocking strategies and natural occurrences [2]. From [1] we found a relevant approach towards their data sets and their simulation of data.

REFERENCES

1. Harms, S. K., & Deogun, J. S. (2004). Sequential Association Rule Mining with Time Lags (M. S. Hacid, Ed.). Journal of Intelligent Information Systems, 22:1(7–22). Retrieved September 1, 2008.

2. Goldin, D.Q. and Kanellakis, P.C. (1995). On Similarity Queries for Time-Series Data: Constraint Specification and Implementation. In Proceedings of the 1995 International Conference on the Principles and Practice of Constraint Programming (pp. 137–153). Marseilles, France.

3. Thiele-Eich, I., Burkart, K., & Simmer, C. (2015). Trends in Water Level and Flooding in Dhaka, Bangladesh and Their Impact on Mortality (K. L. Ebi & J. Hess, Eds.). International Journal of Environmental Research and Public Health, 1196-1215. doi:10.3390/ijerph120201196

4. Ert¨oz L., Steinbach M., and Kumar V., "Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data", In Proc. of 3rd SIAM International Conference on Data Mining, San Francisco, CA, USA, 2003

5. Liu Y., "A framework of data mining application process for credit scoring", InstitutfürWirtschaftsinformatik, Georg-August-University G.ttingen, http://www.wi2.wiso.unigoettingen.de/getfile?DateiID=394 ,2002

6. Box, G.E.P., Jenkins, G.M.," Time Series Analysis: Forecasting and Control", Holden Day, San Francisco, CA, 1976.

7. El Niño/Southern Oscillation (ENSO) Technical Discussion. Retrieved December 22, 2017, from https://www.ncdc.noaa.gov/teleconnections/enso/enso-tech.php

8. Equatorial Pacific Sea Surface Temperatures. Retrieved December 22, 2017, from https://www.ncdc.noaa.gov/teleconnections/enso/indicators/sst.php

9. M.Kannan, S.Prabhakaran, P.Ramachandran "Rainfall Forecasting Using Data Mining Technique" International Journal of Engineering and Technology Vol.2 (6), 2010, pp. 397-401.

10. Khaled Hammouda ,Prof. Fakhreddine Karray ''A Comparative Study of Data ClusteringTechniques'' SYDE 625: Tools of Intelligent Systems Design. Course Project.

11. S. Kannan , Subimal Ghosh, "Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output, Stochastic Environmental Research and Risk Assessment. May 2011, Volume 25, Issue 4, pp 457-474.

M.Sc

ORIGINALITY REPORT

21%	14%	7%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	paper.ijcsns.org Internet Source	8%
2	research.ijcaonline.org Internet Source	5%
3	Submitted to Daffodil International University Student Paper	3%
4	Bruce Clapham, Thomas S. Reger, Kim D. Janda. "Polymer-supported catalysis in synthetic organic chemistry", Tetrahedron, 2001 Publication	3%
5	Sherri K. Harms. "chapter 206 Temporal Association Rule Mining in Event Sequences", IGI Global, 2005 Publication	1%
6	Submitted to Howard Community College Student Paper	1%
7	Submitted to Swinburne University of Technology Student Paper	1%

8	Submitted to Lindenwood University Student Paper	<1%
9	link.springer.com Internet Source	<1%
10	eprints.ums.edu.my Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		