Inspiring Excellence

# PREDICTION OF RAINFALL USING DATA MINING TECHNIQUES

## Supervisor: Suraiya Tairin

## Submitted by:


**FAHMIDA TASNIM PREMA**

**12201069**

**SHARMIN AHAMED**

**17141001**

**MD RIFAT ISLAM**

**13110015**

**ADVISOR**

**-------------------**

**SURAIYA TAIRIN PAKHI**

**CO-ADVISOR**

**--------------------**

**MD. SAIFUL ISLAM**

**---------------------**

**FAHMIDA TASNIM PREMA**

**---------------------**

**MD RIFAT ISLAM**

**---------------------**

**SHARMIN AHAMED**

# INDEX

# Acknowledgement

First and foremost, we would like to thank our thesis supervisor Suraiya Tairin Pakhi and our co-supervisor Md. Saiful Islam, lecturer of the School of Computer Science and Engineering at BRAC University. We came up with the idea behind this thesis project with their help. Without their assistance and dedicated involvement throughout every step, this thesis would have been incomplete. They encouraged us to contribute to our country's weather system using data mining techniques with rainfall, El Nino and SOI (Southern Oscillation Index) data. We would like to thank everyone for their support and understanding over the past year.

Three of us- Sharmin Ahmed, Fahmida Tasnim Prema and Md. Rifat Islam have conducted a research on this topic and decided to carry on with the project. The members were fully dedicated with their goal that is, completing the thesis flawlessly.

We would once again thank our thesis supervisors for sharing their knowledge, experience, and valuable insights regarding our thesis. Without their valuable time and patience we would not have reached towards the completion of this project. We would also like to thank our family for their overwhelming support.

We would also extend our gratitude towards Bangladesh Meteorological Department (BMD) for providing the rainfall data necessary for conducting simulation.

Finally, we would like to express our gratitude sincerely to the honorable faculty members of Computer Science and Engineering department for enhancing our minds with their technical know-how. We are very thankful to BRAC University, Bangladesh for giving us a chance to complete our B.Sc. degree in Computer Science and Computer Science and Engineering.

# ABSTRACT

The research for this paper concentrates on finding inter-relations between various climatic indices and predict precipitation consequently. And since rainfall is the prominent reason behind flood, our study can aid immensely in predicting flood and designing a proper risk management system. Flood has been a major hindrance in the path of development for Bangladesh. Being a riverine country, flood occurs in Bangladesh almost every other year. Predicting flood accurately can help us in developing our economy. Our study shows how the climatic parameters (SOI,El Nino) are responsible for major rainfall in Bangladesh. Though many other researches on predicting rainfall have been conducted using other climatic factors, the southern oscillation index and the El nino 3.4 show stronger correlation with rainfall in our country than the others. For establishing a relationship among rainfall ,SOI and El Nino , we have applied Data Mining technique. The specific data mining algorithms that we have implemented in our paper are K-clustering, Decision tree and Regression model. The outputs of these algorithms give us a straightforward relationship between rainfall and the input parameters. Implementing our method on the dataset of rainfall for the past couple of years, our estimated rainfall is almost the same as the actual ones of those years. So in designing a feasible rainfall prediction model for Bangladesh, our work can play a significant role due to its high efficiency.

# Thesis Contribution:

The main purpose behind this thesis was to predict rainfall using data mining techniques. The objective of this thesis was to make the farmers aware about the rainfall possibility long before it takes place. As a result they will be able to face the consequences and the damages will be lessen. Mainly this thesis will contribute to the agricultural sector of Bangladesh. Each year several areas in Bangladesh are submerged underwater due to heavy rainfall. As a consequence, farm land are severely damaged and many cattle farmers suffer from this. Several housing are also deteriorated 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. Moreover, 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. Such algorithms will enable the farmers know when any tropical catastrophe will take place. Not only that, the farmers can prepare a month or a week early before they can see their hard work being drowned or blown away by cyclones.

We went to Bangladesh Meteorological Department (BMD) and talked to several experts regarding 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.

# Introduction

Bangladesh is a riverine country. It falls under the active monsoon region of the world. Among the total area of 1,47,570 sq. kms of this deltaic land, the river area alone holds 9399 sq. kms (DOE 2002). Having an agro based economy, almost 40% of the total population of this country are related to agriculture professionally. Hence, rainfall plays a crucial part in the economy of Bangladesh. The erratic heavy rainfall events may affect ecosystems, agriculture, food security, urban. drainage, water availability, water quality and health and livelihood of people of the country. On the monsoon of 2007, The country received unprecedented heavy rainfall together with the onset of flooding by the Himalayan rivers which led to severe flood in almost half of Bangladesh. Heavy rainfall at 14th September of 2004 and 28th July of 2009 resulted in severe urban flooding and drainage congestion in the capital city of Dhaka. Landside fatality also was observed during 2008 and 2010 due to heavy rainfall events over the country. Recent studies also suggest that the frequency and magnitude of heavy rainfall events have already been increased under the global warming scenario including the high altitude areas of Bangladesh. It is therefore, imperative to have an efficient rainfall estimation system for this natural disaster prone country.

The El Nino and SOI (Southern Oscillation Index) are classifications of similar processes related to SST( Sea Surface Temperature ) and Pressure anomalies for sea surface and surface level of the atmosphere and together it is known as ENSO ( El-Nino and Southern Oscillation ). ENSO is a dominant mode of interannual climate variability that develops from air-sea interactions in the tropical Pacific, but effects weather patterns globally. Since Bangladesh is such a country which is a little far from the area where these climatic factors are dominant, finding co relations among rainfall and the factors has always been a challenge. Though most of the parameters are weakly related to rainfall in our country, ENSO shows that it has a significant role to play behind the erratic precipitation occurences in this country.

The research conducted in our paper mainly uses the two above mention parameters as input and find out the estimated rainfall. Since these two factors occur long before the rainfall, the prediction can be made early and quite conveniently. We have applied our algorithms of Decision tree, K-means clustering and regression model to estimate the rainfall of two stations, Dhaka and Khulna. We have selected these stations based upon their correlation factor with ENSO. Running our algorithm on the years 1980-2016, our output shows that the rainfall estimated via our proposed method matches almost identically with the actual rainfall data. We have also compared the accuracy of these three algorithms regarding rainfall estimation.

# 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 major thorn in the path of our economic development. 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.

# Problem statement:

In our research, nearly thousand data instances containing values like rainfall, temperature, air pressure, year etc have been used and analyzed with the help of different algorithms for getting a proper prediction. Collecting all the data was quite difficult as we had to go to the Bangladesh Meteorological Department (BMD) for rainfall data. We went to the center several times and tried to talk to some experts and there were some times they could not make time. After that we collected other data from different reliable sites. Getting more accurate data was important for us. Finding out the site where we could get the actual data was difficult. Moreover, none of us was so expert in programming. As a result we had to go through tough times to understand the codes well and work based on that. Giving input in an accurate way and knowing what is the expected output and what we are getting was also an important task for us. We had to go through many trials and errors in this regard during working with those algorithms. Extensive amount of research was done to understand the basic of the algorithms. Furthermore, learning how to work with big data in excel was a bit new experience.

# Literature review:

According to AMS (2014), all solid or liquid aqueous particles that originate in the atmosphere and fall into the earth surface are precipitation. Rainfall, hail, snow, drizzle or sheets are the main form of precipitation. Rainfall can be considered as the amount of liquid precipitation that reach earth surface and collected by rain gauge. Rainfall is the most dominant element of the climate in Bangladesh. The country has a tropical monsoon climate with high amount of rainfall. Distinctive seasonal pattern exists in annual rainfall cycle which is more prominent than temperature cycle.

Forecasting heavy rainfall events has been a challenging job for Bangladesh due to its weak correlations with the major factors behind south Asian monsoon. There are several methods available to forecast rainfall events.
Rahman et al. (2013) used simple regression model to forecast summer monsoon rainfall over Bangladesh.

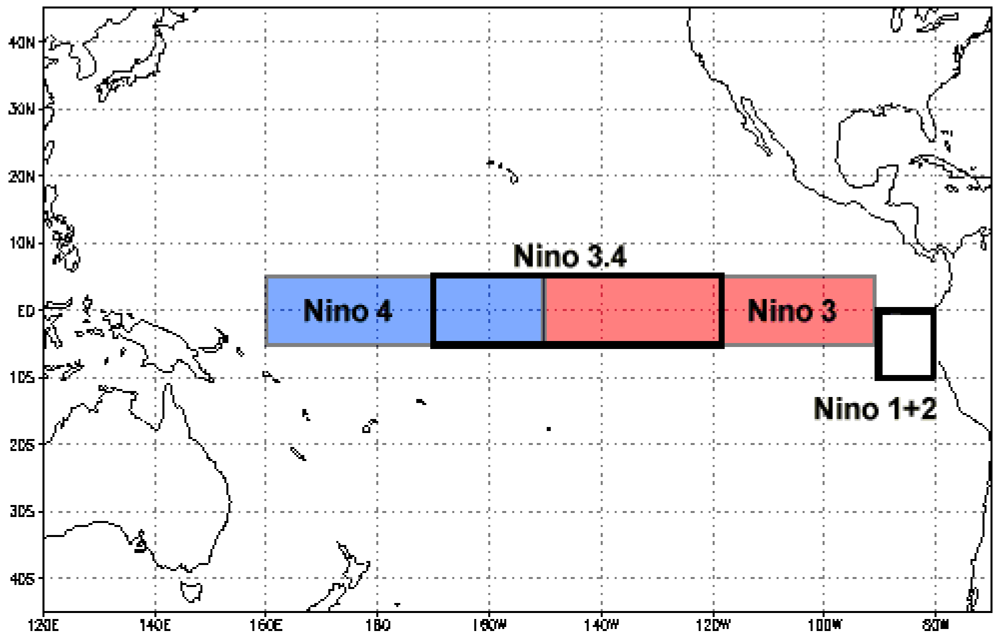
Many regional and event based study has been done so far using NWP models to improve the high intensive rainfall forecast. And in Bangladesh, these studies have been done more seldom. Till date, only a few systems have been designed for rainfall estimation. Among them, data mining has been a major tool since most of these systems have been implemented using one or another data mining model.

# Methodology:

**ENSO**

The episodic fluctuation in sea surface temperature, known as El Nino and the air pressure of the superimposing atmosphere that is, Southern Oscillation throughout the equatorial Pacific Ocean combines to form ENSO [1]. El Nino reaches its full strength at the end of every year usually. It is an occurrence in the equatorial Pacific Ocean written off as five sequential 3 month running average of sea surface temperature (SST) anomalies in the Nino 3.4 region which is below and above the threshold of +0.5 degree celcius to -0.5 degree celcius. Oceanic Nino Index (ONI) is the standard measurement [2].

Niño Regions

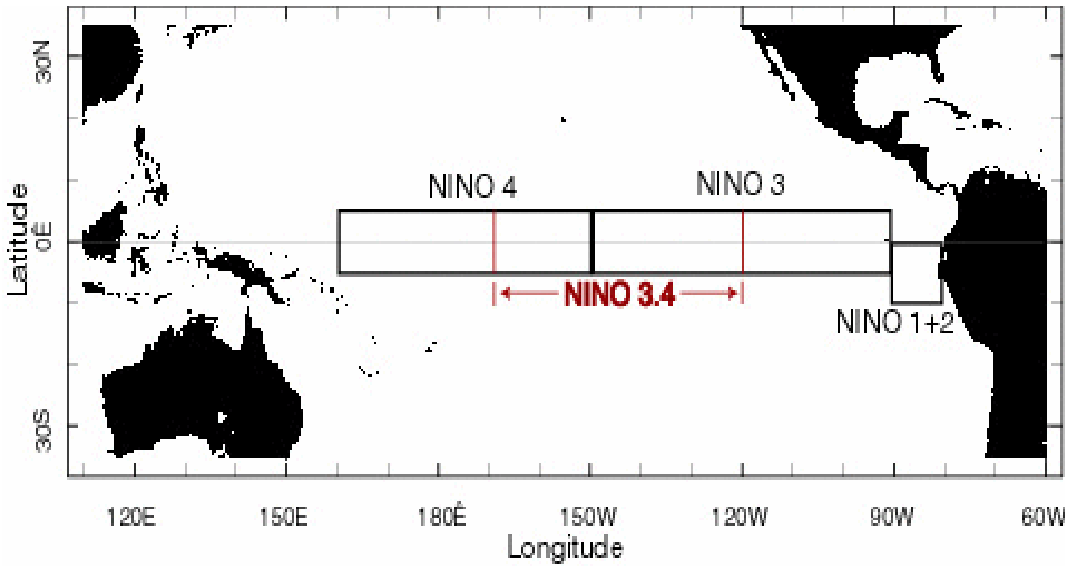



The Southern Oscillation is more of a dual peak variation in sea level barometric pressure between two stations – Darwin and Tahiti. It is enumerated in the Southern Oscillation Index (SOI) which is basically the change between lower pressure over Darwin and higher pressure over Tahiti. The pressures promote a round-up of air from east to west that brings out warm surface water towards the west makes it rain in Australia and the western Pacific. Heavy precipitation can cause floods when pressure difference lowers in parts of western Pacific. Southern Oscillation is strongly unpredicted with El Nino conditions [1].

The two components of ENSO which are sea surface temperature and atmospheric pressure are strongly related although its specific initiating causes are not fully understood. During the El Nino event, the eastern trade winds slow down across the equatorial Pacific. This in turn slows the ocean current that draws surface water away from the western coast of South America and decreases the upsurge of cold, nutrient–rich water from the deeper ocean, crushing out the thermocline and allowing warm surface water to make in the eastern part of the basin.

There is a relation between the Southern Oscillation and rainfall which is also clear in the amount of long radiation waves that leaves the atmosphere. A great deal of long radiation waves exists under the clear skies which are released into the atmosphere from the surface and escape into space. Under the cloudy skies, some radiation is denied from escaping. These activities are measured using satellites.

To monitor the ENSO condition, we first need to focus on the sea surface temperature (SST) in 4 geographic regions of the equatorial pacific. SST anomalies 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, anomalies less than or equal to -0.5 degree Celsius which are cool and are called La Nina condition. If the Oceanic Nino Index (ONI) shows warm or cool phase condition in less than or equal to five sequential values, it becomes an El Nino or La Nina event [1].




The reason why we have chosen Southern Oscillation Index instead of other climate indices 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.

# Results and Analysis:

## K-means clustering

## Dhaka June:

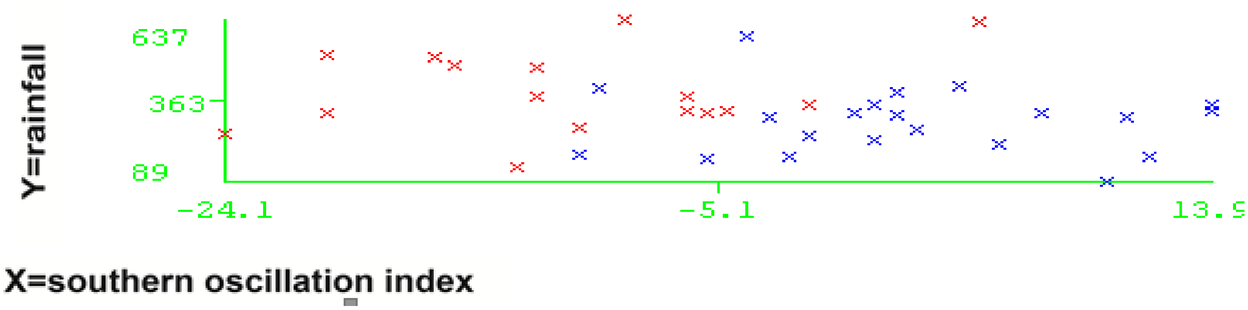



**Figure 1**

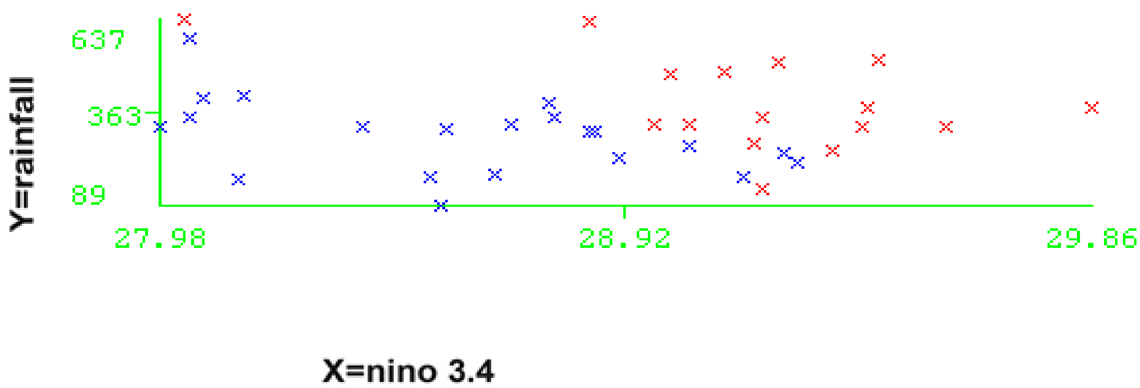



**Figure 2**

## Khulna June:

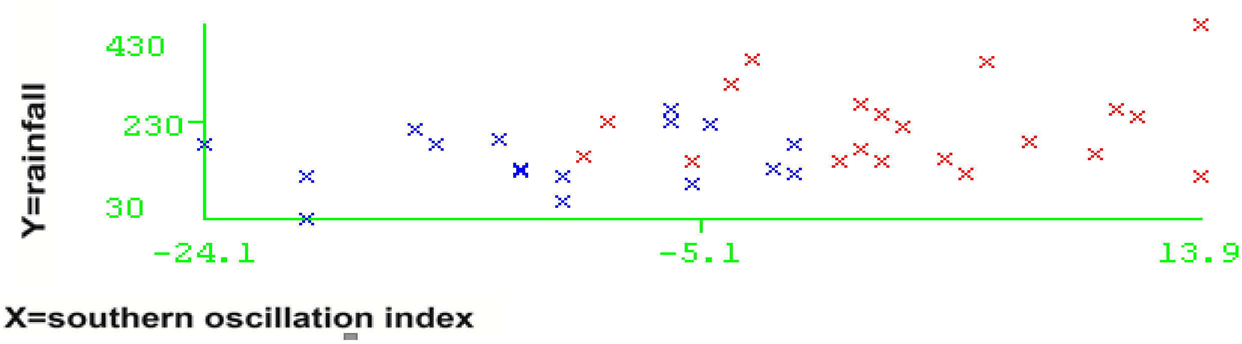
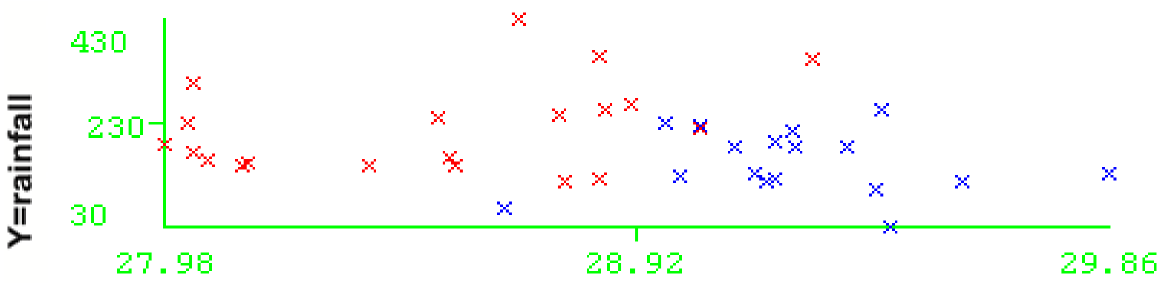


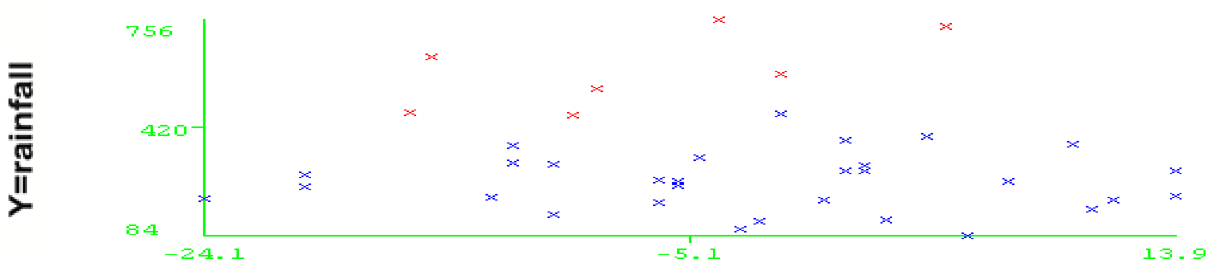

**Figure 3**




X=nino 3.4

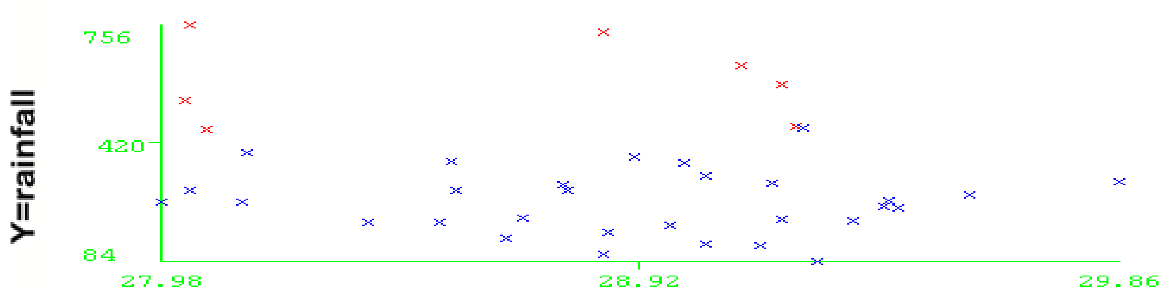
Figure 4

# Bogora June:




X=southern oscillation index

Figure 5




X=nino 3.4

Figure 6

## Barisal June

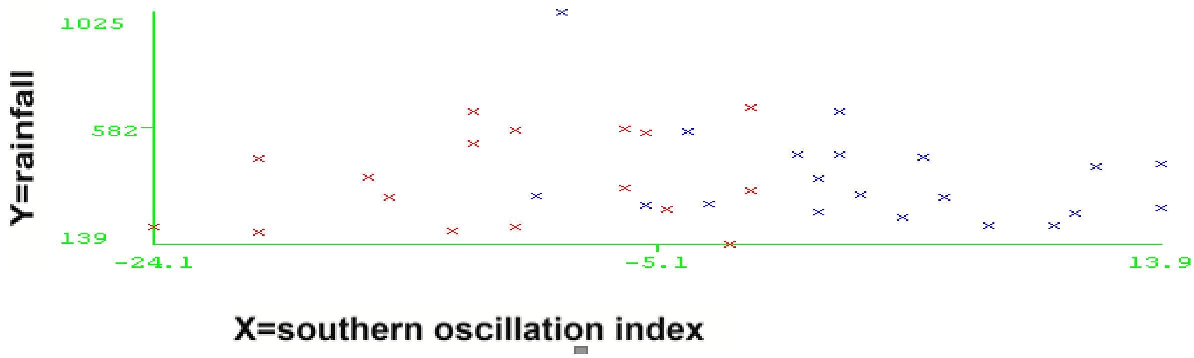



Figure 7

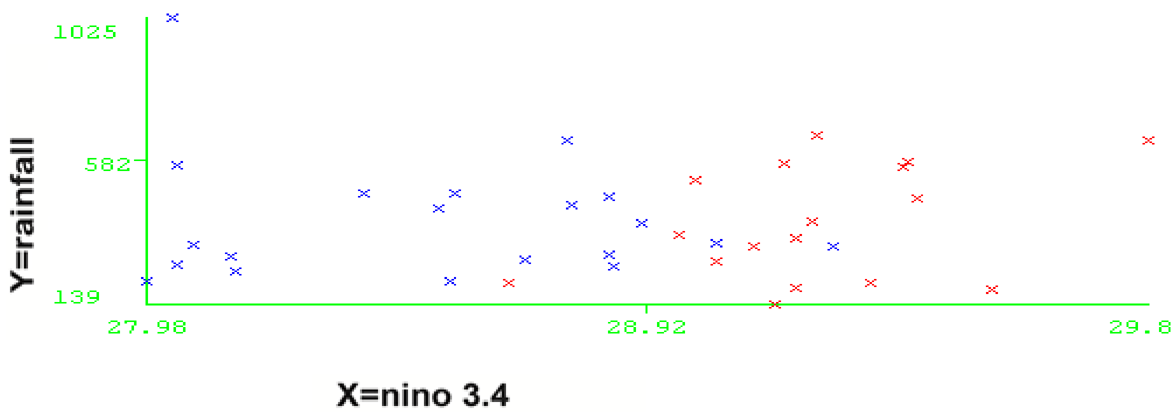



Figure 8

## Bhola June:

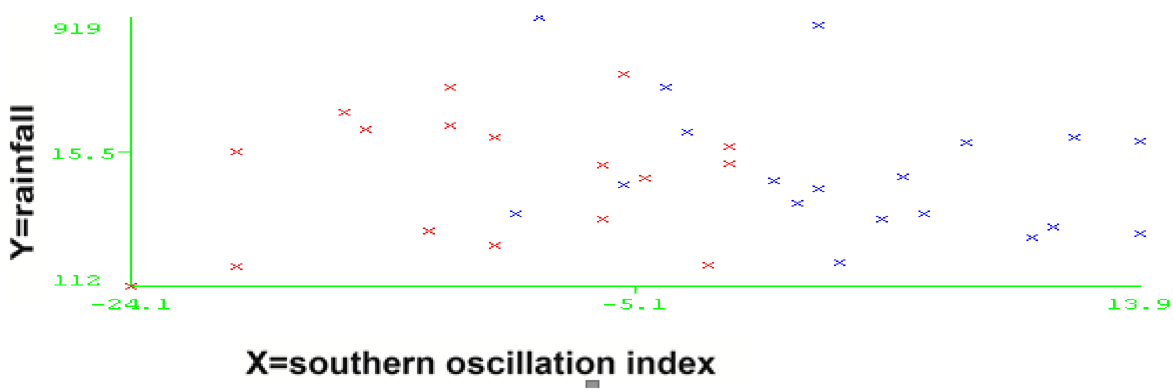



Figure 9

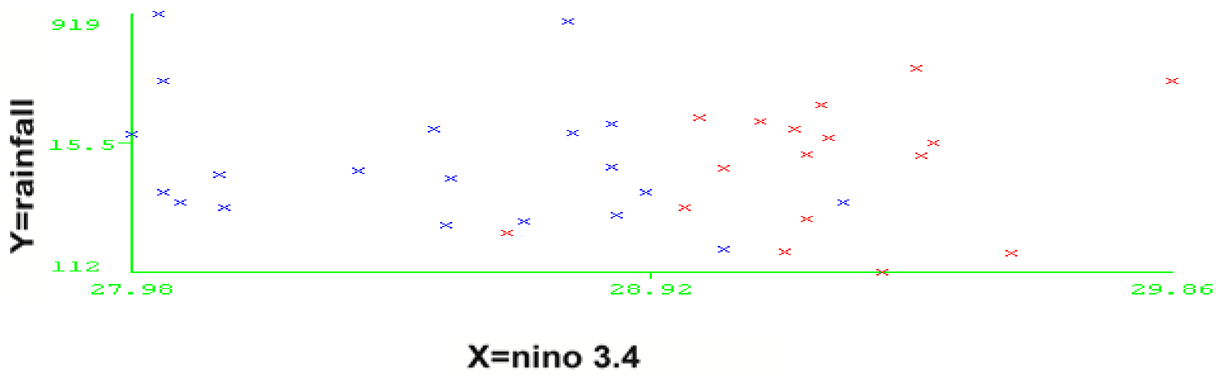



**Figure 10**

# Cox's Bazar June:

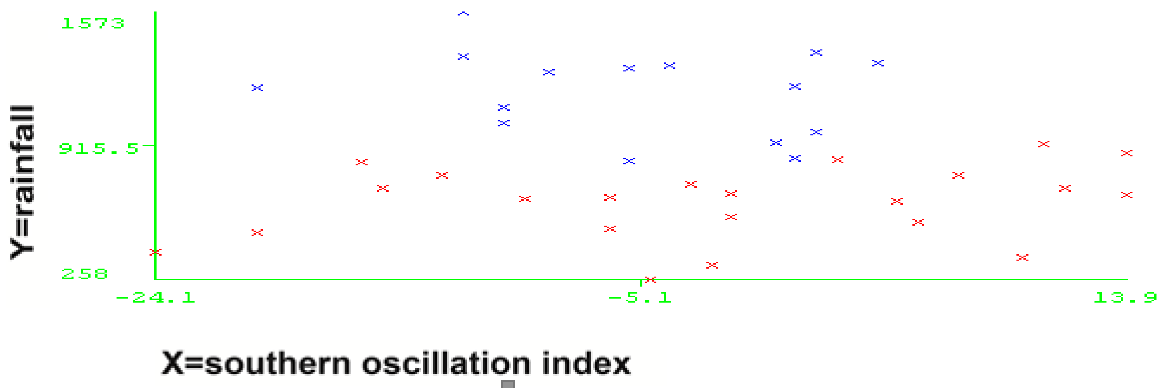



**Figure 11**

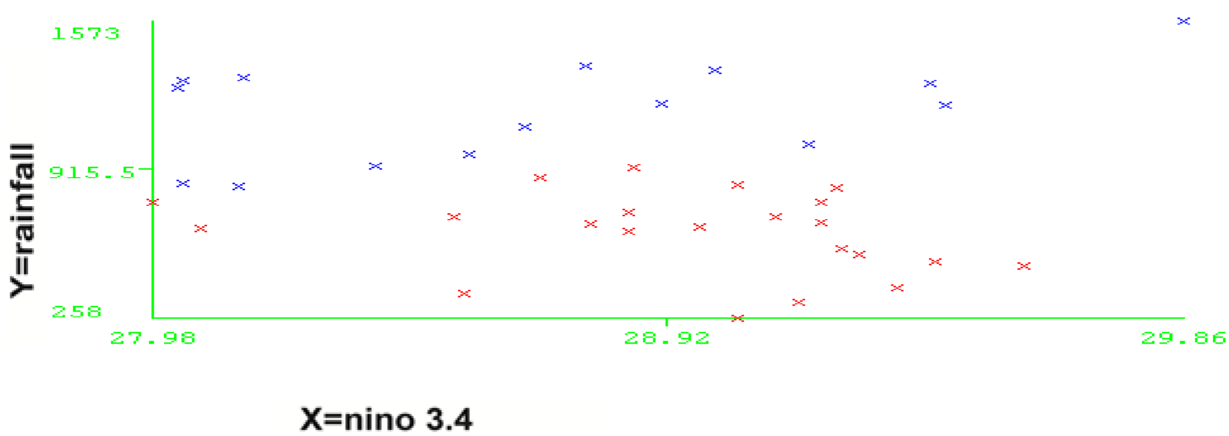



**Figure 12**

# Dhaka July

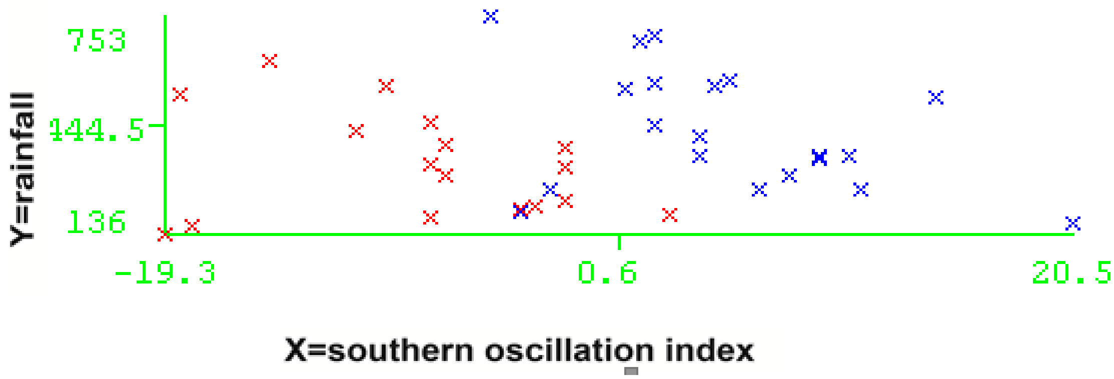



**Figure 13**

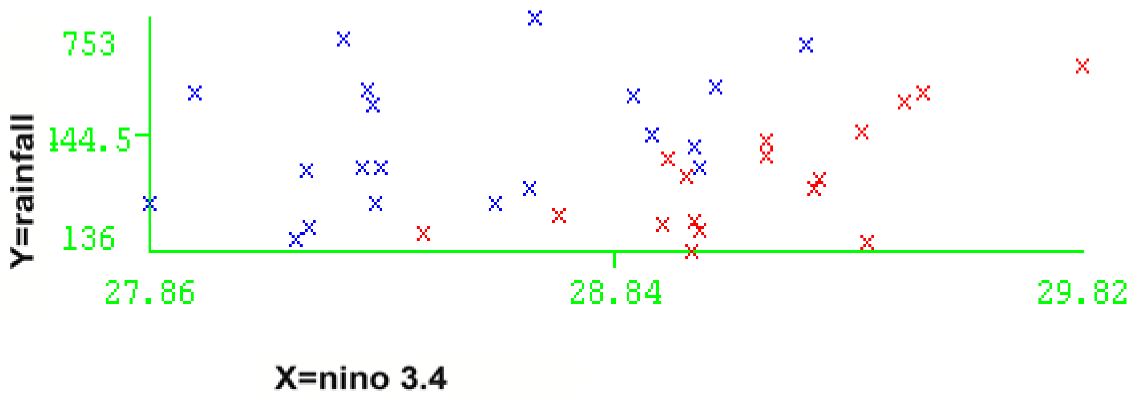



**Figure 14**

# Khulna July:

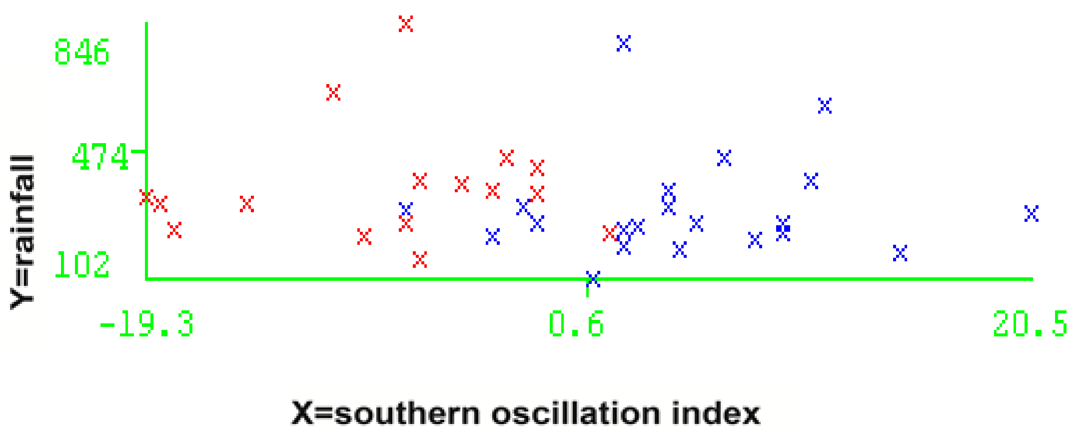



**Figure 15**

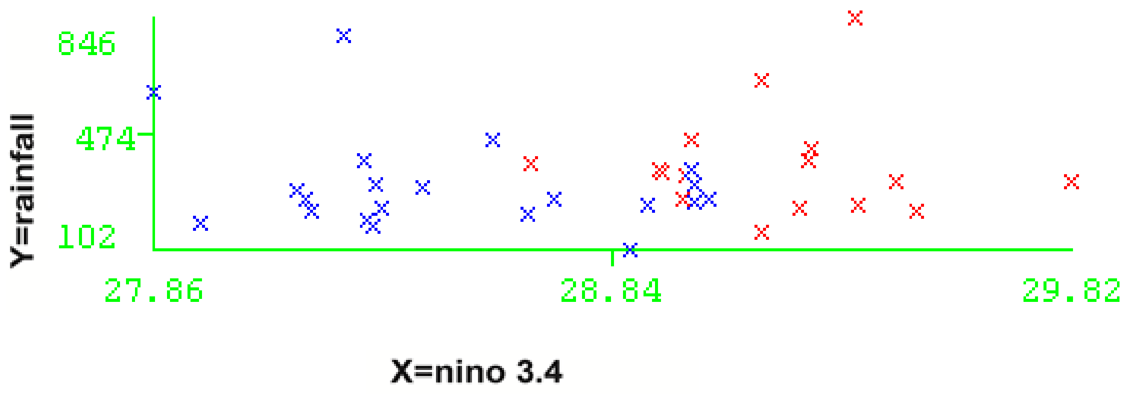



Figure 16

# Bogra July:

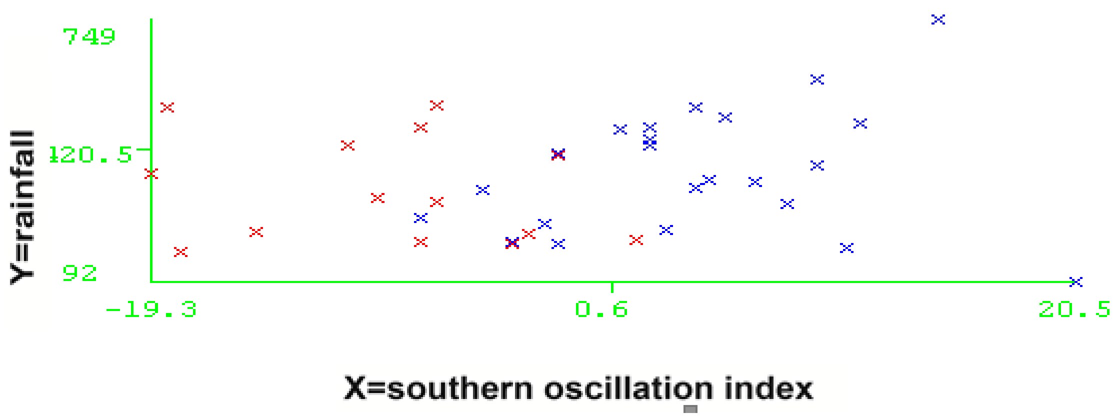



Figure 17

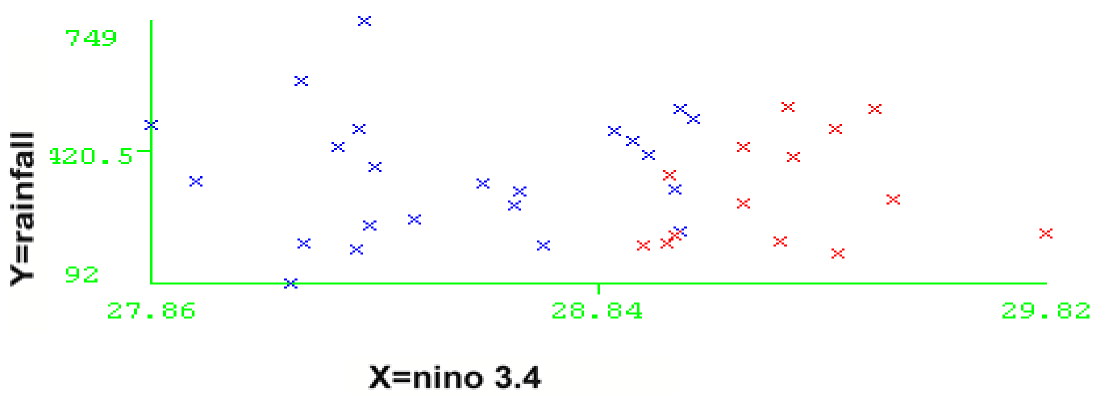



Figure 18

# Barisal July:

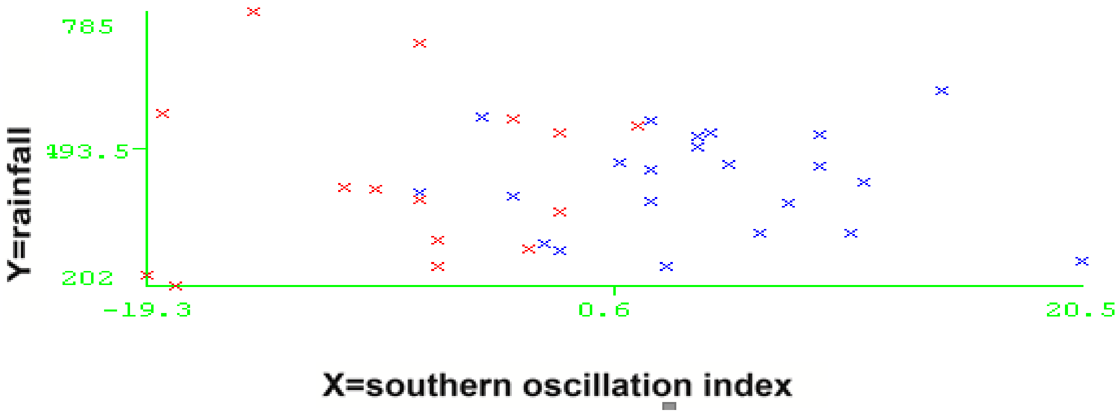



Figure 19

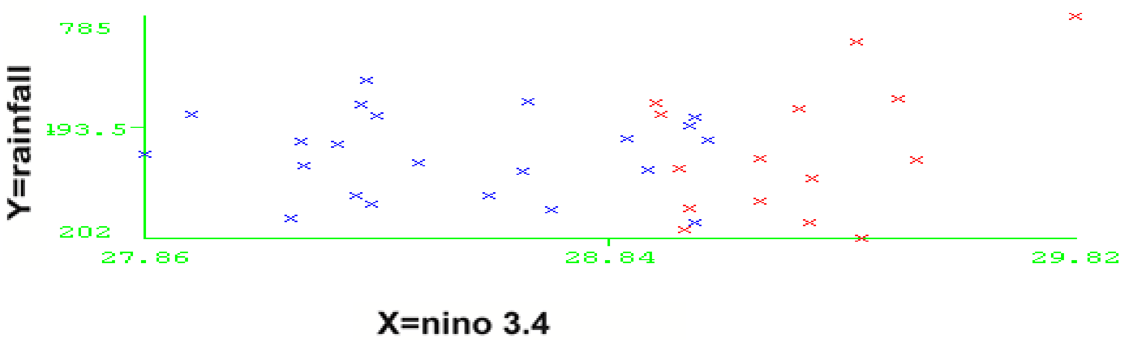



Figure 20

# Bhola July:

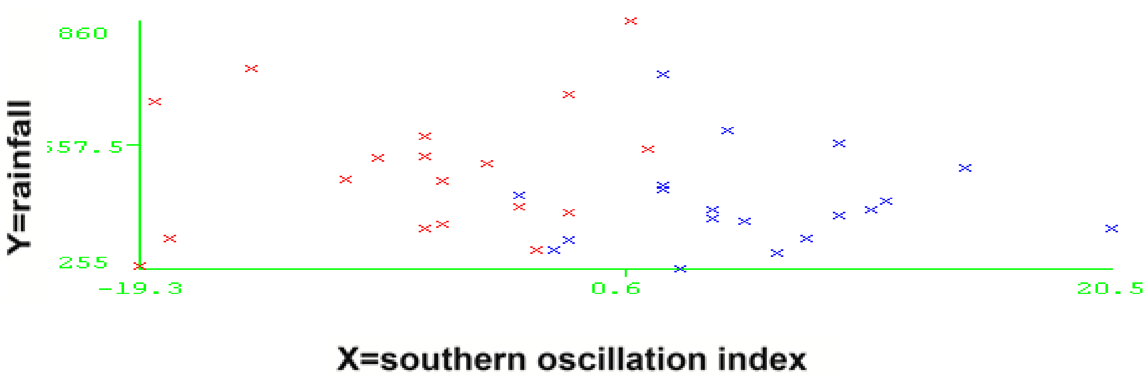



Figure 21

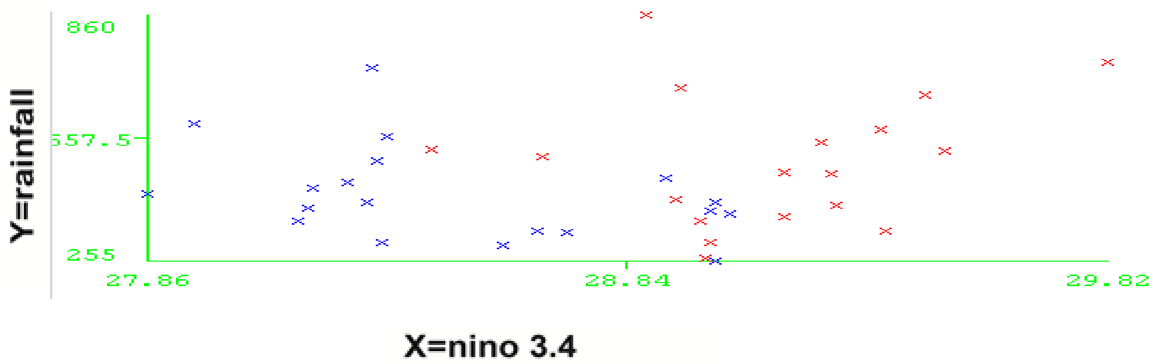



**Figure 22**

# Cox's Bazar July:

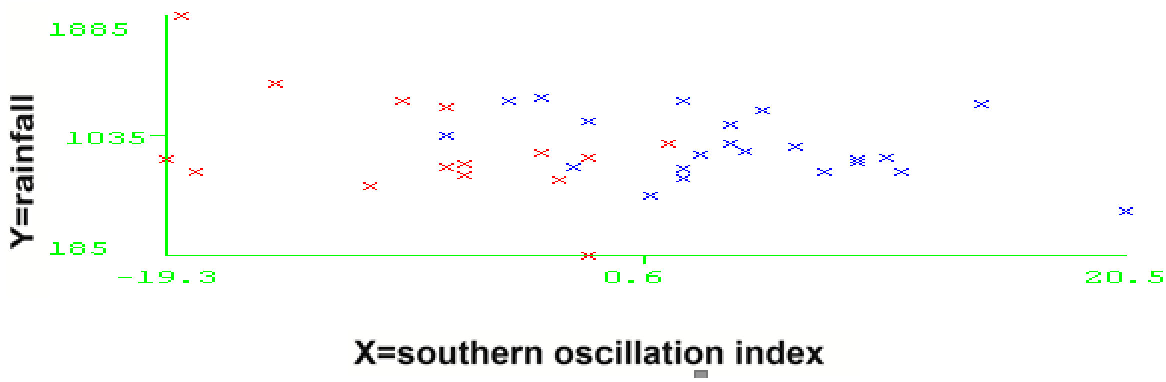



**Figure 23**

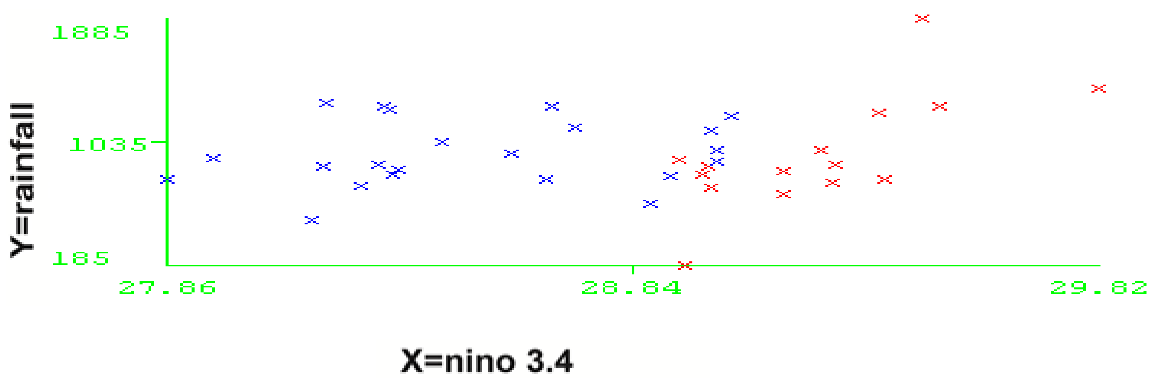



**Figure 24**

# Dhaka august

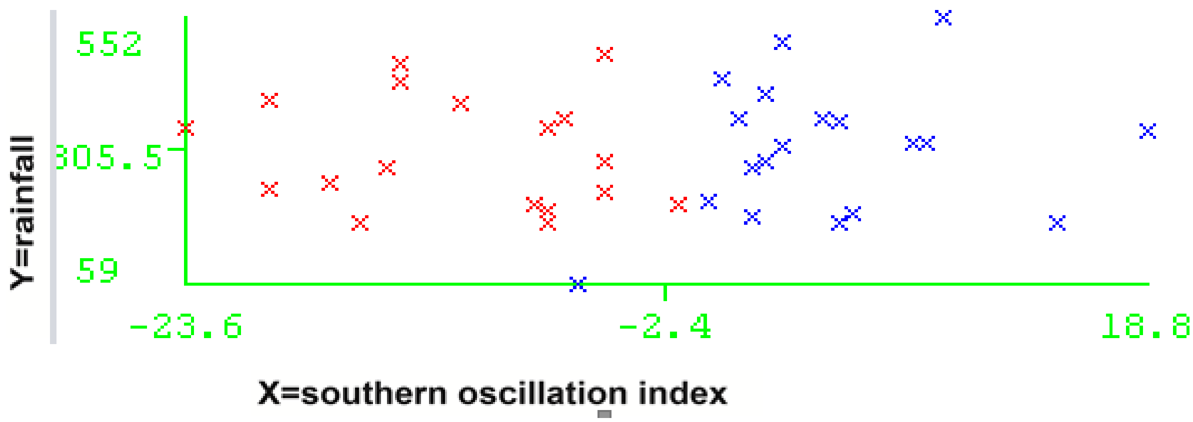



Figure 25

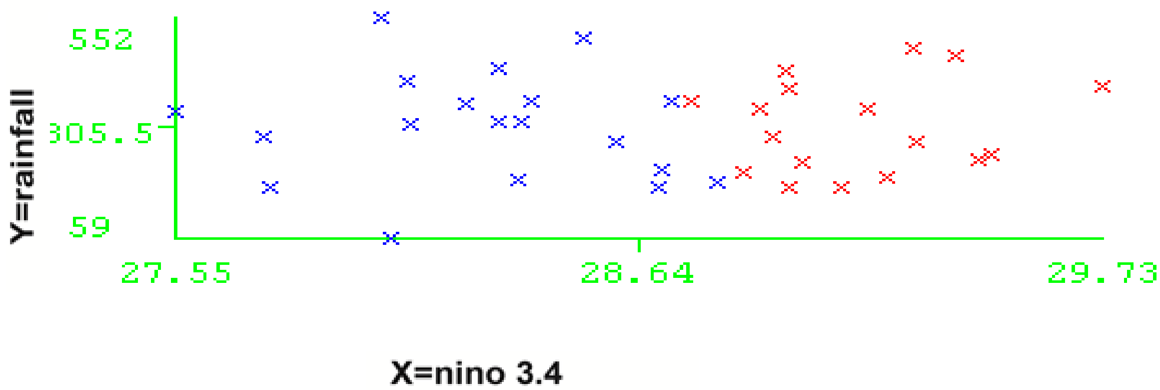



Figure 26

# Khulna August:

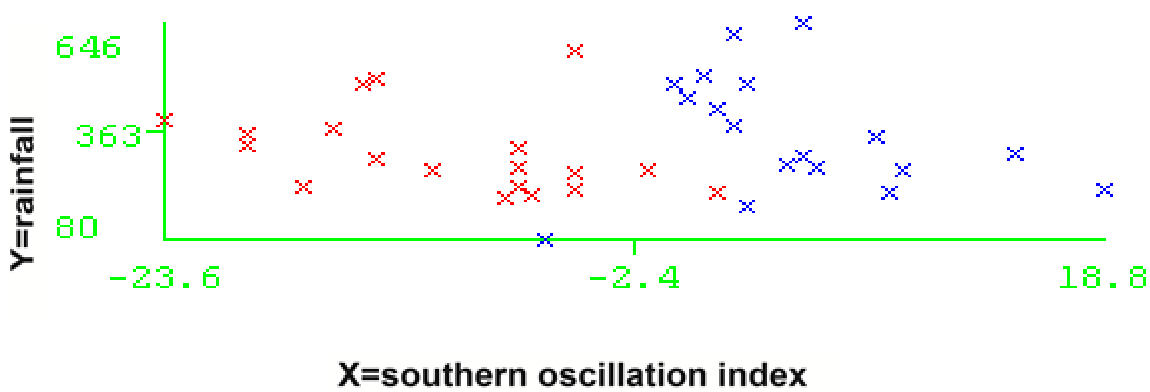



Figure 27

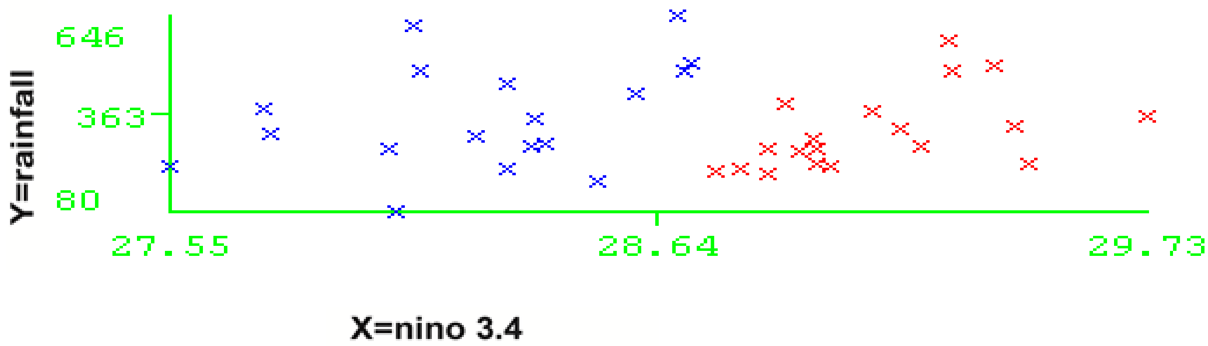



Figure 28

# Bogra August:

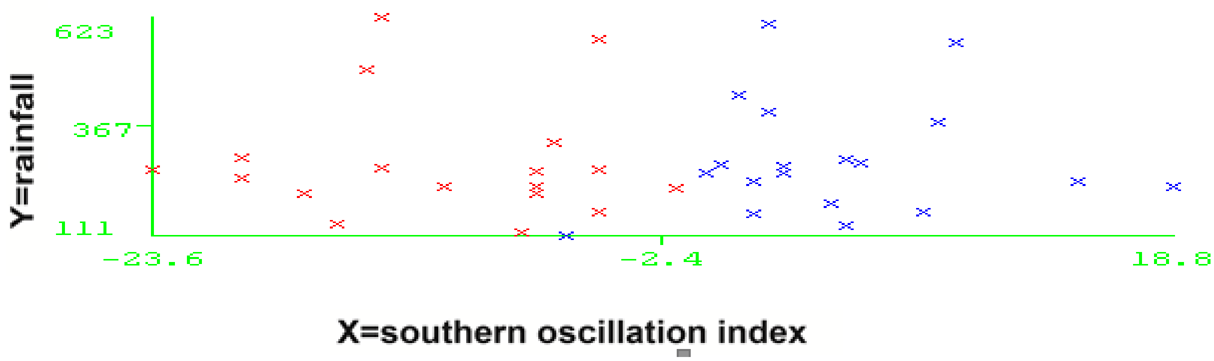



Figure 29

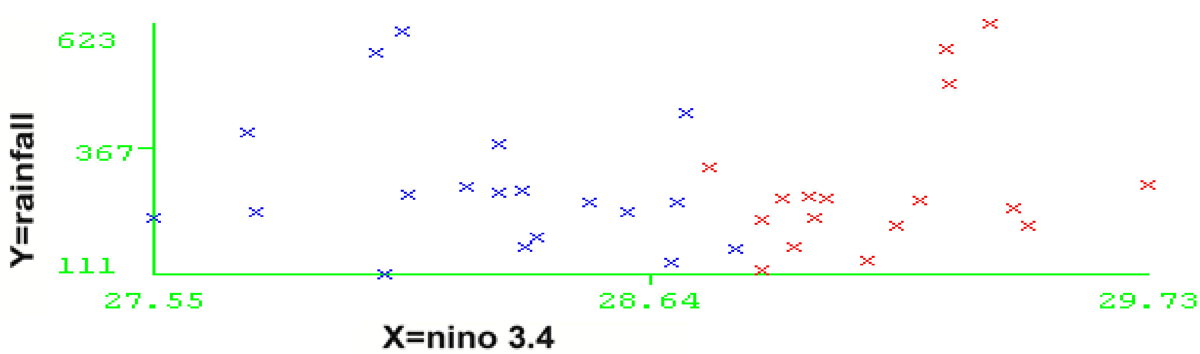



Figure 30

# Barisal August:

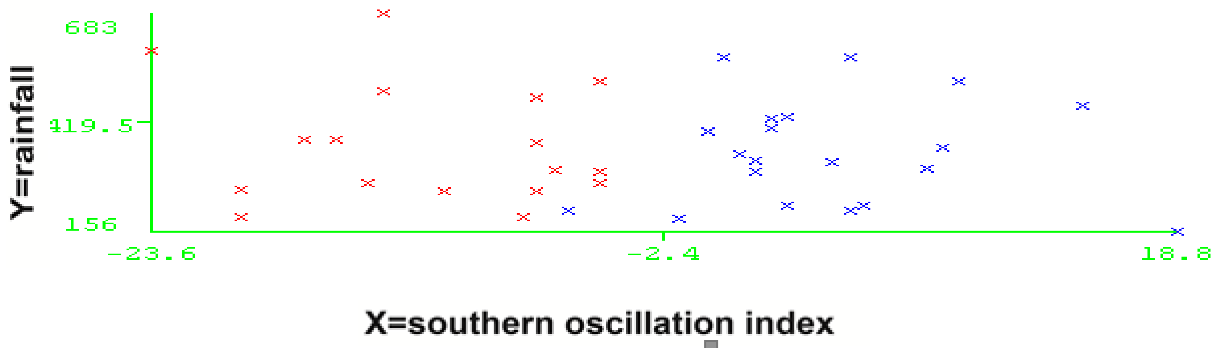



Figure 31

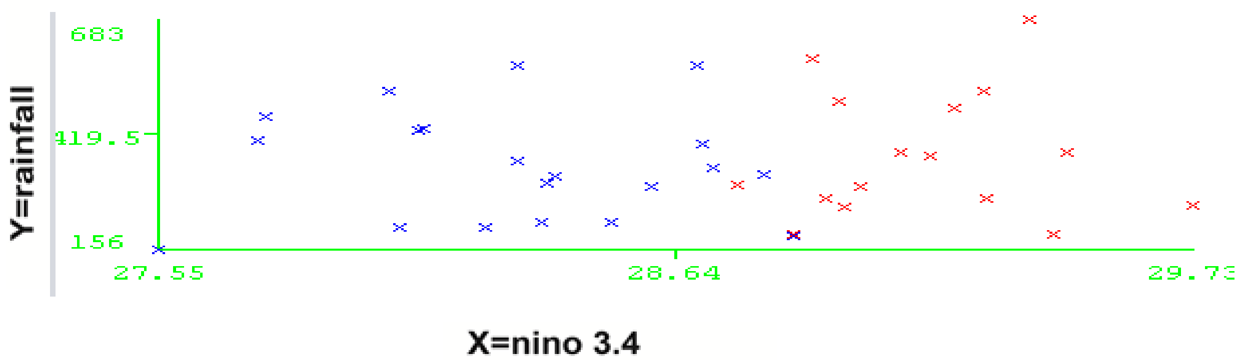



Figure 32

# Bhola August:

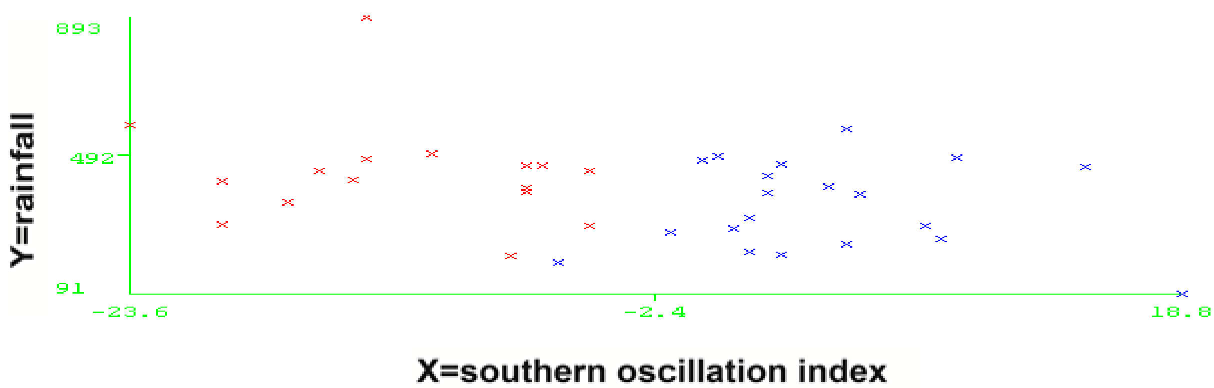



Figure 33

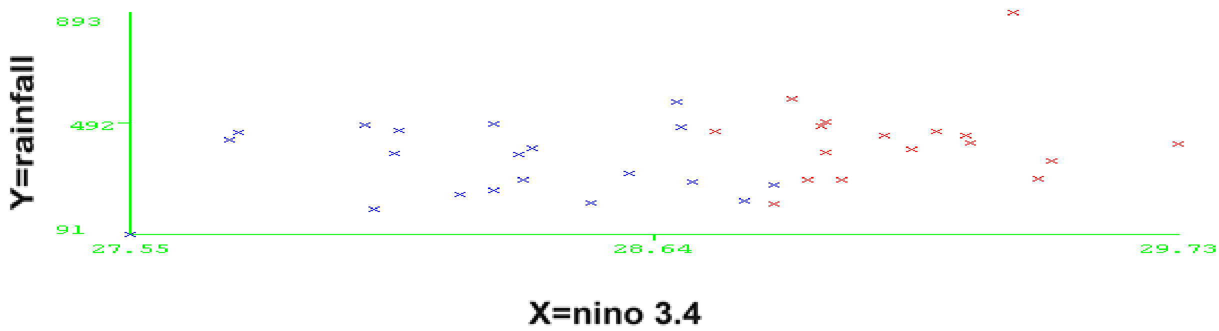



Figure 34

# Cox's Bazar August:

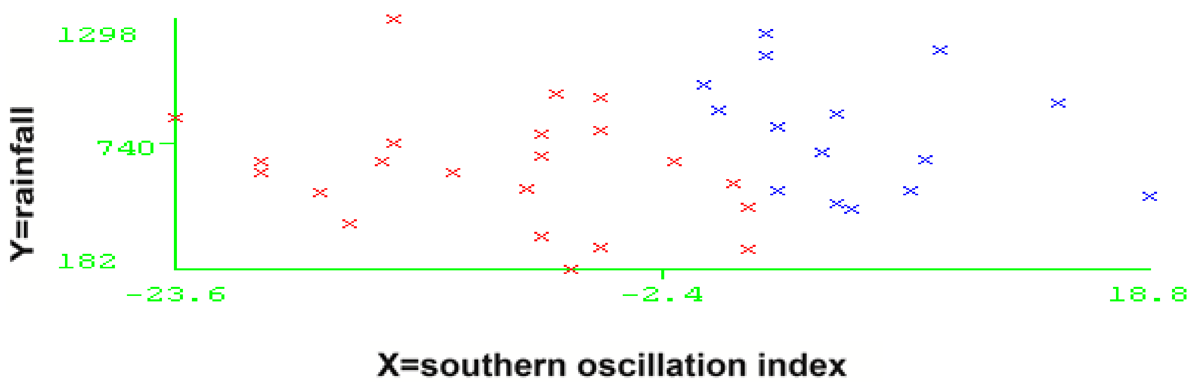



Figure 35

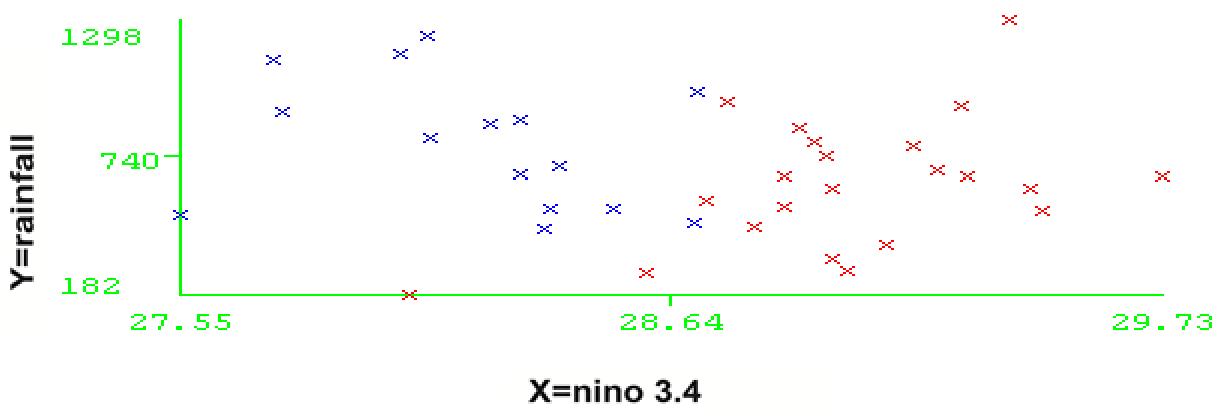
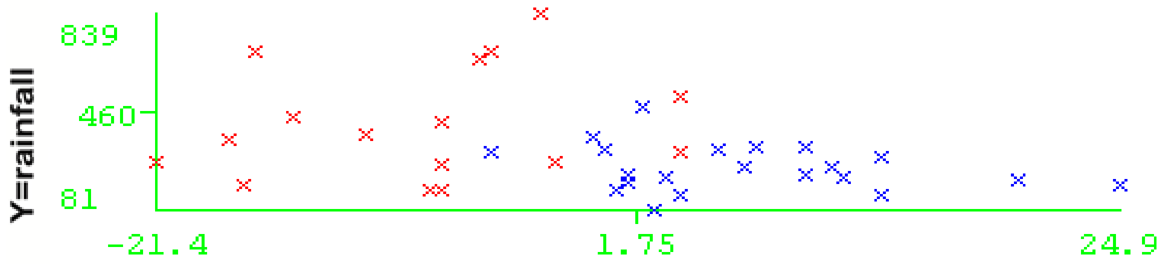


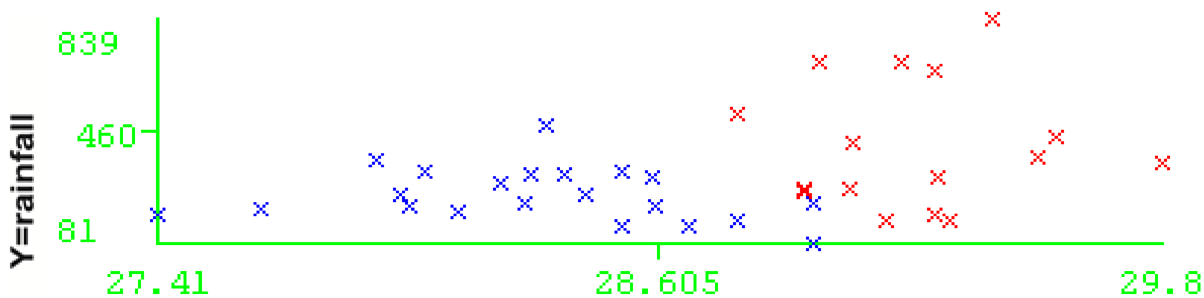

Figure 36

## Dhaka September:




X=southern oscillation index

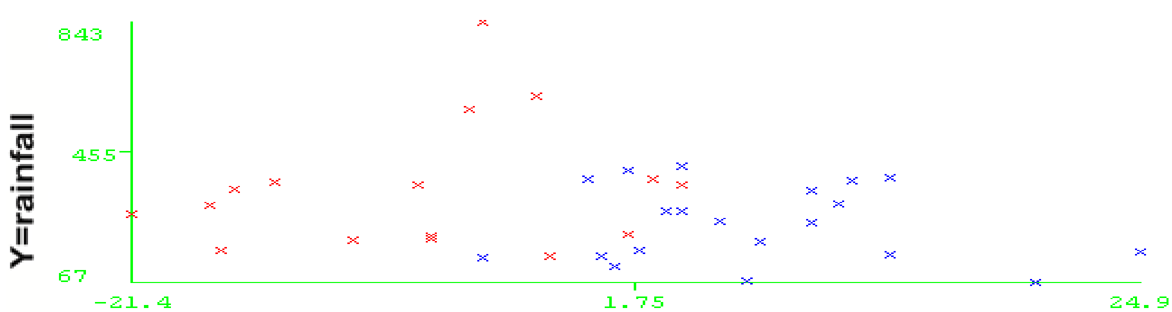
**Figure 34**




X=nino 3.4

**Figure 35**

## Khulna September:




X=southern oscillation index

**Figure 36**

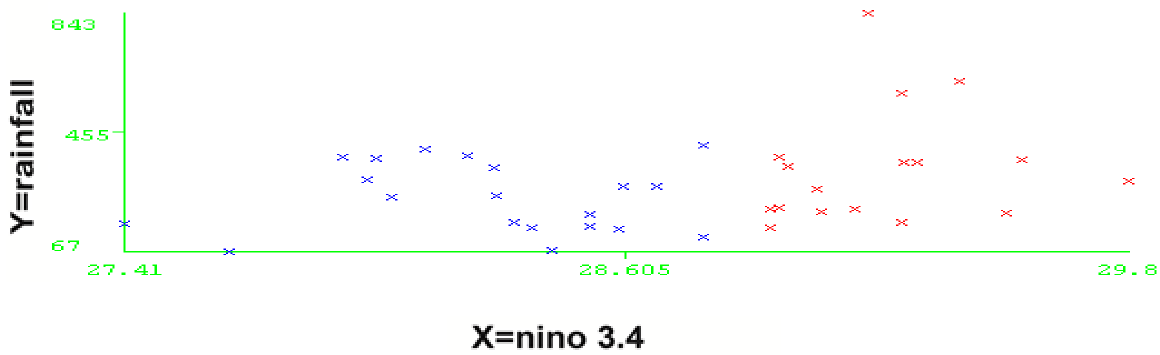



Figure 37

# Bogra September:

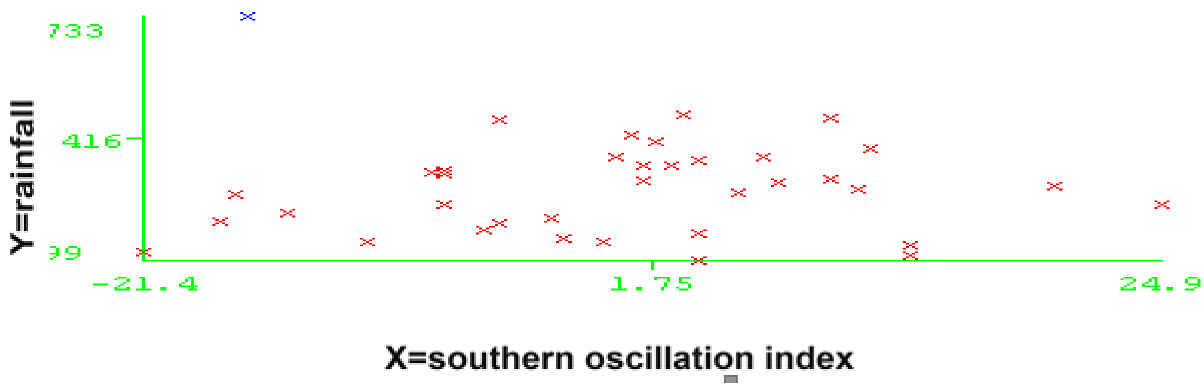



Figure 38

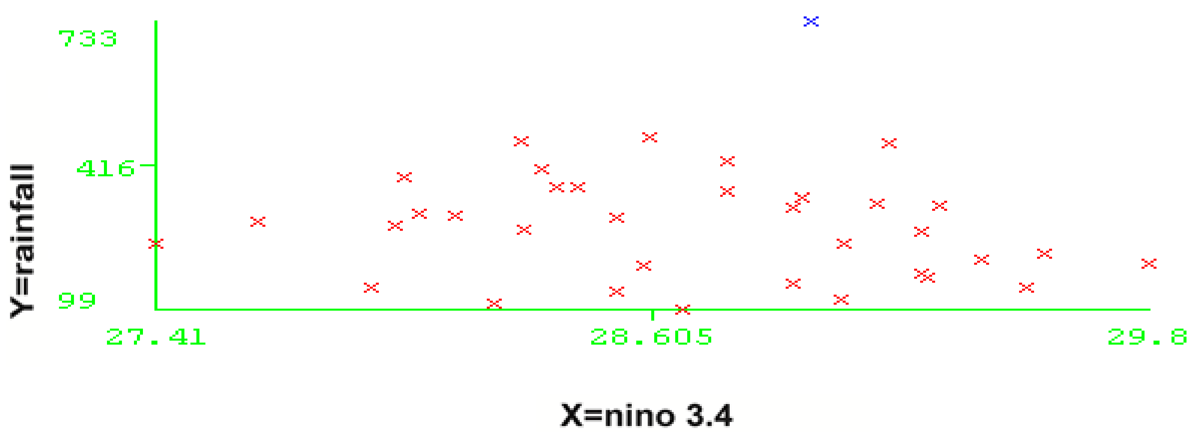



Figure 39

## Barisal September:

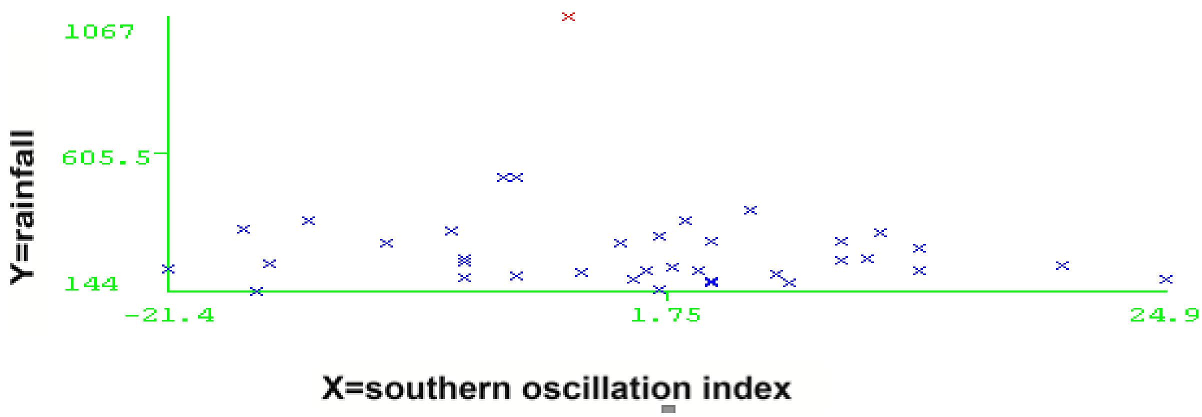



**Figure 40**

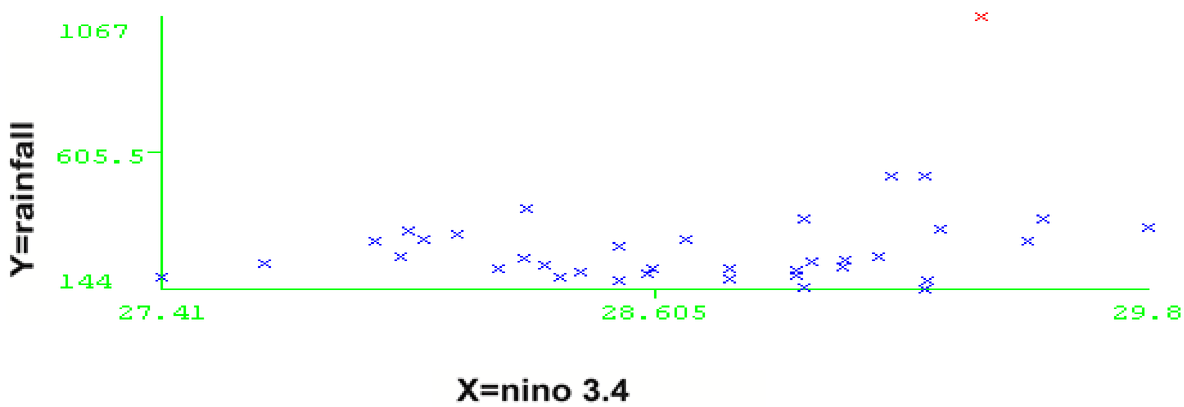



**Figure 41**

## Cox's bazar September:

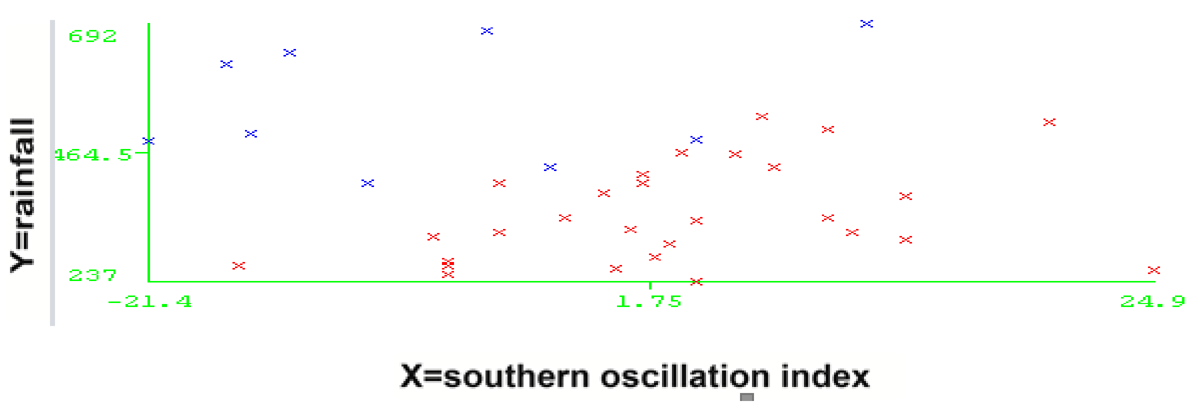



**Figure 42**

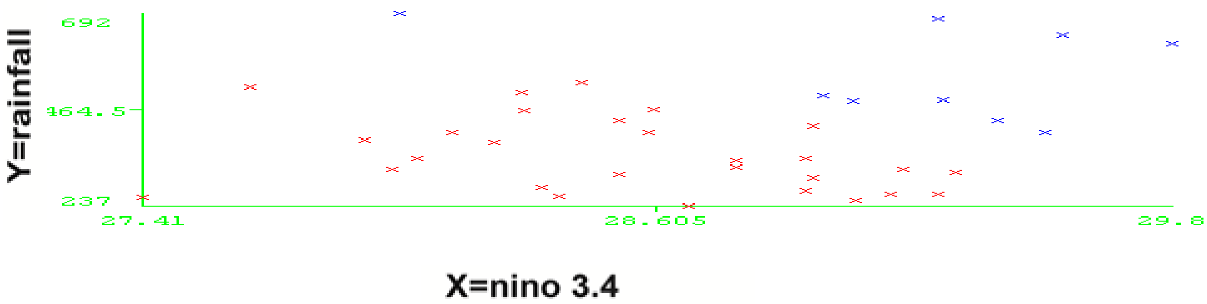



**Figure 43**

# Dhaka October :

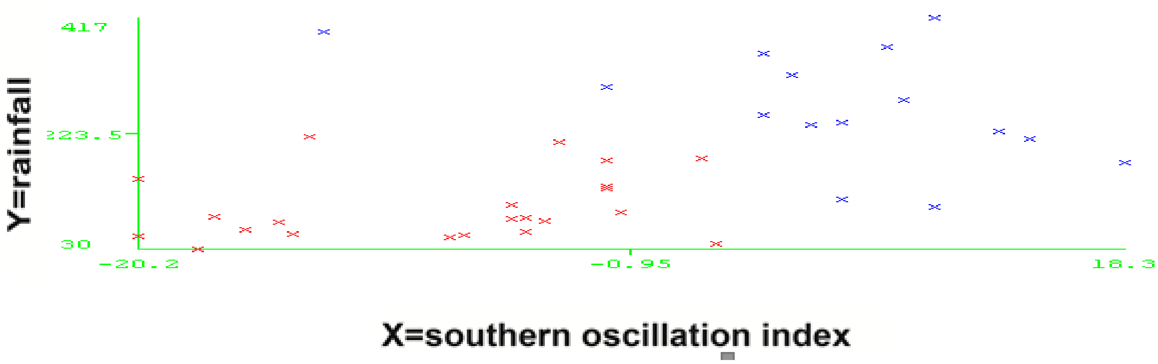



**Figure 44**

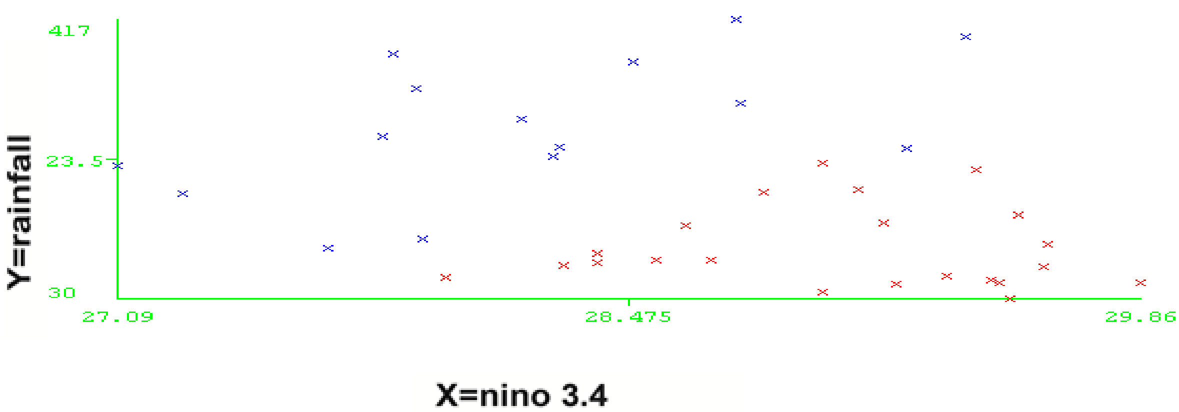



**Figure 45**

# Khulna October:

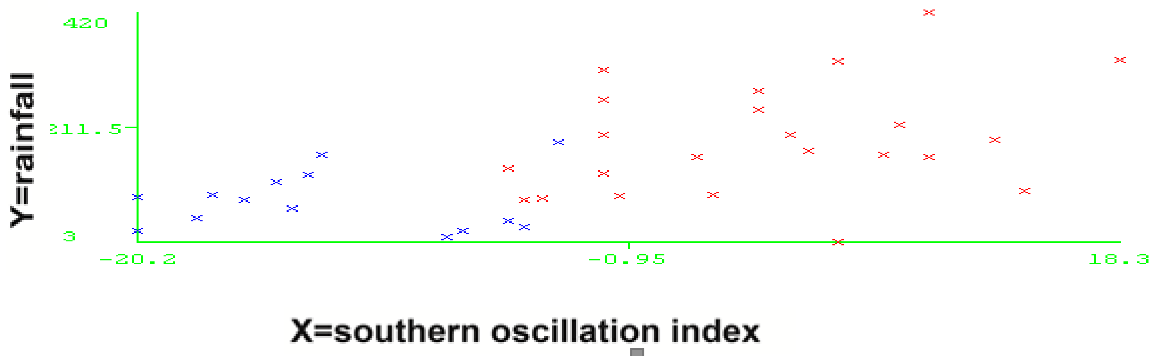



**Figure 46**

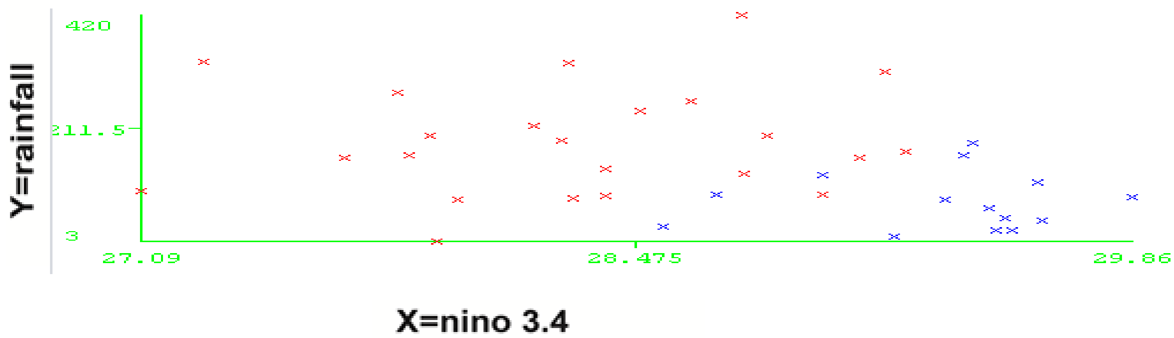



**Figure 47**

# Bogra October:

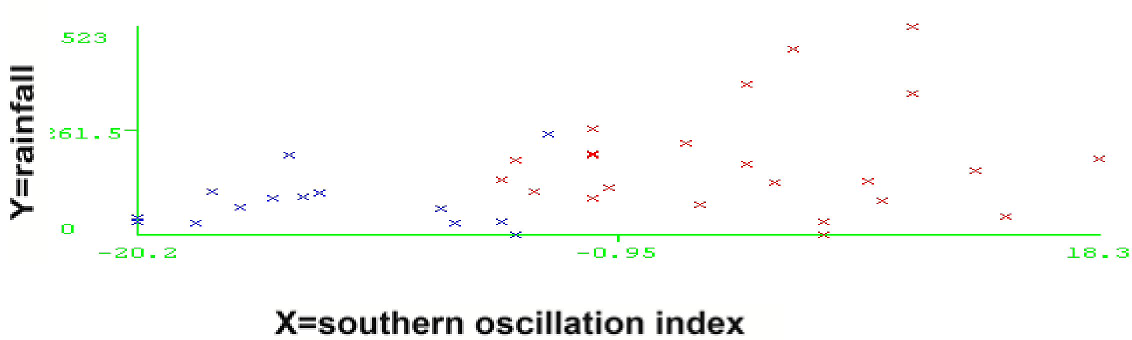



**Figure 48**

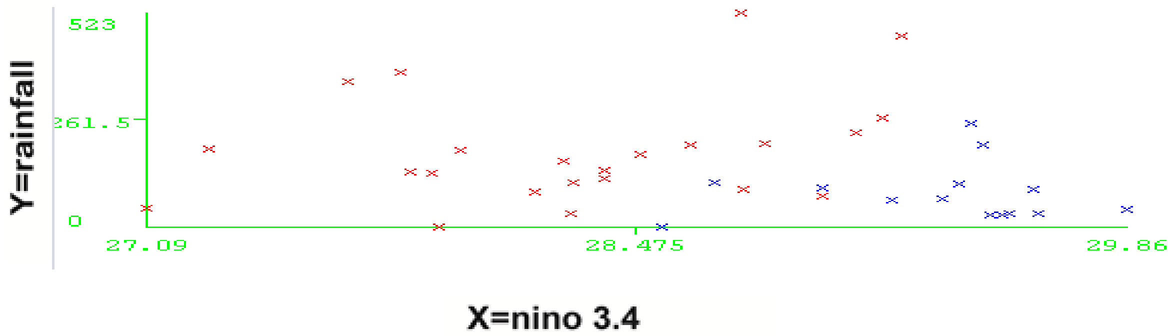



**Figure 48**

# Barisal October:

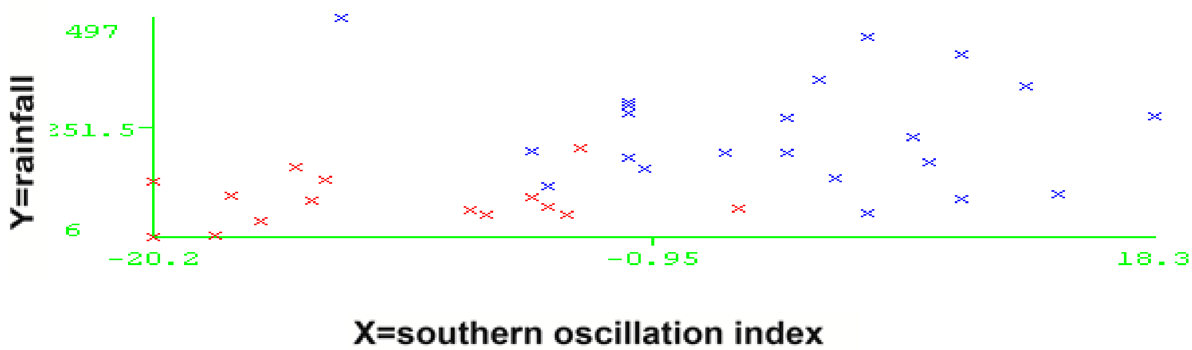



**Figure 48**

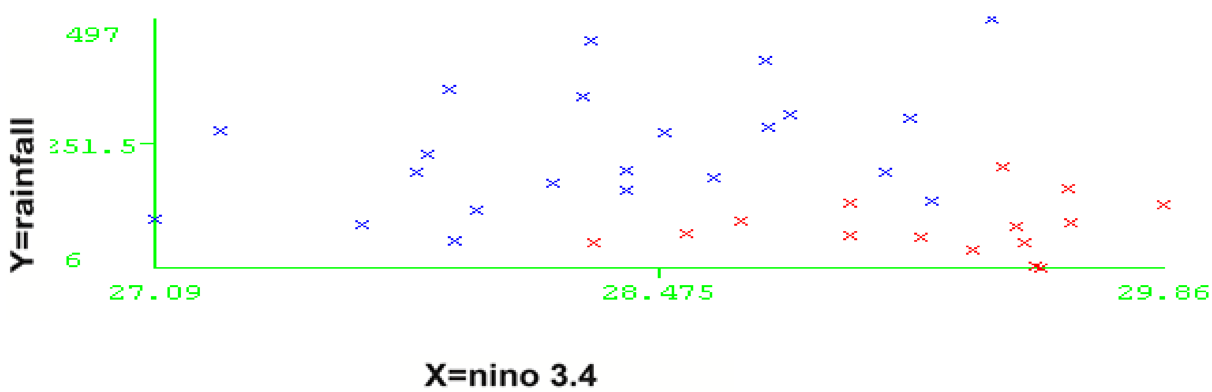



**Figure 49**

# Bhola October:

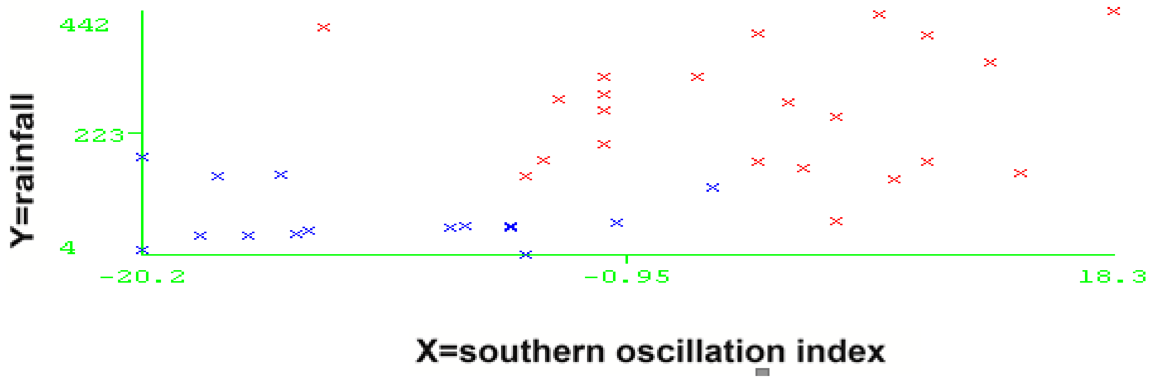



Figure 49

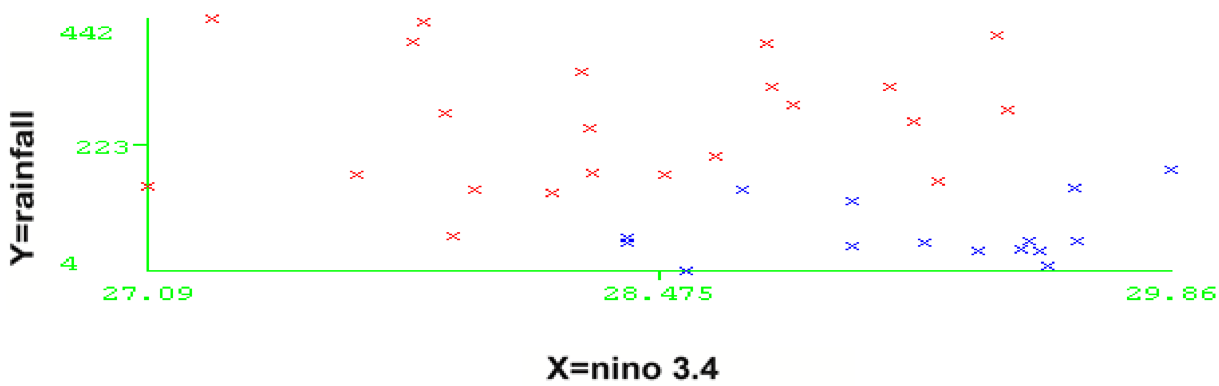



Figure 50

# Cox's Bazar October:

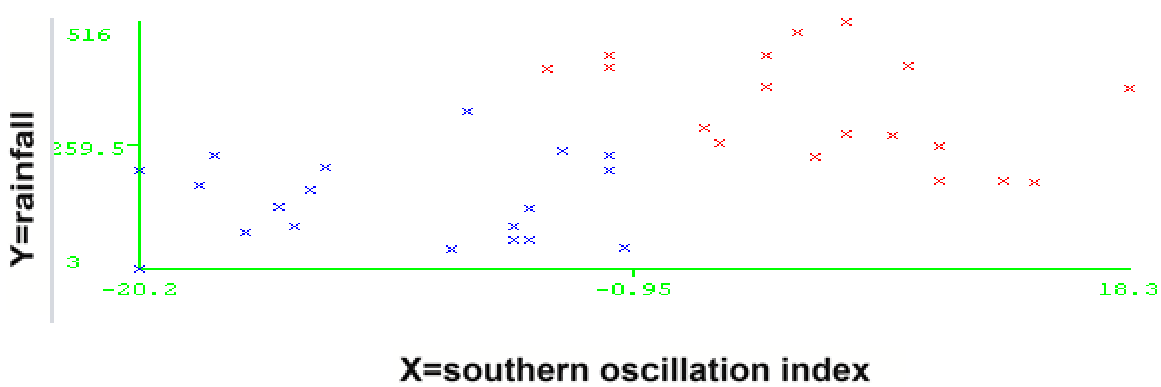



Figure 51

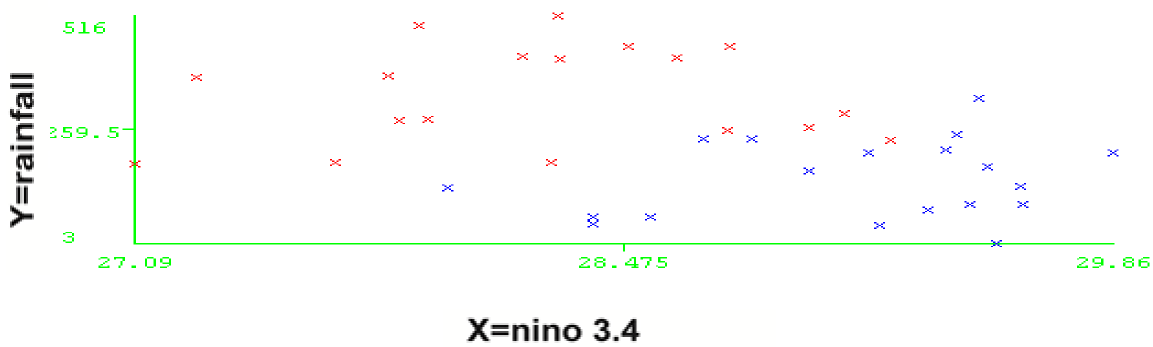



**Figure 52**

| 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 | 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 | 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 do 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 .

# 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. We 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. We 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.

# 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. A graph is shown in (a) for better understanding.

   The MOWCATL approach has more mechanisms for: (1) constraining the search space during the discovery process, (2) allowing a time lag between the antecedent and consequent of a discovered rule, and (3) working with episodes from across multiple sequences. As shown in the algorithm below,

1. ***MOWCATL Algorithm***:
2. Generate Antecedent Target Episodes of length 1 ($ATE_{1,\mathcal{B}}$);
3. Generate Consequent Target Episodes of length 1 ($CTE_{1,\mathcal{B}}$);
4. Input sequence $\mathcal{S}$, record occurrences of $ATE_{1,\mathcal{B}}$ and $CTE_{1,\mathcal{B}}$ episodes;
5. Keep episodes from $ATE_{1,\mathcal{B}}$ and $CTE_{1,\mathcal{B}}$ that meet *min_sup*;
6. $k = 1$;
7. *while*($ATE_{k,\mathcal{B}} \neq \emptyset$) *do*
8. Generate Antecedent Target Episodes $ATE_{k+1,\mathcal{B}}$ from $ATE_{k,\mathcal{B}}$;
9. For each episode
10. record minimal occurrences of length $\leq win_a$ timestamps;
11. Keep episodes from $ATE_{k+1,\mathcal{B}}$ that meet *min_sup*;
12. k++;
13. Repeat or execute in parallel, steps 6–12 for consequent episodes,
14. using $CTE_{k+1,\mathcal{B}}$ and $win_c$ instead of $ATE_{k+1,\mathcal{B}}$ and $win_a$, respectively;
15. Generate combination episodes $CE_{\mathcal{B}}$ from $ATE_{\mathcal{B}} \times CTE_{\mathcal{B}}$;
16. For each combination
17. record occurrences with *lag* between antecedent start and consequent start;
18. *return* supported episode rules in $CE_{\mathcal{B}}$ that meet the *min_conf* threshold;
19. End pseudo-code.

The MOWCATL algorithm needs a single database pass. MOWCATL starts by searching the target episodes that includes some events from the insertion constraint set and episodes that do not result in minimum support limit are cut off [1].

3. Next we would like to predict the increase in water level rise. The rising water level that causes most areas of Bangladesh, not only floods most of the land, but also affects the mortality rate. Back in the day, high mortality counts have been involved with extreme flood events [3].
4. We would also like to calculate our own Southern Oscillation Index and El Nino using the data mining techniques. Since data mining techniques contain a variety of methods: predictive modeling, clustering, associationmining, and change and deviation detection, predictive modeling includes classification forcategorical predictions and regression analysis for numerical predictions [4,5]. A few extra models can be used for the predictive purposes. The models can be both parametric or data driven. Parametric methods basicallyuse parameter estimation in statistics, it includes regression, discriminant analysis techniques, and autoregressive integrated moving average model [6].Data driven approach,

uses solving problems less idea about the statistical properties of the data and is better for difficulties with complex nonlinear data relationships [4]. The method uses computer strength to iterate as long as it establishes a good fit in the data sheet. Examples are like decision trees, neural networks, k- nearest neighbors and genetic algorithm.

5. Next we want to introduce a multi-platform app for the farmers. It will contain a database of the past records of flood, which of the months have the most rainfall, which crops are beneficial for the weather with respect to the season since rainfall is happening almost every month of the year.
6. We also want to include Bangla language for the app as it will be easier for the farmers to read them.

# Reference:


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 Specificationand Implementation. In Proceedings of the 1995 International Conference on the Principles and Practice ofConstraint 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”, HoldenDay, 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