A Multiple Linear Regression Model between different climatic variables for the district of Purulia located on Lower Ganges Basin

[1]Kush Kumar Dey, Research Scholar, National Institute of Technology Durgapur, India.
kushdey242@gmail.com
[2]Dr.Vijay Kumar Dwivedi, Professor, National Institute of Technology Durgapur, India.
vkdwivedi10725@yahoo.co.in
[3]Dr. Satanand Mishra, Scientist, CSIR, Advanced Materials Processes Research Institute (AMPRI), Bhopal, India. snmishra07@gmail.com

Abstract: **Water is the prime natural element for all life in the Earth. Water stress is the most challenges matter at present day of Civil Engineer. In this paper, we established a Statistical Model on Multiple Linear Regression Analysis. This analysis is evaluating the relative impact of predictor (independent) variables such as Temperature and Cloud Cover on a particular outcome (dependent variable) as rainfall. We also workout the Multiple Correlation Coefficient 'R' to measure and interpret the strength of linear relationship between the variables. The Correlation Coefficient value 0.100 indicates the direction and strength which is weak. . In this research work, the daily rainfall, temperature and cloud cover data of different stations for the period 1901 to 2016 of Purulia district with long data series have been considered. The data are collected from National Data Centre, office of the Additional Director General of Meteorology (Research), Pune. In this analysis we focused on the direction and strength of correlation between climatic elements such as Rainfall, Temperature and Cloud Cover for the district of Purulia located on Lower Ganges Basin.**

Keywords — Multiple Linear Regression analysis, Pearson Correlation Coefficient, Rainfall, Temperature and Cloud Cover.

I. INTRODUCTION

Climate is weather averaged of 30 years intervals are typically used in climatology. Climate change is a change in the statistical distribution of weather pattern for a long-term average condition, whether due to natural variability or as a result of human activity. Due to increase of concentration of Green House Gases (GHG) particularly CO_2 in atmosphere the following factors responsible for prominent climate changes (i) the increase of temperature of atmosphere and earth surface. (ii) Change in precipitation pattern and (iii) Rise in sea level. An analysis done by NIH Roorkee of temperature data of 125 stations distributed all over India shows that the changes in temperature in India and Indian subcontinent over last century as global trend of increase in temperature.

The district Purulia located in Lower Ganges Basin and eastern slope of Chhotanagpur Plateau fall under Drought Prone Area Programme (DPAP) 2008.They have clearly mentioned, 21 blocks in Purulia district classified as water scarce area. In this district the distribution of rainfall is highly non-uniform both in terms of time and space. As a result water is required to be stored and utilized for meeting the increasing demands of that area.

Temperature drives the hydrological cycle which influence the hydrological processes in a direct or indirect way. The warmer climates lead to higher rates of evapotranspiration and increase of precipitation. The processes in association with shifting pattern of precipitation, runoff, soil moisture, humidity, surface and ground water resources etc. and may boost the frequency of drought and floods. It has been found that in India 12% and 16% of total geographical area are flood prone and drought prone area respectively.

In this study, analysis has been made the direction and strength of correlation between three major climatic elements rainfall, temperature and cloud cover for the district of Purulia located on Lower Ganges Basin. In this research work, the daily rainfall, temperature and cloud cover data of different stations for the period 1901 to 2016 of Purulia district with long data series have been considered. The data are collected from National Data Centre, office of the Additional Director General of

 DOI : 10.35291/2454-9150.2019.0128

Meteorology (Research), Pune. From this investigation we observed and summarized that the Correlation Coefficient value 'R' 0.100 indicates the direction and strength which is weakly positive.

From literature review, several researchers have analysed and studied different aspects of this area over last century but they have not given prominent signals. For this reason we have started to investigate and analyse on the district of Purulia of W.B, India. So, this analysis will give the fruitful suggestion and information about Climate Change on water resources for the district Purulia.

II. Study Area

Geographically the district Purulia is located in between $22^0 42^{/} 35^{//}$ and $23^0 42^{/} 00^{//}$ N latitude and $85^0 49^{/} 25^{//}$ and $86^0 54^{/} 37^{//}$ E longitude. The district boundary follows the Damodar in the north and Subarnarekha in the West. Purulia has a long history of water scarcity. Purulia is one of the most backward districts in West Bengal in terms of agriculture, economy and human development. The district is very poor in ground water resources because of hard crystalline basement (Halder and Saha, 2015). Ground water is available in localized areas covered by cracks and crevices in the hard rock and also from the upper weathered zone of the bed rock. The discharge from the well is generally satisfactory. A majority of shallow wells go dry or retain scanty water during the summer. There are more or less no function of ground water in agriculture development, about 1% area is irrigated from ground water (Ghosh, 2015). Purulia district has been described as the eastern slope Chhotanagpur Plateau. The major portion of the district falls under Drought Prone Area Programme (DPAP) 2008, Govt. of India.

III. Materials and Method

Data collection and data range

In this research work, the daily rainfall, temperature and cloud cover data of different stations for the period 1901 to 2016 of Purulia district with long data series have been considered. The data are collected from National Data Centre, office of the Additional Director General of Meteorology (Research), Pune. All calculation we have done by Microsoft Excel.

Methodology

The purpose of Multiple Linear Regression analysis is to determine the relative impact of predictor variables on a particular result. This is different from a correlation analysis, where the purpose is to examine the strength and direction of the relationship among random variables. Here, Multiple Linear Regression model contains two independent (explanatory) variables, X_1 and X_2 represents temperature and cloud cover respectively.

The general equation of the model expressed as $Y_i = b_0 + b_1X_1 + b_2X_2 + \ldots\ldots + b_KX_K + e_i$ where $i = 1, 2, \ldots\ldots n$, the regression parameter 'b_0' is the intercept and the regression coefficient 'b_1' and b_2 are the slope of the regression line. The random error term e_i is assumed to be uncorrelated with a mean of '0' (zero) and constant variance. Long-term mean temperature, cloud cover and mean rainfall is denoted by

$\bar{X}_{1,}$ $\bar{X}_2$ and $\bar{Y}$ respectively. The estimate of intercept 'b_0' and regression coefficient b_1 and b_2 by the least square method.

$$b_1 = \frac{(\sum X_2^2)(\sum X_1 Y) - (\sum X_1 X_2)(\sum X_2 Y)}{(\sum X_1^2)(\sum X_2^2) - \sum X_1 X_2} \quad (1)$$

$$b_2 = \frac{(\sum X_1^2)(\sum X_2 Y) - (\sum X_1 X_2)(\sum X_1 Y)}{(\sum X_1^2)(\sum X_2^2) - \sum X_1 X_2} \quad (2)$$

$$b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 \quad (3)$$

Multiple Correlation Coefficients:

Correlation Coefficient 'R'finds out the strength of linear relationship among the variables. It always takes a value -1 and +1 with 1 or -1 indicating the perfect Correlation (all points would be along a straight line). A Correlation Coefficient equal to zero indicates no relationship between the variables. A positive Correlation Coefficient indicates a positive ie. Upward relationship and a negative Correlation Coefficient indicates a negative ie. Downward relationship between the variables. Given the values of X_1 denotes the temperature (independent variable), X_2 denotes the cloud cover (independent variable), and Y denotes the rainfall (dependent or variable).

$$R = \sqrt{\frac{{r_{YX_1}}^2 + {r_{YX_2}}^2 - 2 r_{YX_1} r_{YX_2} r_{X_1X_2}}{1 - {r_{X_1X_2}}^2}} \quad \ldots\ldots\ldots\ldots\ldots(4)$$

r_{YX_1} = Correlation Coefficient of Y (Rainfall) and X_1 (Temperature),

r_{YX_2} = Correlation Coefficient of Y (Rainfall) and X_2 (Cloud Cover),

 DOI : 10.35291/2454-9150.2019.0128

ISSN : 2454-9150 Vol-05, Issue-02, May 2019

$r_{X_1X_2}$ = Correlation Coefficient of X_1 (Temperature) and X_2 (Cloud Cover),

General formula, Correlation Coefficient given by r = $\dfrac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}$ …………(5)

IV. Result and Discussion

Multiple Linear Regression model **$Y_i = b_0 + b_1X_1 + b_2X_2 + \ldots\ldots + b_KX_K + e_i$**

Table - 1

$b_1 = \dfrac{(\sum X_2^2)(\sum X_1Y)-(\sum X_1X_2)(\sum X_2Y)}{(\sum X_1^2)(\sum X_2^2)-\sum X_1X_2}$	$b_2 = \dfrac{(\sum X_1^2)(\sum X_2Y)-(\sum X_1X_2)(\sum X_1Y)}{(\sum X_1^2)(\sum X_2^2)-\sum X_1X_2}$	$b_0 = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$	**Multiple Linear Regression Model**
1.460	**0.132**	**1289.99**	**$Y=1289.99+1.460X_1+0.132X_2$**

Correlation Coefficient r = $\dfrac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}$

Table – 2

Correlation Coefficient	Correlation Coefficient Value	Individual Direction and Strength of Correlation	Multiple Correlation Coefficient Value $R = \sqrt{\dfrac{r_{YX_1}^{\;2} + r_{YX_2}^{\;2} - 2r_{YX_1}r_{YX_2}r_{X_1X_2}}{1 - r_{X_1X_2}^{\;2}}}$	Overall Strength and Direction of 'R'
r_{YX_1}	0.1108	Weakly Positive	**0.100**	**Weakly Positive**
r_{YX_2}	(-)0.1686	Weakly Negative		
$r_{X_1X_2}$	(-)3.0223	Perfectly Negative		

This analysis is evaluating the relative impact of predictor (independent) variables as temperature and cloud cover on a particular outcome (dependent variable) as rainfall. We also workout the Pearson Correlation Coefficient 'R' to measure and interpret the strength of linear relationship among the variables. The Correlation Coefficient value 0.100 indicates the direction and strength which is weakly Positive.

Data Set of climatic variables for Graphical Representation of Multiple Linear Regression Model.

Table - 3

Sl no	X_1 (Temperature in ^{0}C)	X_2 (Cloud Cover in %)	Y (Rainfall in mm)
1	15.00	5.00	1312.55
2	20.00	13.00	1320.91
3	25.00	21.00	1329.26
4	30.00	29.00	1337.62
5	35.00	37.00	1345.97
6	40.00	45.00	1354.33
7	45.00	53.00	1362.69

DOI : 10.35291/2454-9150.2019.0128

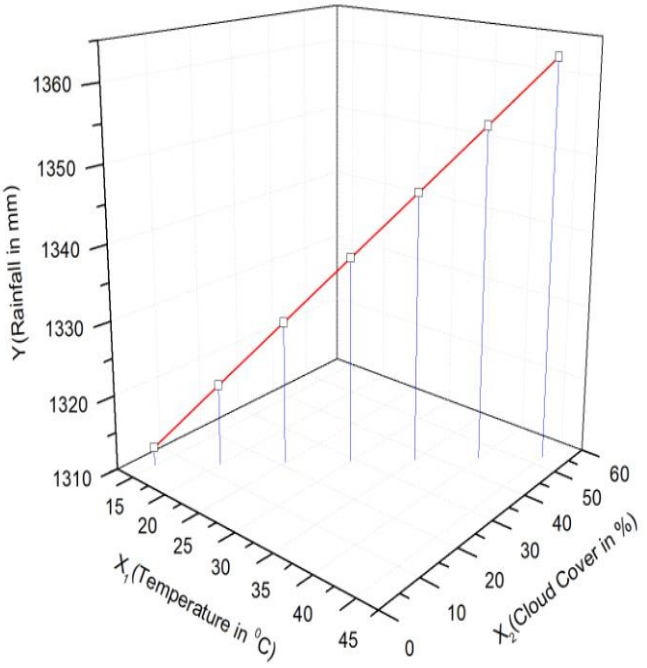

Figure:- 1 Graphical representation of Multiple Linear Regression Model.

V. CONCLUSION

i) From the equation of the Multiple Linear Regression model we concluded that temperature and rainfall are weakly positive interdependent with each other.

ii) Rainfall and cloud cover are weakly negative correlated with each other.

iii) Temperature and cloud cover are strongly negative correlated.

iv) Overall Correlation Coefficient 0.100 indicates that the rainfall is weakly positive (upward) correlated among the variables - temperature and cloud cover.

REFERENCES

[1] Prabakaran S, et al (2017), "Rainfall prediction using Modified Linear Regression", Asian Research Publishing Network Journal of Engineering and Applied Sciences, volume – 12, issue – 12, pp. 3715-3718.

[2] Rajeevan M et al (2001), "Prediction of Indian summer monsoon: Status problem and Prospects", Current Science, 81, pp.1451-1457.

[3] Kundu S, Khare D, Mondal A and Mishra P.K.(2014), "Future rainfall analytic thinking (1871-2011) for whole India", Temperature Change and Bio-diversity, Vol. 20, No. 02, pp. 45-60.

[4] Kundu S, Mishra P.K, Mondal A, Khare, D.(2015), "Analysis of abstraction and temperature change in rainfall trend of M.P in India (1901- 2011)", Environmental Earth Sciences, Vol. 73, No.12, pp.8197-8216.

[5] Meena P.K, Khare D, Kundu S, Mishra P.K.(2015), "Future analytic thinking of mega cities in North India exploitation precipitation knowledge", Indian Journal of science and technology, Vol. 08, No.03, pp. 247-253.

[6] Fowler H.J and Archer D.R.(2005), "Conflicting signals of temporal change within the Higher Indus Basin", Journal of Climate, Vol.19, No. 05, pp.4276-4293.

[7] Arora M, Goel N.K. and Sing P. (2005), "Analysis temporal trend over India", Hydrological Science Jornal, Vol. 50, No. 05, pp. 81-93.

[8] Jamaludin S and Abdul A.J.(2007), "Fitting the statistical distributions to the daily precipitation amount in Peninsular Malaysia." Journal Teknologi, vol. 46(C),pp. 33-48.

[9] Bhutiyani M.R, Kale V.S. and Power N.J.(2007), "Semi-permanent trends in most, minimum and mean annual surface air temperature across the North Western Himalayas throughout the 20 th century", Climatic Modification, Vol. 85, No. 08, pp. 159-177.

[10] Duhan D and Pandey A. (2013), "Applied mathematics analysis of future abstraction and Temperature trend of rainfall (1901-2002) in M.P, India", Atmospherically, analysis, Elsevier, Vol. 122, No. 6, pp.136-149.

[11] Guhathakurta P (2005), "Long-range monsoon prediction of 2005 for the districts and sub-division Karala with Artificial Neural Network", Current Science, 90, pp.773-779.

 DOI : 10.35291/2454-9150.2019.0128