International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Extreme Precipitation Events in Chennai Metro City Using Data Mining

R.Senthil Kumar, C.Ramesh

Abstract*: A information mining approach is displayed and connected to examine the climatic reasons for outrageous climatic occasions. Our methodology involves two primary strides of information extraction connected progressively, so as to decrease the trouble of the first informational index. The objective is to recognize an a lot littler subset of climatic factors that may in any case have the option to portray or even anticipate the outrageous occasions. The initial step applies a class correlation strategy. The subsequent advance comprises of a choice tree learning calculation utilized as a prescient model to outline set of measurably most huge atmosphere factors recognized in the past advance to classes of precipitation quality. The procedure is utilized to the investigation the climatic reasons for two outrageous occasions happened in India the most recent decade: the Chennai 2015 extraordinary precipitation disaster and the Tamilnadu(except Chennai) inadequacy of 2016. In the two cases, our outcomes are in great concurrence with investigations distributed in the writing.

Keywords* : Extreme event, Drought, Intense rainfall, KDD (Knowledge Discovery in Databases), Data mining, Classification, Decision tree.

I. INTRODUCTION

Tamil Nadu, located in southeast peninsular India, receives the major part of its annual rainfall during the northeast monsoon season (the three-month period from October to December). Coastal Tamil Nadu (including Chennai, Cuddalore, Nagapattinam, Ramanathapuram and Kanyakumari) receives about 60% of its annual rainfall and interior Tamil Nadu receives about 40-50% of annual rainfall during northeast monsoon. In comparison with Indian summer monsoon, the Northeast monsoon is characterized by limited aerial extent and average lesser rainfall amount[1-3]. During northeast monsoon season, Tamil Nadu generally receives rainfall due to the formation of tough of low, cyclonic circulation, easterly waves, low pressure area in the Coastal, depression and cyclonic storm over Bay of Bengal, because, the northeast monsoon season is the major rainy season in the Tamilnadu state. The vicissitudes of the rainfall of Tamil Nadu state has led to considerable and widespread interest among the public, farmers as well as in government circles in recent years, in view of the frequent failure of northeast monsoon rainfall over Tamil Nadu. There are several papers and documents to explain the relation between OLR and Northeast monsoon rainfall. The inter-annual variation of the outgoing long-wave radiation for the summer monsoon period showing a close association with the large-scale monsoon rainfall over India has been mentioned by Prasad and Verma (Prasad and Verma, 1985)[4-6]. They have concluded that the satellite-derived outgoing long-wave radiation can be used to monitor more comprehensively, the large-scale monsoon circulation and its year-to year variability, in view of its spatial coverage over oceanic areas. Prasad and Bansod have found the relationship between averaged OLR for west central India and the Indian summer monsoon rainfall to be stable (Prasad and Bansod, 1964). The inter-annual variability of Indian summer monsoon rainfall and Northeast monsoon rainfall is determined by external forcing and nonlinear internal dynamics. Surface air temperature is one of the factors that influence monsoon variability. The distribution of surface air temperature over land and sea determines the locations of heat source and sink which in turn affect circulation patterns through thermal and latent heat energy exchange between atmospheres and the surface beneath. A number of studies addressed the relationship between Indian summer monsoon and land and sea surface temperatures (Sikka, 1980; Verma. et al, 1985)[7].

Many studies (Rajeevan.et al, 1998; Pai, 2003) examined the global land surface air temperature anomaly patterns in association with inter annual variability of Indian summer monsoon rainfall[8]. Balachandran et al (2006) suggested that, in the correlation coefficient patterns, the positive correlation coefficient regions indicate that when the surface air temperature over these areas are Southeast India, especially Tamil Nadu and Puducherry experienced unprecedented rainfall activity during November and early December 2015 leading to devastating flood over Tamil Nadu[9]. Chennai was worst affected during the end of November and early part of December. The extremely heavy rainfall over north coastal Tamil Nadu including Chennai occurred in three different spells, viz., 8-9 Nov, 16-17 November and 30 Nov.-1- Dec., 2015. Details of the synoptic situations for these three spells are presented. It was mainly due to a Deep depression over southwest Bay of Bengal, well marked low pressure area over southwest bay of Bengal and a trough of low with embedded cyclonic circulation extending up to middle troposphere level respectively. The performance of the IMD GFS and WRF model for these three extreme spells are evaluated and presented[10-13].

s

Manuscript published on 30 September 2019.
***Correspondence Author(s)**
R. Senthil Kumar*, Research Scholar, Department of Computer Science and Engineering, Satyabama University, Chennai.
Dr. C. Ramesh, Research Supervisor, Satyabama University, Chennai.

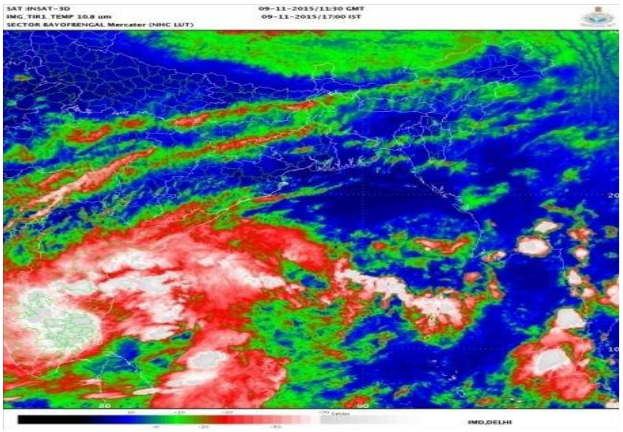


Fig.1a INSAT-3D satellite imagery of DD (08-10 Nov 2015) based on 09th/1130 UTC, Doppler Weather Radar, Chennai imagery based on 09th/1400 UTC

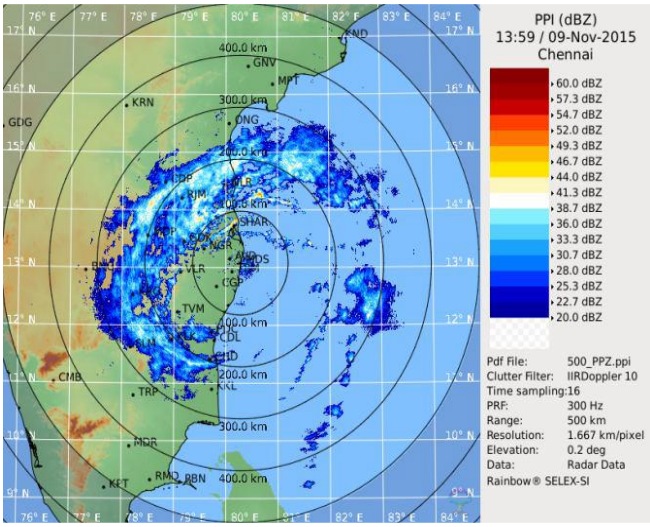


Figure 1-b- Spatial distribution of 24-hr heavy rainfall occurences as on 0300 UTC of 08-11 November 2015).

Data mining − a step in the more general process of knowledge discovery in databases (KDD) − attempts to uncover hidden patterns in large data sets. Its main goal is to extract information from a data set and transform it into an understandable structure for further use, in order to facilitate a better interpretation of existing data [8]. These patterns can be seen as a kind of summary of the input data and may be used in further analysis. Data mining may, for instance, identify multiple clusters or subsets in the data, which can then be used to obtain more accurate prediction results by a decision support system. For more several decades, climatologists have been using data mining techniques in an assortment of studies. For a review see [9] and references therein. However, within the particular context of extreme rainfall-associated events, data mining technologies were applied in a relatively small number of studies[10,11]. Here it is present an innovative data mining approach to investigate the climatic causes of extreme events such as the Chennai of 2015 tragedy, and the Chennai of 2016 droughts. Our approach comprises two main steps of knowledge extraction, applied successively in order to reduce the complexity of the original dataset, and identify a much smaller subset of climatic variables that may explain the event being studied. In the first step, it is follow along the lines of [14], and apply a class comparison technique commonly used as a tool to analyze large data sets of genome-wide studies. This step results in a series of-spatial fields that identify which climatic variables behave differently across pre-defined classes of rainfall intensity. More generally, it permits to identify coherent spatial patterns that might indicate the existence of plausible links between different climate subsystems. The second step consists of a decision tree (DT) learning algorithm used as a predictive model to map the set of statistically most significant climate variables identified in the previous step to classes of rainfall intensity. A decision tree is a flowchart-like structure in which internal nodes represent tests on attributes, each branch represents the outcome of a test, and each leaf node represents a class label. A path from the root to a given leaf represents a set of classification rules [15]. In the present context, the final result identifies a small subset of climatological variables that may explain or even forecast the extreme event in study.The remainder of this paper is organized as follows. Section 2 presents the methodology and data sets used in this investigation. Section 3 presents our results, while in Section 4 conclusions and discuss some further developments.

II. METHODOLOGY

The data mining approach here employed comprises two main steps of knowledge extraction: 1)Class-Comparison, and 2)Decision Trees. These methods are applied successively to reduce the difficulty of the original dataset (obtained from Chennai Meteorological Department: http://www.imdchennai.gov.in/) and recognize a much smaller subset of climatic variables that may explain the event being studied.

A.. Class-Comparison

Class comparison methods are used for comparing two or more pre-defined classes in a data set. Here, it is apply the class-comparison to time series of climatic grid box values or indices, but not to the entire fields. The objective is to determine which variables in our data set behave differently across pre-defined classes of rainfall strength (“high”, “neutral”, and “low”, for example). The “no-difference” case corresponds to a null hypothesis. The classes are defined in such a way so as to captured in the correct class the main episodes of drought or extreme rainfall that occurred during the period being evaluated. There are several methods for checking whether differences in variable values are statistically significant[16]. The F-test is a generalization of the well-known t-test, which measures the distance between two samples in units of standard deviation. Large absolute values of the F-statistic suggest that the observed differences among classes are not due to chance, and that the null hypothesis can therefore be rejected. Supposing there are J1 data points of class 1 and J2 data points of class 2, the t-test score is computed as:

Retrieval Number: J99780881019/19©BEIESP
DOI: 10.35940/ijitee.J9978.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s_p^2\left(\frac{1}{J_1} + \frac{1}{J_2}\right)}}, \qquad (1)$$

where,

$$s_p^2 = \frac{(J_1-1)s_1^2 + (J_2-1)s_2^2}{J_1+J_2-2}, \qquad (2)$$

and for i=1,2,

$$s_i^2 = \frac{1}{J_{i-1}} \sum_{j=1}^{J_i} \left(x_{ij} - \overline{x_i}\right)^2, \qquad (3)$$

where $\overline{x_1}$ = mean of samples class 1,
$\overline{x_2}$ = mean of samples class 2.

For more than two classes, an F-statistic shall be computed. In this case, the alternative to the null hypothesis is that at least one of the classes has a distribution that is different from the others. The t-test and F-test scores may be converted into probabilities, known as p-value. A p-value is the probability that one would observe under the null hypothesis a t-statistic (or F-statistic) as large as or larger than the one computed from the data. Both the t-test and F-test assume that the means are normally distributed, which may not hold, particularly when the number of data points is small. In this case, one could use the non-parametric counterparts of these tests, such as the Wilcoxcon test, the Kruskal-Wallis, or a permutation method. The probability of observing an F-statistic as large as or larger than the one computed from the data is called a "p-value". It is a measure of statistical significance in the sense that one expects to observe, under the null hypothesis, p-values less than 0.01 only 1% of the time. Permutations methods, which do not rely on data normality assumptions, are commonly used for computing p-values [17]. For this, after calculating t-test scores for each variable, the class labels of the J1 and J2 are randomly permuted, so that a random J2 of the samples are temporarily labeled as class 1, and the remaining J2 samples are labeled as class 2. Using these temporarily labels, a new t-test score is calculated, say t*. The labels are then reshuffle many times again, with a t* being computed at each permutation. The p-value from the permutation t-test is given by:

p-value from the permutation t-test is given by:

$$\text{p-value} = \frac{1 + \#\ of\ random\ permutation\ where\ |*| \geq |t|}{1 + \#\ of\ random\ permutation} + \cdots (4)$$

B. Decision Tree

Usually, DT learners use the divide-and-conquer strategy to construct a suitable tree from a training set. For this, the problem is successively divided into smaller sub-problems until each subgroup addresses only one class, or until one of the classes shows a clear majority not justifying further divisions. Most algorithms attempt to build the smallest trees without loss of predictive power. To this end, the J4.8 algorithm relies on a partition heuristic that maximizes the "information gain ratio", the amount of information generated by testing a specific attribute. This approach permits to identify the attributes with the greatest discrimination power among classes, and select those that will generate a tree that is both simple and efficient.

The information gain is measured in terms Shannon's entropy reduction. Given a set A with two classes P and N, the information content (in bits) of a message that identifies the class of a case in A is then

$$I(p,n) = -\frac{p}{p+n} log_2\left(\frac{p}{p+n}\right) - \frac{n}{p+n} log_2\left(\frac{n}{p+n}\right), \text{ ---------- (1)}$$

where is the total number of objects belonging to class , and n is total number of the objects into the classes N. If A is partitioned into subsets A1, A2, … , Av by a given test T, the information gained is given by

$$G(A;T) = I(A) - \sum_{i=1}^{v} \frac{p_i + n_i}{p+n} I(A_i), \text{ ----------(2)}$$

where Ai has pi objects from the class P, and ni from the class N. The algorithm chooses the test T that maximizes the information gain ratio G(A;T)/P(A;T), with

$$P(A;T) = -\sum_{i=1}^{v} \frac{p_i + n_i}{p+n} log_2\left(\frac{p_i + n_i}{p+n}\right), \text{ ---------(3)}$$

being the information gain from the partition itself. The process is repeated recursively to obtain the other nodes, structuring the decision tree with the rest of the subsets.

III. RESULTS

The climatic causes of the Chennai 2016 tragedy and the Tamilnadu droughts of 2005 and 2010 are investigated. The entire data sets used in the analysis can be freely downloaded from the Web(obtained from Chennai Meteorological Department: http://www.imdchennai.gov.in/). Surface- and pressure-level atmospheric fields have a spatial resolution of 2.5ox2.5o and were extracted from NCEP/NCAR Reanalyzes. Sea Surface Temperatures (SSTs) on a 2ox2o grid were obtained from the NOAA Optimum Interpolation SST Analysis, version 2[20]. The objective of this study is to determine which variables in our dataset behave differently across pre-defined classes of precipitation intensity. The "no-difference" case corresponds to the null hypothesis for the applications considered here.

A. Extreme Rainfall over Chennai of Tamilnadu

The data set used in this study comprises 3.781 time series (Table 1). Gridded data cover a region delimited by latitudes 20oS and 50oS, and longitudes 30oW and 60oW. Since the episode of extreme rainfall in Chennai was an event of short duration, pentad-averaged anomalies were used in the analysis.

Table 1: Data set used in this study

Variable	Units	Observations	No. of Time Series
Sea Surface Temperature	oC	Surface	147
Sea Level Pressure - SLP	Pa	1000 hPa	170
Air Temperature	oC	Surface	170
Specific Humidity	g/kg	At 850 and 1000hPa	331
Omega	Pa/s	At 100,200,300,400,500,600,700,850, and hPa	1478
Geopotential Height	m	At 1000hPa	174
Zonal Wind	m/s	At 200,500,850 hPa	512
Meridional Wind	m/s	at 200,500,850 hPa	516
Cloud Cover	%	Surface	174

A. Class-Comparison

The goal is to identify variables that might correlate with observed differences among classes of rainfall in the region of Chennai (red colored dots in Figs.3 to 5), one of the most affect areas by the 2015 disaster. To this end, we analyzed 12 years (January 1994 up to December 2016) of pentad averages, comprising 3,781 environmental variables. Precipitation data in the region of Chennai (Fig.2) is an average of five measurement stations of Chennai Meteorological Department. For classification purposes, the pentads of this time series were divided in three classes of precipitation intensity: “strong”, “moderate”, and “light” rainfall. The standard t-test (eq.1) was applied, as recommended for applications with two classes: “strong” (precipitation greater than 8), and “moderate” (precipitation between 0 and 8). Results for the most significant variables identified by this procedure are presented in Figs.3 to 5. These results represent p-value fields, where coherent spatial patterns of low p-values indicate the existence of a possible links between omega and zonal/meridional wind anomalies, at different levels, and the precipitation intensity in the region of Chennai (Fig.2). The red isolines in Fig. 3 and 4 correspond to omega anomalies averaged over the period November 22 up to 26, 2015, the period of most intense precipitation in and around Chennai (delimited by the red bars in Fig.1). The wind fields in Fig.4 are also anomalies averaged over the same period.

Table 2:Time Series Data Of Rainfall By Seasons (Last 20 Years)-2015-2016

S.NO	Year	Hot Weather Season		South West Monsoon		North East Monsoon		Winter Season		Total		Deviationof Percentage
		Normal	actual	Normal	actual	Normal	actual	Normal	actual	Normal	actual	
1.	1994-95	135.1	164.4	158.2	754.3	328.2	567.6	25.6	0.1	647.2	956.3	+309.1
2.	1995-96	135.1	177.3	158.2	579.9	328.2	194.1	25.6	0.5	647.2	956.3	+309.1
3.	1996-97	135.1	126.6	158.2	181.4	328.2	340.5	25.6	1.2	647.2	649.7	+2.5
4.	1997-98	135.1	75.0	158.2	167.7	328.2	571.9	25.6	0.6	647.2	890.2	+243.0
5.	1998-99	135.1	69.8	158.2	229.7	328.2	434.8	25.6	16.0	647.2	750.3	+103.1
6.	1999-00	135.1	92.3	158.2	87.1	328.2	504.7	25.6	68.7	647.2	752.6	+105.4
7.	2000-01	135.1	141.9	158.2	339.0	328.2	179.8	25.6	5.0	647.2	665.7	+18.5
8.	2001-02	135.1	66.20	158.3	152.4	328.2	327.0	25.6	6.1	647.2	551.8	-14.74
9.	2002-03	135.1	69.6	158.3	78.6	328.2	62.8	25.6	17.6	647.2	228.6	-64.67
10	2003-04	148.4	202	192.9	90.1	327	205.4	26.1	16.7	694.4	514.2	-25.9

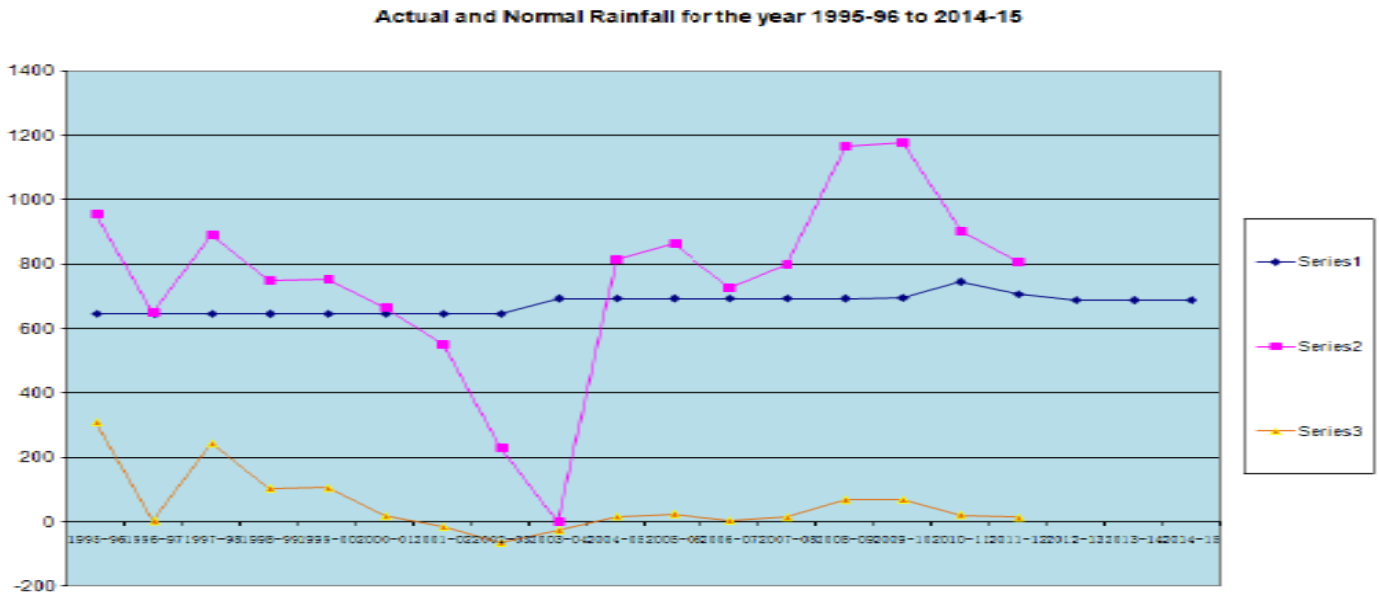
Retrieval Number: J99780881019/19©BEIESP
DOI: 10.35940/ijitee.J9978.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

11	2004-05	148.4	294.7	192.9	233.3	327.0	260.2	26.1	26.6	694.4	814.8	17.3
12	2005-06	148.4	162.1	192.9	177.6	327.0	505.7	26.1	17.7	694.4	863.1	24.3
13	2006-07	148.4	128.4	192.9	141.5	327.0	444.3	26.1	11.1	694.4	725.3	4.4
14	2007-08	148.4	190.8	192.9	204.3	327.0	378.0	26.1	25.0	694.4	798.1	14.9
15	2008-09	148.4	157.3	192.9	695	327.0	312.2	26.1	1.3	694.4	1165.8	67.89
16	2009-10	150.2	101.5	192.9	765.4	327	306.1	26.1	4.8	696.2	1177.8	69.2
17	2010-11	150.3	194.6	233.1	188.0	341.9	437.0	20.3	82.9	745.6	902.5	21.0
18	2011-12	168.0	140.4	189.8	252.9	328.9	410.7	20.3	2.6	707.0	806.6	14.1
19	2012-13	150.3	121.2	189.8	162.4	320.9	278.5	20.3	57.8	689.3	619.9	-11.2
20	2013-14	150.3	141	189.8	596.7	328.9	257.9	20.3	5.1	689.3	1000.7	45.2
21	2014-15	150.3	342.3	189.8	764.2	328.9	311.6	20.3	0	689.3	1418.1	105.7
22	2015-16	150.3	118	189.8	309.4	328.9	341.1	20.3	23	689.3	791.5	15

Source: Directorate of Economics and Statistics, Chennai.

mm and Extremely Heavy: ≥244.5 mm; Spatial distribution: Isolated (ISOL): 1-25% of stations reporting rainfall,



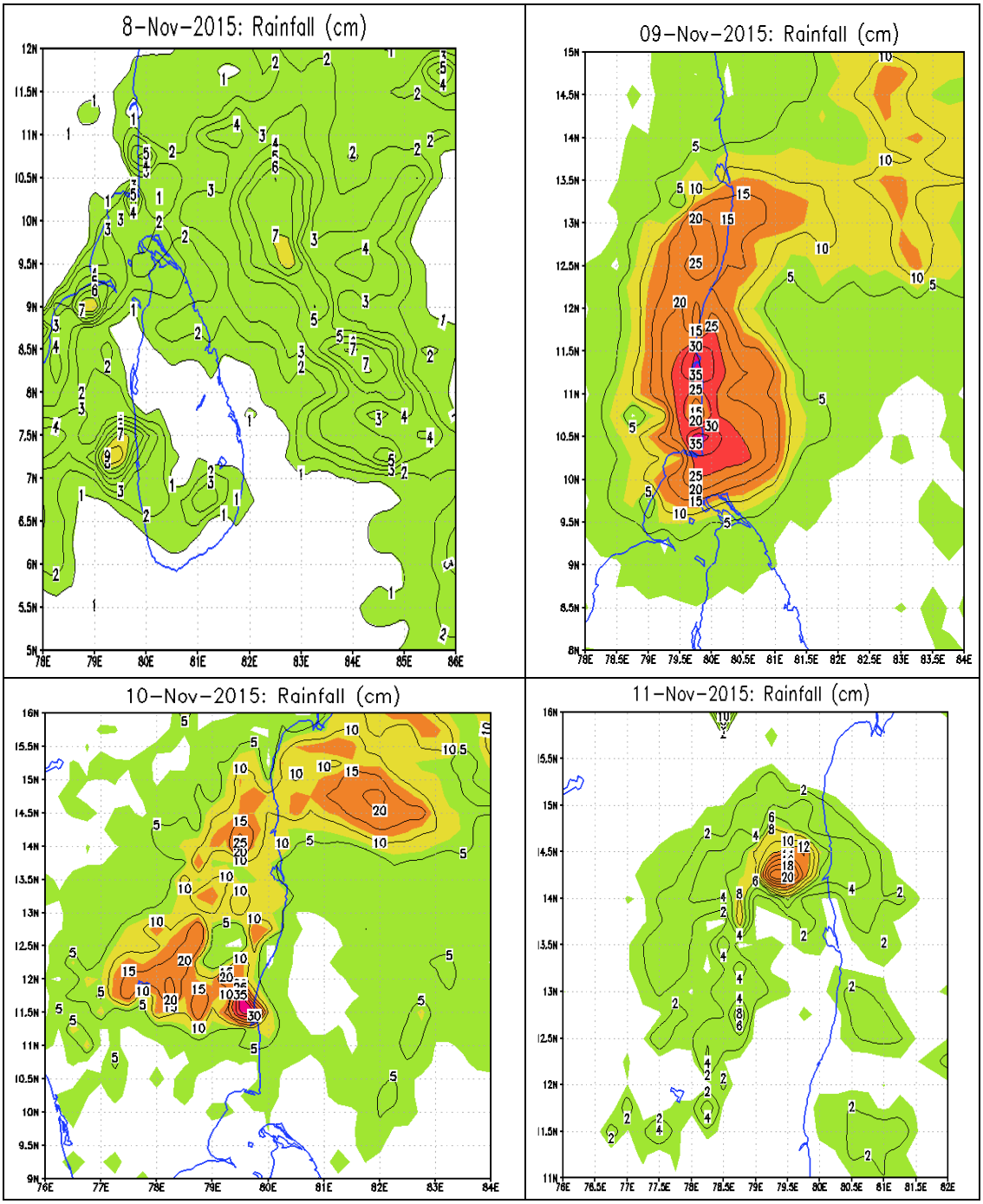
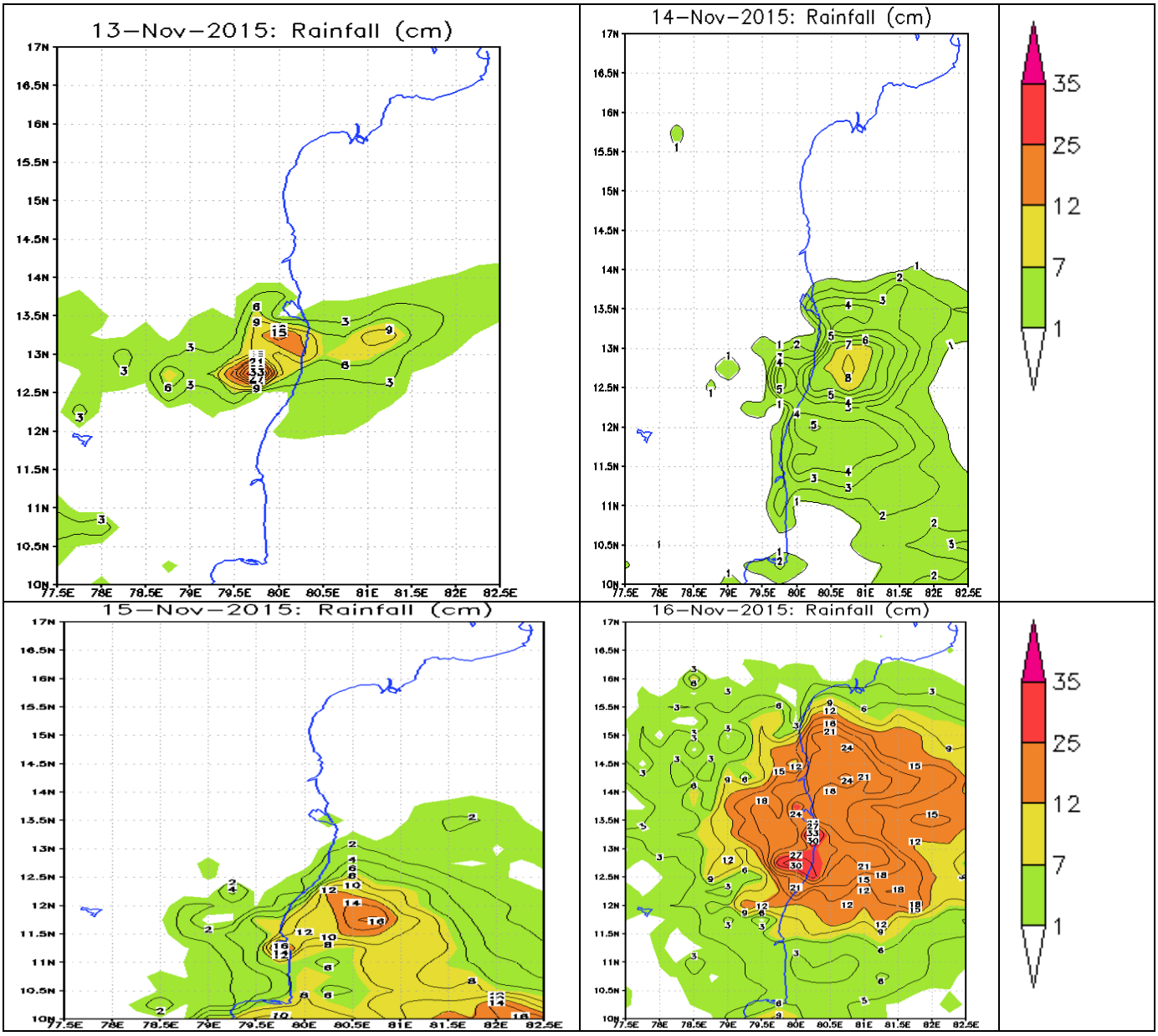


Fig.2c Rainfall realised in association with Deep Depression (08-10 Nov 2015) (based on IMD-NCMRWF GPM gauge merged rainfall data)

Well marked Low pressure area over South West Bay of Bengal (12-18 November 2015) Associated with this system, rainfall activity occurred over coastal Tamil Nadu during 12-18 November. Heavy to extremely heavy rainfall occurred along the entire north Tamil Nadu coastal belt during the 24-hr ending 0300 UTC of 16th and isolated heavy to very heavy rainfall on the other days. Spatial distribution of 24-hr heavy rainfall occurences as on 0300 UTC of 13-19 November 2015) is depicted in Fig.2b.

Retrieval Number: J99780881019/19©BEIESP
DOI: 10.35940/ijitee.J9978.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)



Retrieval Number: J99780881019/19©BEIESP
DOI: 10.35940/ijitee.J9978.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

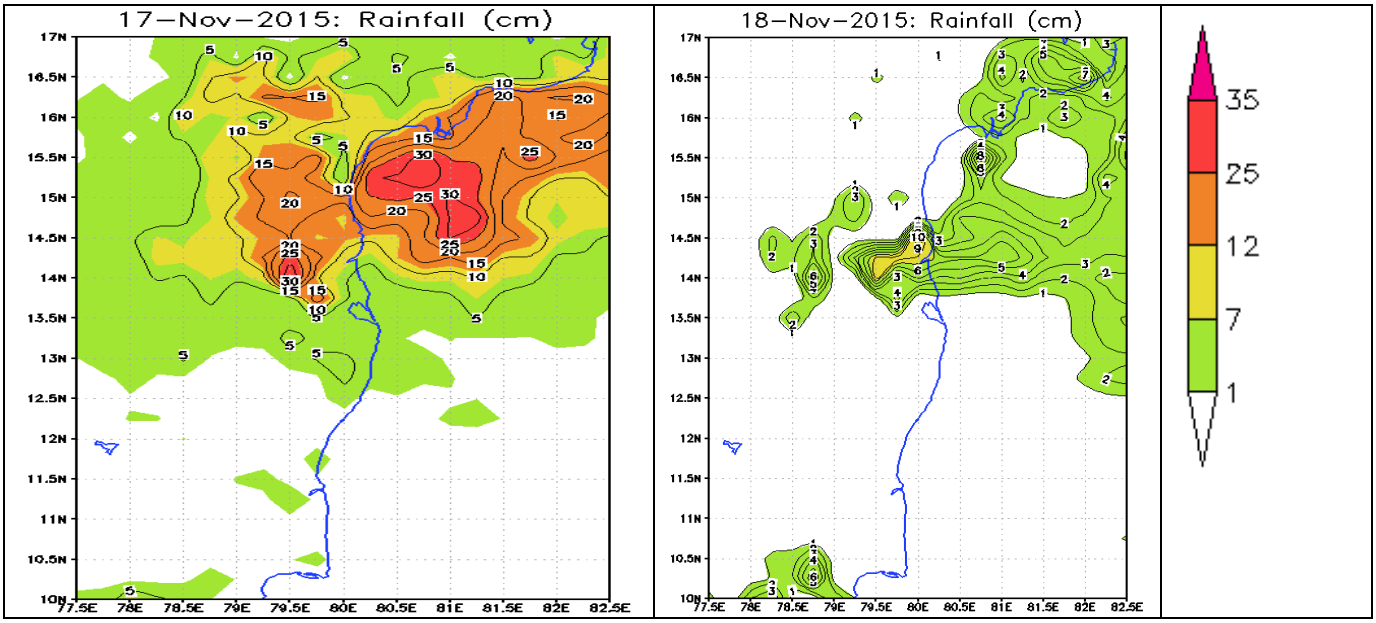


Fig.2b 24-hr accumulated rainfall as on 0300 UTC of 13-18 November 2015 in association with well marked low of 12-18 Nov 2015 (based on IMDNCMRWF GPM gauge merged rainfall data)

Regions with darker shades indicate the grid parameters with lower p-values. A p-value<0.01, for example, indicates probability lower than 1% of being a false positive. Figures 3 and 4 show a dense dark area of low p-values for omega at different levels, which extends from the Chennai city of Tamilnadu state. During the extreme rainfall episode, it is also observe (see the isolines) that omega values are negative over the continent (upward vertical motion) and positive over the ocean (downward vertical movement). It is well known that upward vertical motion over the continent can result in rainfall. This rainfall is fed by moisture transported from the ocean to the continent by easterly winds that predominated in the area in late November (see Fig. 4). According to the location of a blocking anticyclone on the Bay of Bengal Ocean (with winds that rotate in anti-clockwise on the Southern Hemisphere) determined the occurrence of easterly winds on large part of the South Region coast, resulting in a large scale moisture transport from the ocean to the continent.

B.. Decision Tree

The decision tree with the J4.8 algorithm was created with confidence factor used for pruning (0.25), and number of instances per leaf(8). Several tests were performed: with fixed number of attributes (meteorological variable for different coordinates are considered different attribute) with smallest p-values. The best result was obtained with the 5 different climatological variables, considering 10 different coordinates for each variable, with smallest p-values (total 50 attributes). To this goal, the precipitation time series were divided over the area of Chennai(red dot) in two classes: “light” (values below the median), and “strong” (values above the median), corresponding to episodes of low and high rainfall, respectively. The training set comprised data from 2000 up to 2016. The years of 1999, 2007, 2008, 2009, and 2010 were used to evaluate the tree performance. Figure 1shows two rainfall intense episodes: July 1999, and November 2008. The event at July 1999 was less intense than November 2008.

The resulting tree, displayed in Fig.5, has 7 leafs (4 “strong” and 3 “light”) and 6 decision nodes. The variable with the highest information gain is omega at 500 hPa, and at coordinates 50oW and 25oS. As expected, these coordinates are as near to the disaster zone as the limited spatial resolution of the gridded data permits. Note that all but one decision nodes are also associated with omega, at different pressure levels but always in the vicinity of the affected area. These results highlight the key role played in the episode of extreme rainfall in Chennai 2015 by the vertical transport of the moisture, brought from the ocean by sustained easterly winds. As a predictor, the tree was able to forecast 100% of the cases of extreme rainfall during the evaluation years (1995, 2007-2016), including the episode occurred in November 2015.

Retrieval Number: J99780881019/19©BEIESP
DOI: 10.35940/ijitee.J9978.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

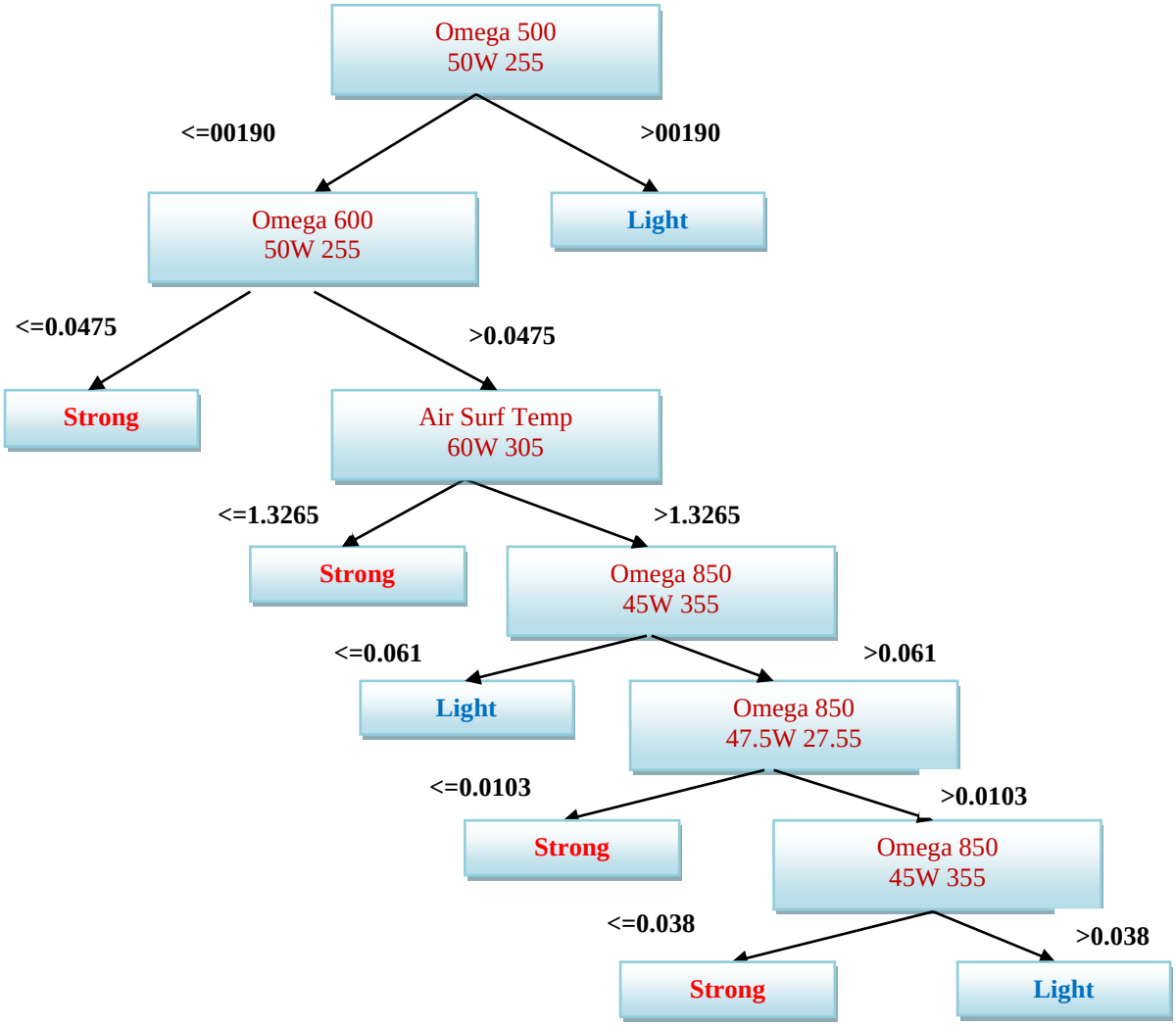


Figure 5. Decision tree generated: training set from 2000 up to 2016; test set:1999, 2007-2016

.

C.Tamilnadu Droughts

This analysis has used climatological data covering the period from January 1995 up to December 2016. Monthly anomalies were computed relative to the mean values over the period. The entire data set used in this illustrative study comprises 2.455 time series.

D. Class-Comparison

Class-classification was based on a time series of monthly accumulated precipitation anomalies, averaged over the area delimited by latitudes 4oS and 8oS and longitudes 68oW and 72oW. This time series was used as proxy of drought in our analysis. This region, located in the east costal area was strongly affected by the droughts of 2005 and 2010. In this time series, the range of anomalies was split into 3 sub-classes: “dry”, “neutral” and “wet”. To this end, the interval is divided between the highest and the lowest precipitation anomaly into three parts, assigning the upper and lower 37% bins to the “wet” and “dry” classes, respectively, and the remaining 26% to the “neutral” class. The results represents class comparison between "dry" and "neutral" classes.

IV. CONCLUSION

In this work our technique was used to generate decision trees and rules for classifying weather parameters such as maximum temperature, minimum temperature, rainfall, evaporation and wind speed in terms of the month and year. The data used was for Tamilnadu metropolis obtained from the meteorological station between 2009 and 2016. The results show how these parameters have influenced the weather observed in these months over the study period. Given enough data the observed trend over time could be studied and important deviations which show changes in climatic patterns identified. This work is important to climatic change studies because the variation in weather conditions in term of temperature, rainfall and wind speed can be studied using these data mining techniques.

REFERENCES

1. Ruivo H. M.; Sampaio G.; Ramos F. M; Knowledge extraction from large climatological data sets using a genome-wide analysis approach: application to the 2005 and 2010 Amazon droughts; Climatic Change; pages 1-15 (2014).
2. Rokach, L., Maimon, O.: Data Mining with Decision Trees: Theory and Applications. World Scientific Publishing. Vol. 270, (2007).
3. Simon, R. M. et al.: Design and analysis of DNA microarray investigations Springer Vol 209, (2003).
4. Hardin J, Mitani A, Hicks L, VanKoten B., A robust measure of correlation between two genes on a microarray. BMC Bioinformatics 8:220. (2007).
5. Witten, I. H., Frank, E. S.: Data mining: Practical machine learning tools and techniques with java implementation Morgan Kaufmann Publishers. (2000).

6. Hunt, E.B. Concept learning: An information processing problem.New York: Wiley (1962).
7. Quinlan, J. R.: C4.5: programs for machine learning. Morgan Kaufmann Publishers. (1993).
8. Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D. et al.: The NCEP/NCAR 40-Year Reanalyses Project. Bull Amer Meteor Soc Vol 77, pages 437-471. (1996).
9. Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C., Wang, W.: An improved in situ and satellite SST analysis for climate Journal Of Climate Vol 15, pages 1609-1625. (2002).
10. NOAA - Earth System Research Laboratory: Multivariate ENSO index (MEI), U.S. Department of Commerce, National Oceanic and Atmospheric Administration, http://www.cdc.noaa.gov/people/klaus.wolter/MEI/, Accessed 19 jul. 2010 (2007).
11. Huffman GJ, Bolvin DT (2011) TRMM and Other Data Precipitation Data Set Documentation. Laboratory for Atmospheres, NASA Goddard Space Flight Center and Science Systems and Applications Inc. ftp://meso.gsfc.nasa.gov/pub/trmmdocs/3B42$_$3B43$_$doc.pdf. Accessed 30 Jun 2011.
12. Lewis SL, Brando PM, Phillips OL, van der Heijden GMF, Nepstad D: The 2010 Amazon Drought. Science 331:554--554. doi: 10.1126/science.1200807. (2011).
13. Sistema Nacional de Informações sobre Recursos Hídricos (SNIRH) – Agencia Nacional de Águas (ANA) - Available in: http://ana.gov.br/portalsnirh/. Accessed March 2010.
14. Chao WC: Multiple quasi equilibria of the ITCZ and the origin of monsoon onset. Journal of the atmospheric sciences, 57(5), 641-652.(2000).
15. Cook B, Zeng N, Yoon and J-H: Climatic and ecological future of the Amazon: likelihood and causes of change. Earth Syst. Dynam. Discuss. (2010).
16. Marengo JA, Tomasella J, Alves LM, Soares WR, Rodriguez DA: The drought of 2010 in the context of historical droughts in the Amazon region. Geophys Res Lett 38 L12703, doi:10.1029/2011GL047436.(2011).
17. Marengo JA, Nobre CA, Tomasella J, Cardoso MF, Oyama MC: Hydro-climatic and ecological behaviour of the drought of Amazonia in 2005. Philos Trans R Soc B 363:1773--1778. (2008).
18. R.Senthil Kumar, Dr. C.Ramesh," Analysis of Extreme Rainfall Events in Chennai Using a Novel Data Mining Approach", Sylwan Journals, 163(7), Pp.14-29.