*International Journal of Emerging Research in Management &Technology*
*ISSN: 2278-9359 (Volume-6, Issue-6)*

**Research Article** **June 2017**

# Rainfall Prediction in South-Eastern Part of Bangladesh by Linear Regression Method

**[1]Mohammad Shohidul Islam[*], [2]Sultana Easmin Siddika, [3]S M Injamamul Haque Masum**
[1,3]School of Electronics and Information Engineering, [2] School of Economics and Management
[1, 2, 3] Beihang University, Beijing, China

***Abstract—***

***Rainfall forecasting is very challenging task for the meteorologists. Over the last few decades, several models have been utilized, attempting the successful analysing and forecasting of rainfall. Recorded climate data can play an important role in this regard. Long-time duration of recorded data can be able to provide better advancement of rainfall forecasting. This paper presents the utilization of statistical techniques, particularly linear regression method for modelling the rainfall prediction over Bangladesh. The rainfall data for a period of 11 years was obtained from Bangladesh Meteorological department (BMD), Dhaka i.e. that was surface-based rain gauge rainfall which was acquired from 08 weather stations over Bangladesh for the years of 2001-2011. The monthly and yearly rainfall was determined. In order to assess the accuracy of it some statistical parameters such as average, meridian, correlation coefficients and standard deviation were determined for all stations. The model prediction of rainfall was compared with true rainfall which was collected from rain gauge of different stations and it was found that the model rainfall prediction has given good results.***

***Keywords***— ***Bangladesh, Correlation Coefficient, Rainfall, Prediction, Linear Regression***

## I. INTRODUCTION

Bangladesh is primarily a low-lying plain land of about 144,000 square kilometres, situated on deltas of large rivers flowing from the Himalayas, has a sub-tropical humid climate characterized by wide seasonal variations [1]. The geographical location of Bangladesh is highly discussed for generating weather phenomena and disasters. Bangladesh is a deltaic land at the end of the funnel shaped Bay of Bengal (Figure 1) [2]. This special geographical configuration is dominating for the formation of cyclones, floods, droughts, tornadoes, heavy rainfall and so on. In view of this, the rainfall is being measured on instant basis using the weather radar of Bangladesh Meteorological Department (BMD) and 35 rain gauges located at different places of the country.


Fig. 1 The geographical location of Bangladesh in the World and Asia map [2]

In this study, the main purpose is to find out a forecasting trend of rainfall which will be helpful for flood forecasting, heavy rainfall estimation as well as early warning of harsh weather. In this regard rainfall estimation is important in critical moments; monsoon, pre-monsoon or post-monsoon are highly focused [2].

## II. OVERVIEW AND RELATED WORKS

### *A. Overview*

There are mainly 4 meteorological seasons in Bangladesh; they are winter, pre-monsoon, monsoon and post-monsoon. Actually, pre-monsoon and post-monsoon are the transitional period between SW-monsoon (monsoon) and NE-monsoon (winter). Average temperature of the country ranges from 17 to 20.6°C during winter and 26.9 to 31.1°C during summer. Average relative humidity for the whole year ranges from 70.5% to 78.1% in Bangladesh [3]. The humidity is over 84% during the monsoon months and about 55% in other months. The Seasons of Bangladesh: a.

 

Winter (December-February) b. Pre-monsoon (March-May) c. Monsoon (June-September) d. Post-monsoon (September-October). Heavy rainfall, tea garden, dense bamboo and cane bushes, high flood plain and the flashy rivers; all the features are very related and contributed by the hills of this region, e.g., [4]

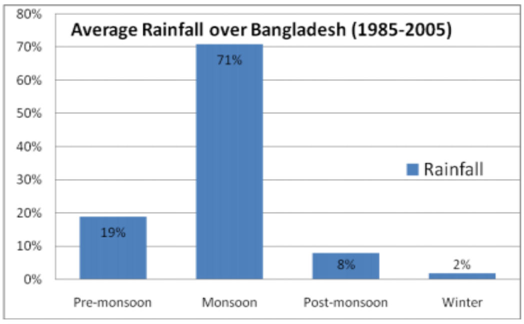


Fig. 2 The percentage of average precipitation over Bangladesh (1985-2005) [2]

### *B. Related Works*

N. Sen. has presented long-range summer monsoon rainfall forecast model based on power regression technique with the use of Ei Nino, Eurasian snow cover, north west Europe temperature, Europe pressure gradient Wind pattern, Arabian sea SST, east Asia pressure and south Indian ocean temperature in previous year [5]. The experimental results showed that the model error was 4%. S. Nkrintra, described the development of a statistical forecasting method for SMR over Thailand using multiple linear regression and local polynomial-based nonparametric approaches [6]. M. T. Mebrhatu modelled for prediction categories of rainfall (below, above, normal) in the highlands of Eritrea [7]. Human height prediction model based on multiple polynomial regressions that were used successfully to forecast the growth potentials of height with precision and was helpful in children growth study [8]. Vaccari modelled plant motion time series and Nutrient recovery data for advanced life support using multi variable polynomial regression [9].

## III. STUDY AREA & DATA

### *A. Study Area*

The study area for this research was selected the south-eastern part of Bangladesh i.e. a part of Chittagong region. The topography of Chittagong region is very interesting. The region is covered by fantastic lakes, beautiful hills, small forests, isolated islands and plain lands. Chittagong is the second largest city of Bangladesh and a busy international seaport. It is called the commercial capital of Bangladesh. With a picturesque hinterland of large hill-forests and lakes, Chittagong is a good vacation spot. Its green hills and forests; broad, sandy sea beaches; and fine, cool climate always attract holidayers. It combines the hum of a restless seaport with the pleasure of a charming hill town with its undulating topography.

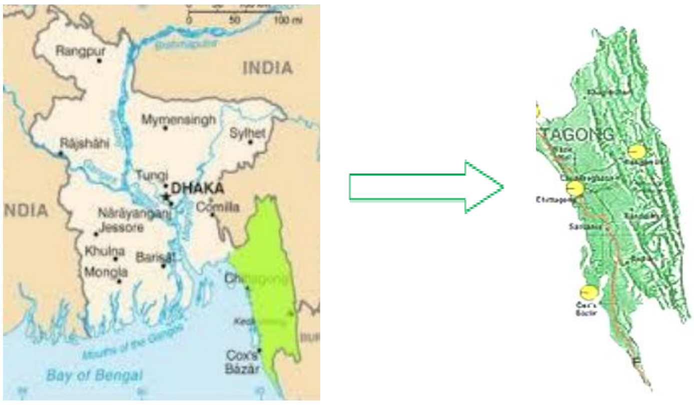


Fig. 3 Study area (a) light green colour in map, (b) Chittagong region [2]

### *B. Data*

In the current analysis, long-term recorded climate dataset was used. The recorded rainfall data was collected from Chittagong region of Bangladesh. It is actually rain gauge rainfall over the study area which is provided by Bangladesh Meteorological Department (BMD) for a certain time period of 2001- 2011. The data included observations of both 03 hourly and daily rainfall accumulations at several rain gauge sites were collected from BMD stations.

## IV. METHODOLOGY

Linear regression (LR) method is used to model the linear relationship between a predicated (dependent variable) and predictors (one or more independent variables). This model is based on least squares and it is widely used in climatology for developing models to reformation climate variables from tree-ring series. The regression model is used to originate estimates of the predicted variable outside the period used to fit the data. The uncertainty in the reformation is summarized by confidence intervals, which can be computed by various alternative ways.

 

### *A. Linear Regression Model*

The linear regression line is fitted using the most common method of least squares. This method measures the best fitting line for the observed data by minimizing the sum of the squares of the vertical deviations from each data point to the line. If a point lies exactly on the straight line then the algebraic sum of the residuals is zero [1]. Residuals are defined as the difference between an observation at a point in time and the value read from the trend line at that point in time. The equation of a linear regression line is given a

$$y = a + bx \quad \ldots\ldots(1)$$

where, y is the observation on the dependent variable(s)

x is the observation on the independent variables

'a' is the intercept of the line on the vertical axis and 'b' is the slope of the line.

The estimate of intercept 'a' and the regression coefficient 'b' by the least square method

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \quad \ldots\ldots(2)$$

$$\hat{b} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})} \quad \ldots\ldots(3)$$

and Coefficient of determination, $R^2$ = SS due to regression / total SS =

$$\frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \quad \ldots\ldots(4)$$

In order to fit regression lines of the in rainy season monthly average Rainfall (dependent variables) against time (independent variable) in years were plotted. Linear regression lines were then fitted to determine the trends of rainfall. The drawing of the diagrams and the fitting of the regression lines were done in Microsoft Excel.

### *B. Intensity Trend*

By secular trend or simply trend we mean the general tendency of the data to increase or decrease during a long period of time [1]. Temperature, rainfall and agriculture production data are made over time and therefore are referred to as time series data, which is defined as a sequence of observations that varies over time. The time series is made up of four components known as seasonal, trend, cyclical and irregular (Patterson, 1987). Trend is defined as the general movement of a series over an extended period of time or it is the long-term change in the dependent variable over a long period of time (Webber and Hawkins, 1980). Since the trend variation occurs over a substantial extended period of time, the stations 30 years of available data were considered suitable for the trend analysis. Therefore Cox'sBazar, Hatiya and Sandwip stations were excluded from this analysis. Trend is determined by the relationship between the two variables (rainfall and time). To observe that the trend of monsoon average Rainfall for the selected stations and trend values have been calculated by using least square method. The simple regression coefficient indicates that on an average the rainfall in Chittagong is increasing by 0. 314 for per year.

### *C. Correlation Coefficient*

The correlation coefficient determines the strength of linear relationship between two variables. It always takes a value between –1 and +1, with 1 or –1 indicating a perfect correlation (all points would lie along a straight line in this case and having a residual of zero). A correlation coefficient close to or equal to zero indicates no relationship between the variables. A positive correlation coefficient indicates a positive (upward) relationship and a negative correlation coefficient indicates a negative (downward) relationship between the variables. The correlation coefficients between rainfall and time were calculated as follows. Given the pairs of values $(x_1, y_1)$, $(x_2, y_2)$, ………. $(x_n, y_n)$, the formula for computing the correlation coefficient is given by

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad \ldots\ldots(5)$$

The correlation coefficients for Chittagong station was calculated using the above formula. The results are shown in Table 3 and 4.

## V. RESULTS ANALYSIS AND DISCUSSIONS

### *A. Analysis of Monthly Rainfall*

In table 1, monthly average rainfall was calculated from daily rainfall of 8 stations for 2001-2011 and trend values of rainfall was calculated for each year. Figure-4 shows that the trend of rainfall for Chittagong is increasing which indicates there is a positive linear relationship between rainfall and time. The $R^2$ value 0.0985 means that only 9.85 percent variation in rainfall is explained by time. The strength of the linear relationship between the variable and time was then calculated to determine the trend of rainfall. These relationships are measured by the correlation coefficient. According to monthly rainfall, the correlation coefficient for Chittagong region is 0.314.

 

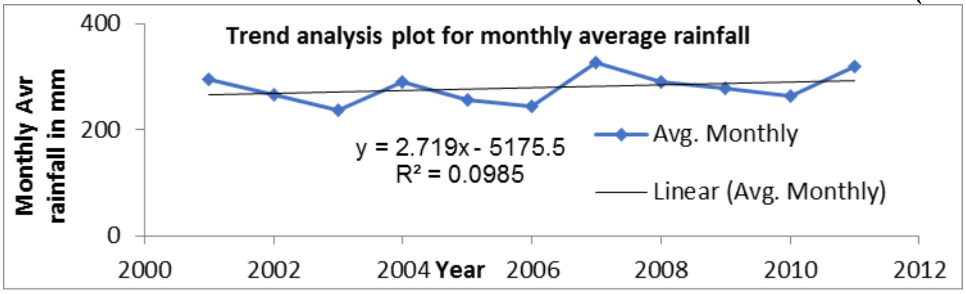


Fig. 4 Trend Analysis Plot for monthly average rainfall of Chittagong region during period last 11 years (2001-2011).

Table 1 Computation of Trend Values of Monthly Average Rainfall over Chittagong

| Year | Monthly. Avg. RF | Trend Values | Correlation Coefficient |
|---|---|---|---|
| 2001 | 295.05 | 265.32 | |
| 2002 | 266.25 | 268.04 | |
| 2003 | 237.93 | 270.76 | |
| 2004 | 290.86 | 273.48 | 0.314 |
| 2005 | 255.26 | 276.19 | |
| 2006 | 244.81 | 278.92 | |
| 2007 | 326.18 | 281.63 | |
| 2008 | 291.19 | 284.75 | |
| 2009 | 278.31 | 287.07 | |
| 2010 | 262.48 | 289.79 | |
| 2011 | 319.34 | 292.51 | |

Table 2 Some Statistical Monthly Rainfall Overall Station

| Statistical Parameters | Value |
|---|---|
| Average | 278.8782 |
| Median | 278.31 |
| Correlation Coefficient | 0.314 |
| Standard Deviation | 28.73186 |

In figure 4, it is also seen that linear line is upward direction which means the rainfall activities is increasing with time. At the same time, some statistical parameters i.e. average, median, correlation coefficient, standard variation are calculated for monthly average rainfall overall stations.

### *B. Analysis of Yearly Rainfall*

Figure 5 shows that the trend of yearly rainfall for Chittagong is also increasing which indicates that there is a positive linear relationship between rainfall and time. The $R^2$ value is also same. The linear relationship between the rainfall and time was then calculated to determine the trend of rainfall. These relationships are measured by the correlation coefficient.

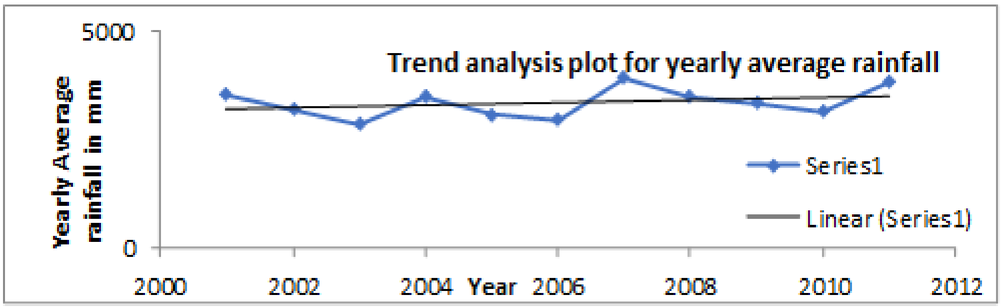


Fig. 5 Trend Analysis Plot for yearly average rainfall over 08 individual stations of Chittagong region for last 11 years (2001-2011)

Table 3 Step By Step Standard Deviation Finding

| Year | Average Rainfall | Trend Values |
|---|---|---|
| 2001 | 3540.625 | 3182.63 |
| 2002 | 3195 | 3215.25 |

| | | |
|---|---|---|
| 2003 | 2855.25 | 3247.88 |
| 2004 | 3490.375 | 3280.51 |
| 2005 | 3063.125 | 3313.14 |
| 2006 | 2937.75 | 3345.76 |
| 2007 | 3914.125 | 3378.39 |
| 2008 | 3494.375 | 3411.02 |
| 2009 | 3339.625 | 3443.65 |
| 2010 | 3149.875 | 3476.28 |
| 2011 | 3832.125 | 3508.9 |

Table 4 Some Statistical Parameters for Monthly Average Rainfall for Individual Stations

| **Statistical Parameters** | **Value** |
|---|---|
| Average | 3346.568 |
| Median | 3339.625 |
| Correlation Coefficient | 0.313882 |
| Standard Deviation | 344.7676 |

The Trend of rainfall was analyzed also for 08 stations individually. It is seen that trend of rainfall increasing mode for most of the stations. There was only exception of Hatiya and Rangamati. Over these two stations trend is negative mode which means that there is a negative relationship between rainfall and time.

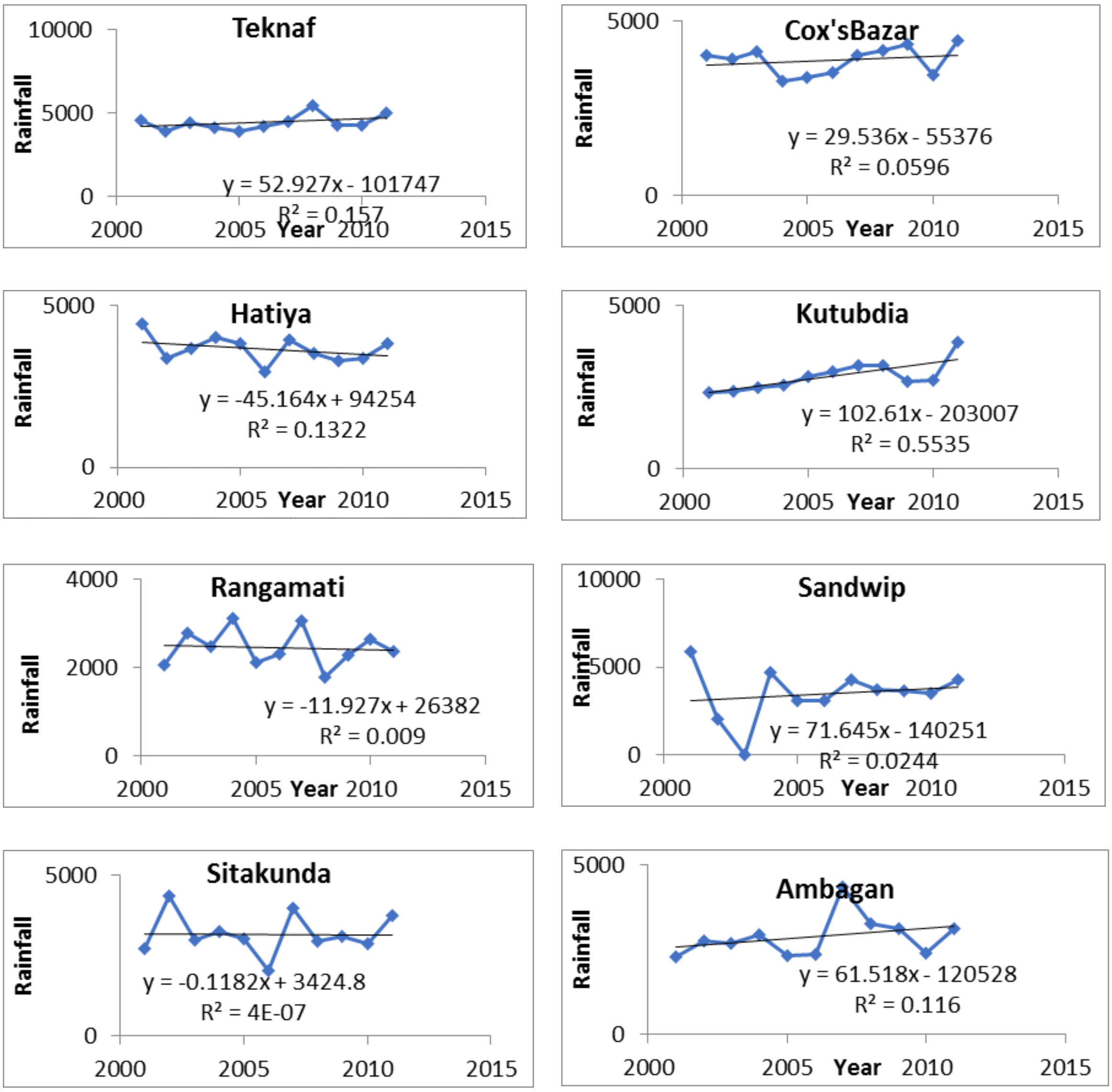


Fig. 6 Trend Analysis Plot for yearly average rainfall over 08 individual stations of Chittagong region for last 11 years (2001-2011)

 

Table 5 Step By Step Standard Deviation Finding

| **Station** | **Monthly Avg Rain (X)** | **Mean Avg Rain (M)** | **X-M** | **(X-M)** $^2$ | Standard Deviation |
|---|---|---|---|---|---|
| Teknaf | 414.83 | 319.38 | 95.45 | 9110.7 | 67.07516 |
| Cox'sBazar | 370 | | 50.62 | 2562.38 | |
| Hatiya | 320.16 | | 0.78 | 0.6084 | |
| Kutubdia | 324.5 | | 5.12 | 26.2144 | |
| Rangamati | 196.75 | | -122.63 | 15038.1 | |
| Sandwip | 355 | | 35.62 | 1268.78 | |
| Ambagan | 260.66 | | -58.72 | 3448.04 | |
| Sitakunda | 313.16 | | -6.22 | 38.6884 | |

Table 6 Some Statistical Parameters for Monthly Average Rainfall for Individual Stations

| **Statistical Parameters** | **Values** |
|---|---|
| Average | 319.3825 |
| Median | 322.33 |
| Standard Deviation | 67.07516 |

In figure 6, it is seen that linear line is upward direction for most of the stations like Cox'sBazar, Sandwip, Kutubdia, Ambagan and line is downwards for Hatiya and Rangamati., which means the rainfall activities is decreasing with time on Hatiya and Rangamati. At the same time, some statistical parameters are calculated for monthly average rainfall for individual stations.

## VI. CONCLUSION

Landslide vulnerability assessment and division may be a necessity for property management. Planning, land use and utilization should adhere to the suggested land division and relevant policy and legal provisions. Most of the landslides in Chittagong happen throughout the season once downfall intensity is extremely high. Thus, rainy seasons ought to be monitored closely to assess things, particularly within the landslide prone areas. Just in case of any potential landslide, folks of the involved localities ought to learn through early warning system. Awareness program ought to additionally contain the importance of correct land use still as property land management. A recognized vital component in adaptation to inflated floods caused by global climate change is and adequate flood prediction system ready to give reliable forecasts throughout floods with sufficient time interval. Modification of buildings and structures and their immediate surroundings goes to cut back injury in flooding. Actions have undertaken throughout floods to stop injury to and failure of control structures area unit familiar flood fighting. Flood fighting is associate degree emergency live of mitigating flood impacts on society and surroundings, significantly once control structures have tried ineffective or failing. Evacuation is important wherever the buildings or alternative options don't offer a secure place of refuge throughout a flood. Within the contest of temperature change government ought to preserve an honest volume of cash to assistant associate degreed flood relief as an accommodative live. Flood insurance is often enforced to manage the flood injury prices. It's presently on the market in several countries with well-developed insurance markets.

## ACKNOWLEDGMENT

I'm very grateful to Mr. Abdul Mannan, Scientist, Synoptic Division, SAARC Meteorological Research Center (SMRC) who led me into the world of translation. Lastly, my thanks would go to my beloved family for their loving considerations and great confidence in me all through these years. I also owe my sincere gratitude to my friends and my fellow classmates who gave me their help and time in listening to me and helping me work out my problems.

## REFERENCES

[1] Mallika Roy, "*Time Series, Factors and Impacts Analysis of Rainfall in Bangladesh*,'' International Journal of Scientific and Research Publications, ISSN 2250-3153, Volume 3, Issue 8, August 2013.

[2] Mohammad Shohidul Islam, "*Satellite Rainfall Estimation in the South-Eastern Part of Bangladesh*,'' International Journal of Scientific & Engineering Research, Volume 4, Issue 6, June-2013.

[3] Islam, S, "*National Encyclopedia of Bangladesh*,'' Asiatic Society of Bangladesh, Banglapedia, 2003.

[4] Islam, M.S.; Hasan, G.M.J.; and Chowdhury, M.A.I. "*Destroying hills in the north-eastern part of Bangladesh: A qualitative assessment of extent of the problem and its probable impact*,'' International Journal of Environmental Science and Technology, 2(4), 301-308, 2005.

 

[5] Klaush Juliseh, "*Data mining for Intrusion Detection – A critical review*", Applications of Data mining in computer security, Daniel Barbara, Sushil jajodia, Published by Springer

[6] Guhathakurta, P, "*Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network",* Current Science, 90, 773-779, 2005.

[7] D. Mark, "*Geographical information science; critical issues in an emerging cross – disciplinary Research domain*", URISA Journal, vol. 12, February 1999.

[8] Tamil Nadu Meteorology Department, Chennai.

[9] European Commission, "*Forest Fires in Europe 2007*, Technical report, Report No. 8, 2008

 