International Conference on ***Statistical Data Mining for Bioinformatics Health Agriculture and Environment,*** 21-24 December, 2012.

ISBN: **978-984-33-5876-9**

Changing Trends of Climate in Bangladesh and A Noble Procedure of Distribution of Rainfall by Clustering

Abdus Salam Azad, Md. KamrulHasan, M.ArifImtiazurRahman,

Md. MustafizurRahman and NashidShahriar

Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Email: { azadsalam2611, kamrulhasan326, arif.imtiaz216, nahidcse05, nshahriar }@gmail.com

Abstract. The present study aimed at quantifying the change in surface air temperature and the distribution of rainfall over the years in Bangladesh. The changing trend of temperature was detected using Mann-Kendall trend test and Sen's slope estimator. K-means clustering algorithm was used to identify the rainfall distribution patterns over the years and their changes with time. The analysis was performed using daily temperature and rainfall data of more than last 40 years (till 2009). The study found an increasing trend in maximum temperature during June to November and in minimum temperature during December to January in Bangladesh. There has been seen no significant change in rainfall over the years. However on the western side of the country the amount of rain is significantly less than the eastern side.

General Terms: Data mining, Statistical Analysis, Algorithms.

Keywords: Climate change, Bangladesh, Trend, Mann-Kendall trend test, Sen's slope estimator, Clustering, K-means clustering algorithm.

1 Introduction

Bangladesh is likely to be one of the countries in the world which are most vulnerable to climate change. In recent times natural hazard are more frequent and intense compared to the similar kind of events occurred in one or two decades ago. National governments and IPCC (Intergovernmental Panel on Climate Change) scientists accepted that this climate hazards are the result of climate change at the global and regional level. According to the Fourth Assessment Report(AR4) of IPCC [1], during the last hundred years the global temperature increased by 0.74 +/- 0.18 °C. The model results of AR4 for Bangladesh are appropriate for global scale. But they did not use the local data of 37 stations in Bangladesh operated by the Bangladesh Meteorological Department (BMD). Various researchers have contributed to the study of climate change[2] with long term data. In our study inspection was done on more than 40 years climatic data of all the stations (37) of Bangladesh. Then 5 important places - Dhaka, Cox's Bazar, Khulna, Sylhet and Rajshahi were analyzed with their geographic position in mind, as these stations give a clear picture of the entire country. This study undertook the challenge of finding the trends in daily temperature changes on those stations using Mann-Kendall trend test and Sen's Slope. Monthly rainfall of these selected regions was also investigated to find the distribution of rainfall throughout the year. K-means clustering algorithm was used for this purpose which gives the clear picture of rainfall distribution [4].

2 Methods Used for Analysis

2.1 K Means Clustering

Clustering is a main task of explorative data mining, which is a well known technique for statistical data analysis used in many fields, including pattern recognition, image analysis and information retrieval etc [9][10]. It assigns a set of objects into groups (clusters) so that the objects in the same cluster are more similar. For clustering K-means clustering algorithm was used which can be used for approximating multivariate distributions [3]. It is an algorithm for putting N data points ($x_1, x_2, \ldots\ldots, x_n$) in an I-dimensional space into K clusters.The mean of each cluster is denoted by $m^{(k)}$. Each vector x has I components x_i.Distances between points in real space:

$$d(x,y)=\frac{1}{2}\sum_{i}(x_i-y_i)^2$$

The K means $m^{(k)}$ is initialized to random value.Then it follows two steps:

Assignment Step: Each data point n is assigned to the nearest mean. For the cluster $k^{(n)}$ the point $x^{(n)}$belongs to

$$\hat{k}^{(n)}=\arg\min_{k} d(m^{(k)},x^{(n)})$$

The indicator variable, $r_k^{(n)}$is set to one if mean k is the closest mean to data point $x^{(n)}$; otherwise $r_k^{(n)}$ is zero.

$$r_k^{(n)}=\begin{cases}1, & if\ \hat{k}^{(n)}=k\\ 0, & if\ \hat{k}^{(n)}\neq k\end{cases}$$

Update step: The model parameters, the means, are adjusted to match the sample means of the data points that they are responsible for.

$$m^{(k)}=\begin{cases}\frac{\sum_n r_k^{(n)}\,x^{(n)}}{R^{(k)}}, & if\ R^{(k)}>0\\ Oldest\ m^{(k)}, & if\ R^{(k)}=0\end{cases}$$

where$R^{(k)}$ is the total responsibility of mean k,

$$R^{(k)}=\sum_{n}r_k^{(n)}$$

Repeat the assignment step and update step until the assignments do not change.

2.2 Mann-Kendall Trend Test

The Mann-Kendall test [5] [6] is a non-parametric test for identifying trends in time series data. The test compares the relative magnitudes of sample data rather than the data values themselves (Gilbert, 1987).Here it is assumed that that there exists only one data value per time period. Let ($x_1,x_2,\ldots\ldots,x_n$) represent n data points where x_j represents the data point at time j. Then the Mann-Kendall statistic (S) is given by

$$S=\sum_{k=1}^{n-1}\sum_{j=k+1}^{n} sign(x_j-x_k)$$

Where: $$sign(x_j-x_k)=\begin{cases}+1 & if\quad x_j-x_k>0\\ 0 & if\quad x_j-x_k=0\\ -1 & if\quad x_j-x_k<0\end{cases}$$

A very high positive value of S is an indicator of an increasing trend, and a very low negative value indicates a decreasing trend. However, it is necessary to compute the VAR (S) ,Sen's Slope associated with S and the sample size, n, to statistically quantify the significance of the trend. When n ≥ 8 the S is approximately normally distributed with the mean. The variance of S, VAR(S), by the following equation:

$$VAR(S)=\frac{n*(n-1)*(2n+5)-\sum_{i=1}^{m}t_i\,(i)(i-1)(2i+5)}{18}$$

Where t_i is considered as the number of ties up to sample i.VAR(S) and Sen's Slope estimator both are used to estimate the trend in time series data. A positive S value indicates a positive trend and a negative value indicates a negative trend in time series data.

2.3 Sen's Slope Estimator

In non-parametric statistics Sen's slope estimator [7],is a method for robust linear regression that chooses the median slope among all lines through pairs of two-dimensional sample points. The magnitude of linear trend is predicted by the Sen's estimator. The slope(Q) of all data pair is

$$Q = \frac{x_i' - x_i}{i' - i} \quad \text{fori} = 1, 2, 3, \ldots, N$$

Where:

Q= slope between data points X_i and X_i', X_i'= data measurement at time i′, X_i = data measurement at time i

i′ = time after time i

Sen's estimator of slope is simply given by the median slope (Q ′), shown below as:

$$Q' = \begin{cases} Q\left[\frac{N+1}{2}\right] & if\ N\ is\ odd \\ \frac{Q[N+1] + Q[N+2]}{2} & if\ N\ is\ even \end{cases}$$

Where: N = number of calculated slopes, Then, Q'med is computed by a two sided test at 100(1-α) % confidence interval and then a true slope can be obtained by the non-parametric test. Positive value of Sen's Slope indicates increasing trend and a negative value indicates a decreasing trend.

3 Results and Discussion

3.1 Change in Temperature

Regression analysis and then Mann-Kendall trend test was run on average maximum and minimum monthly temperature on the selected stations. The results show that the maximum temperature (TMAX) of the months June - November has increased and the minimum temperature (TMIN) of winter has increased in Dhaka, Cox's Bazaar and Sylhet. On the contrary in Khulna and Rajshahi TMIN shows negative trend.

The result of Mann-Kendall trend test and Sen's Slope Estimator on average maximum and minimum temperature are shown in Table 3 and 2 respectively. The change in temperature per decade is shown in Figure 2. In Dhaka an increasing trend in maximum temperature is observed during the months June-November (Table 3).On average the maximum temperature has increased 0.19 °C /decade. Figure 1c clearly shows the increasing trend during this time span. On the other hand the TMIN has significantly increased during November to March(Table 2). A very high increase, more than 0.55 °C /decade is observed (Fig 1c).

At Cox's Bazar positive trend is detected throughout the year. Only the month of May has not experienced any trend for TMIN. Specially from the start of the Monsoon (June-September) till the month of November TMAX showed a significant rise, more than 0.29 °C /decade (Fig (1d)). While TMIN is increased significantly in Winter (Dec-Feb) by 0.35 °C /decade. These changes are verified by Mann-Kendall statistics and Sen's slope (Table3 and 2).

Table 1: Mann-Kendall Statistics (S) &Sen's Slope (SS) of Maximum Temperature

Month	Rajshahi	Khulna	Dhaka	Sylhet	Cox's Bazaar
January	S = -227 ss = -0.02	S = -567 ss = -0.026		S = +221 ss = +0.015	S = +427 ss = +0.015
February				S = +231 ss = +0.032	S = +639 ss = +0.029
March		S = -315 ss = -0.016			S = +679 ss = +0.029
April					S= +752 ss = +0.031
May				S = +286 ss = +0.026	S = +680 ss = +0.027

June	S = +219 ss = +0.03	S = +505 ss = +0.019	S = +625 ss = +0.034	S = +525 ss = +0.036	S = +672 ss = +0.029
July	S = +398 ss = +0.03	S = +773 ss = +0.022	S = +732 ss = +0.024	S = +460 ss = +0.033	S = +724 ss = +0.28
August	S = +523 ss = +0.037	S = +653 ss = +0.022	S = +836 ss = +0.032	S = +580 ss = +0.039	S = +887 ss = +0.034
September		S = +330 ss = +0.012	S = +521 ss = +0.022	S = +431 ss = +0.031	S = +929 ss = +0.032
October		S = +435 ss = +0.014	S = +557 ss = +0.028	S = +677 ss = +0.036	S = +955 ss = +0.037
November	S = +268 ss = +0.018	S = +438 ss = +0.017	S = +548 ss = +0.03	S = +634 ss = +0.033	S = +892 ss = +0.041
December				S = +448 ss = +0.025	S = +905 ss = +0.038

* Positive value of SS and S signify positive trend and vice versa

* Empty cell denotes no trend

Table 2: Mann-Kendall Statistics (S) &Sen's Slope (SS) of Minimum Temperature

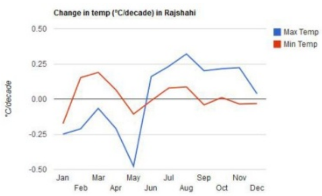
Month	Rajshahi	Khulna	Dhaka	Sylhet	Cox's Bazaar
January	S = -266 ss = -0.034	S = -355 ss = -0.04	S = +727 ss = +0.05	S = +339 ss = +0.025	S = +569 ss = +0.025
February	S = -226 ss = -0.021	S = +786 ss = +0.055	S = +393 ss = +0.025	S = +759 ss = +0.04	
March			S = +0.479 ss = +0.04	S = +358 ss = +0. 029	S = 558 ss = +0.033
April					S = +347 ss = +0.017
May		S = -239 ss = -0.015			
June			S = +319 ss = +0.011		S = +506 ss = +0.016
July	S = +169 ss = +0.007		S = +312 ss = +0.008	S = +244 ss = +0.009	S = +581 ss = +0.012
August	S = +185 ss = +0.009		S = +431 ss = +0.009	S = +372 ss = +0.012	S = +667 ss = +0.015
November			S = +726 ss = +0.048	S = +495 ss = +0.043	S = +439 ss = +0.034
December		S = -263 ss = -0.022	S = +747 ss = +0.056	S = +464 ss = +0.038	S = +559 ss = +0.033

* Positive value of SS and S signify positive trend and vice versa * Empty cell denotes no trend

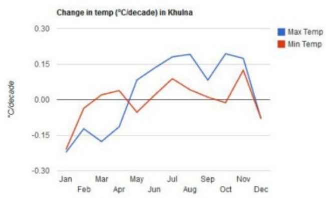
Also in Sylhet the temperature has seen a positive trend almost throughout the year. During June to November the positive trend in TMAX is about 0.29 °C /decade. In winter TMIN has increased by about 0.31 °C /decade (Fig 1e). The Mann-Kendall statistics and Sen's slope also suggests that.

In Rajshahi TMAX has shown a positive trend in June, July, August and November (Table 3). TMIN increased in July, August (Table 2). While January has shown a negative trend both in the minimum and maximum temperature. In monsoon TMAX has shown increasing trend more than 0.2 °C /decade (Fig (1a)). In winter (Dec-Feb) TMIN is found stable (Fig (1a)). Mann-Kendall trend test has shown no trend.

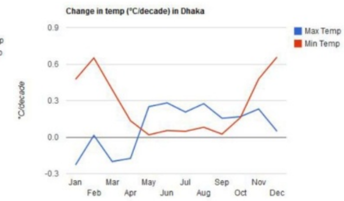
In Khulna TMAX have shown a positive trend in June-November and negative trend in January and March (Table 3). TMAX has increased about 0.16 °C /decade (Fig 1b). Only in Khulna it is found that in winter TMIN has shown negative trend. TMIN has decreased in December - February and May (Table 2). A slightly decreasing trend in TMIN about 0.11 °C /decade is found in winter (Fig 1b).



(a) Rajshahi

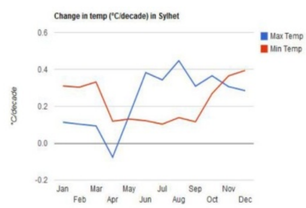


(b) Khulna



(c) Dhaka

(d) Coxbazar



(e) Sylhet

Figure 1: Change in decade(°C /decade) of minimum and maximum temperature

3.2 Clustering of Rainfall

Applying K-means clustering algorithm, monthly total rainfall were partitioned into 5 different clusters. The clusters give five categories of monthly rainfall based on amount. Cluster 1 depicts the months with negligible or no rainfall where cluster 5 represents the highest (Table 3). Thus over the years, the distribution of months in different clusters depicts a clear picture of rainfall of that region on the time period (Fig: 2).

Table 3: Rainfall in different clusters

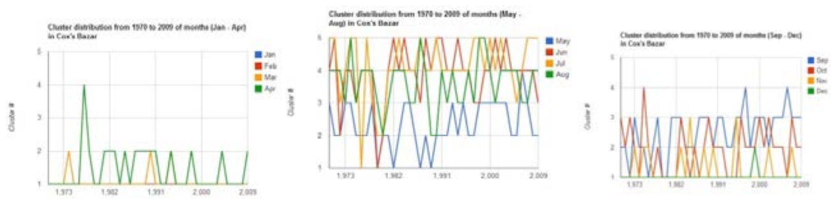
Cluster	Monthly Rainfall (mm)	Class
1	0 – 120	Negligible
2	121 – 322	Less
3	323 – 578	Moderate
4	579 – 966	High
5	967 – 3017	Very High

In Cox’s Bazar there is seen a negligible amount of rain during December to March, falling in Cluster 1. Then it increases from April, reaching its peak rainfall during June to August.In this period huge amount of rainfall is observed, mostly falling in Cluster 4 and Cluster 5. Then it starts decreasing from September (Fig: 2a,2b,2c).

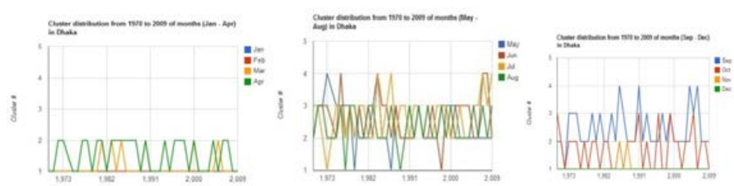
Study for Sylhet shows slightly less rainfall than Cox's Bazar, negligible rain during November to February, the an increase from March. The peak rainfall (Cluster 3,4 and 5) is seen during May to August. Then starts to decrease from September (Fig: 2m,2n,2o).

In Dhaka negligible rainfall is experienced during November to March (Cluster 1). Then it increases from April, reaches peak during May to September (Cluster 2 and 3) and then starts decreasing. (Fig: 2d,2e, 2f).

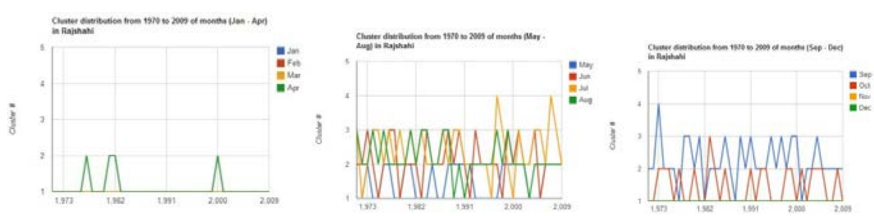
In Khulna during November to February very less rainfall is seen (Cluster 1). Then it increases from March, remains constant in its peak rainfall during May to September (Cluster 2, 3 and 4) and starts decreasing from October (Fig: 2j,2k,2l).



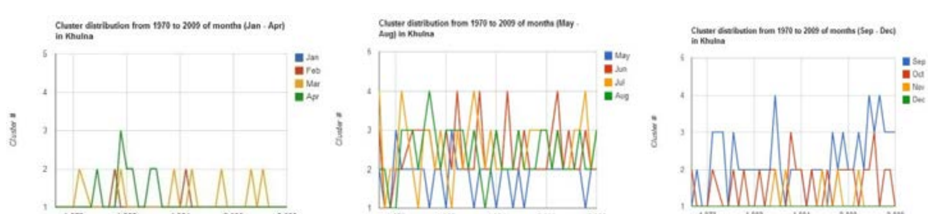
(a) Cox's Bazar(Jan-Apr) (b) Cox'sBazar(May-Aug) (c) Cox'sBazar(Sep-Dec)



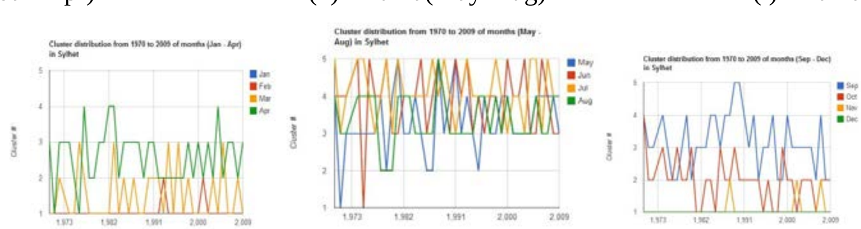
(d) Dhaka(Jan-Apr) (e) Dhaka(May-Aug) (f) Dhaka(Sep-Dec)



(g) Rajshahi(Jan-Apr) (h) Rajshahi(May-Aug) (i) Rajshahi(Sep-Dec)



(j) Khulna(Jan-Apr) (k) Khulna(May-Aug) (l) Khulna(Sep-Dec)



(m) Sylhet(Jan-Apr) (n) Sylhet(May-Aug) (o) Sylhet(Sep-Dec)

Figure 2: Change in decade(°C /decade) of minimum and maximum temperature

The study shows the least amount of rainfall in Rajshahi. Here, the rainfall constantly remains in cluster 1 from November to April spanning half of the year. The peak rainfall is seen during June to September mostly residing in Cluster 2 and 3 (Fig: 2g,2h,2i).

The study also shows that over the years the distribution of rainfall has no significant change. Except only the month of September shows some increase in Khulna from the year 1996.

It is also evident, from the distributions of clusters, in the eastern regions (Cox's Bazar and Sylhet) it rains remarkably more than the western regions (Rajshahi and Khulna).

4 Concluding Remarks

In this study Mann-Kendall trend test and Sen's slope Estimator were used to find the trends in temperature in Bangladesh. K-means clustering algorithm was used to identify the rainfall distribution patterns over the years and their changes with time. The main conclusions of the study are:

(1) Maximum temperature shows remarkable positive trend during June-November in Bangladesh. On the other hand minimum temperature has increased during December-January.

(2) The eastern side has faced more change in temperature than the western side. Cox's Bazar and Sylhet exhibits an increasing trend in almost throughout the year.

(3) The peak rainfall in the country is experienced during June-August.

(4) There has been seen no significant change in rainfall over the years.

(5) Over the years the western side of the country has experienced significantly less rainfall than the eastern side.

In future we will try to find out the correlation pattern or trends between urbanization and temperature increase. The change in temperature will also be analyzed using the empirical mode decomposition method. We have a plan to analyze the climate change behavior in different areas of Bangladesh using a color coded map. We also intend to relate rainfall distributions with floods by means of clustering.

References

[1] IPCC Fourth Assessment Report (AR4) : Climate Change 2007: Synthesis Report .

[2] Dessens J and Bucher A (1995). Changes in minimum and maximum temperatures at the Pic du Midi relation with humidity and cloudiness , 18821984. Atmospheric Research 37, 147162.

[3] a b MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations .Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability.University of California Press.pp. 281297.MR 0214227.Zbl 0214.46201.Retrieved 2009-04-07.

[4] Joo Corte-Real*,BudongQian,HongXu Regional climate change in Portugal: precipitation variability associated with large-scale atmospheric circulation, 1998.

[5] M. Kendall.Rank Correlation Methods .Charles Griffin & Company Limited, 1948.

[6] H . B. Mann. Nonparametric tests against trend. Econometrica ,13(3):245 259, 1945.

[7] Sen PK (1968). Estimates of the regression coefficient based on Kendallstau .Journal of American Statistical Association 39, 13791389.

[8] Bangladesh Meteorological Department, http://www.bmd.gov.bd/ .

[9] Everitt, B,1980: Cluster Analysis. London, UK: Heine-mann, 136 pp

[10] Mo, K.,Ghil,M.,1988: Cluster analysis of multiple planetary flow regions. J. Geophys. Res., 93, 10927-10952.