Impact Factor (SJIF): 4.542

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444

Volume 4, Issue 5, May-2017

Correlation and Linear Regression Analysis of Water Quality Parameters of Mahi River, Gujarat

Salvi Hirva[1],Neha R. Patel[2]

[1] Department of Civil Engineering, Birla Vishvakarma Mahavidyalaya, V.V.nagar, Anand

[2] Department of Civil Engineering, Birla Vishvakarma Mahavidyalaya, V.V.nagar, Anand.

Abstract – ***Statistical analysis, such as correlation and regression analysis were applied to the data on water quality of Mahi River, generated during 2014-2015 year at seven different sites for 14 parameters. Establishing an empirical equations of regression analysis has been found to be a highly useful tool for correlating different parameters and to find monitoring the status of surface water quality. The physico-chemical parameters were measured such as pH, Total Dissolved Solids (TDS), Total Solids (TS), Total Suspended Solids (TSS), Total Hardness (TH), Calcium Hardness (Ca^{2+}), Magnesium Hardness (Mg^{2+}), Chloride (Cl^-), Total Alkalinity (TA), Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD) Temperature and Ammonia (NH_3N). Pearson's correlation coefficient (r) value is determined using correlation matrix to identify the highly correlated and interrelated water parameters. The regression equation established between these parameters to assess the extent of pollution in river. Correlation coefficient find between highly positive correlated parameters such as Total Dissolved Solids (TDS)-Total Solids (TS), Magnesium Hardness (Mg^{+2}) -Total Hardness (TH), Chemical Oxygen Demand (COD) - pH, pH-Total Alkalinity, Total Hardness (TH) -Total Alkalinity (TA), Calcium Hardness (Ca^{+2}) - Total Hardness (TH), Biochemical Oxygen Demand (BOD)- Dissolved Oxygen (DO), pH - Ammonia (NH_3N) Total Suspended Solids - pH. Based on correlation coefficient (r), regression equation establish which has R > 0.6 – 0.8 which shows parameters are dependent or independent on each other and measure standard error for validation.***

Keywords- Physico-chemical water parameters, Correlation of Coefficient, Regression equations.

I. INTRODUCTION

Water is the most important natural resource not only of a state or a country, but of the entire humanity. The prosperity of a nation depends primarily upon the judicious exploitation of this resource. Thus it establishes the importance of rivers, and no other explanation is required to stress their importance (Kannan et al 2014). River basin is as a domain for planning and management has been accepted the world over, as water does not recognize political boundaries. Among the most distinctive features of India are its rivers which hold high religious importance among its people. Covering the vast geographical area of 329 million hectares, Indian rivers have been an important reason for the rural prosperity of India. Being of wider importance in cultural, economic, geographical as well as religious development, its numerous rivers are of great value to India (Vikash et al 2014). The rivers in India are considered as Gods and Goddesses, and are even worshiped by the Hindus. They provide tourists a wonderful insight into the historical, cultural and traditional aspects of India. Among various types of inland fresh water bodies, the riverine system is a unique type of ecosystem (Harendra et al 2014). The size of the drainage basin, the amount of water moving through the system, the proportion of natural versus settled areas, and man's direct impacts are all key factors determining the quality and characteristics of each watershed (Hemant et al 2011).

The correlation coefficient is method by which we can measure the strength of a relationship among variables. If the correlation coefficient is nearer to +1 or –1, it shows the probability of linear relationship between the variables x and y . The correlation and regression analysis attempts to establish the nature of the relationship between the variables and thereby provides a mechanism for prediction or forecasting (Navneet et al., 2010). Regression equations for the parameters having significant correlation was used to estimate the concentration of other constituents (Aziz et al., 2014). Regression model were developed to test the significant correlation coefficient (Muhammad et al., 2014). Regression coefficient measures the degree of association exists between two variables.

II. MATERIALS AND METHODS

Study Area

The Mahi basin extends over states of Madhya Pradesh, Rajasthan and Gujarat having total area of 34,842 Sq.km with a maximum length and width of about 330 km and 250 km. It lies between 72°21' to 75°19' east longitudes and 21°46' to 24°30' north latitudes. It is bounded by Aravalli hills on the north and the north-west, by Malwa Plateau on the east, by the Vindhyas on the south and by the Gulf of Khambhat on the west. Mahi is one of the major interstate west flowing rivers of India. It originates from the northern slopes of Vindhyas at an altitude of 500 m near village Bhopawar, Sardarpur tehsil in Dhar district of Madhya Pradesh. The total length of Mahi is 583 km. The Som is its principal tributary which joins from right, and the Anas and the Panam joins the river from left. It drains into the Arabian Sea through the Gulf of Khambhat. The major part of basin is covered with agricultural land accounting to 63.63% of the total area and 4.34% of the basin is covered by water bodies. The basin spreads over 11 parliamentary constituencies (2009) comprising 6 of Gujarat, 3 of Rajasthan, and 2 of Madhya Pradesh.

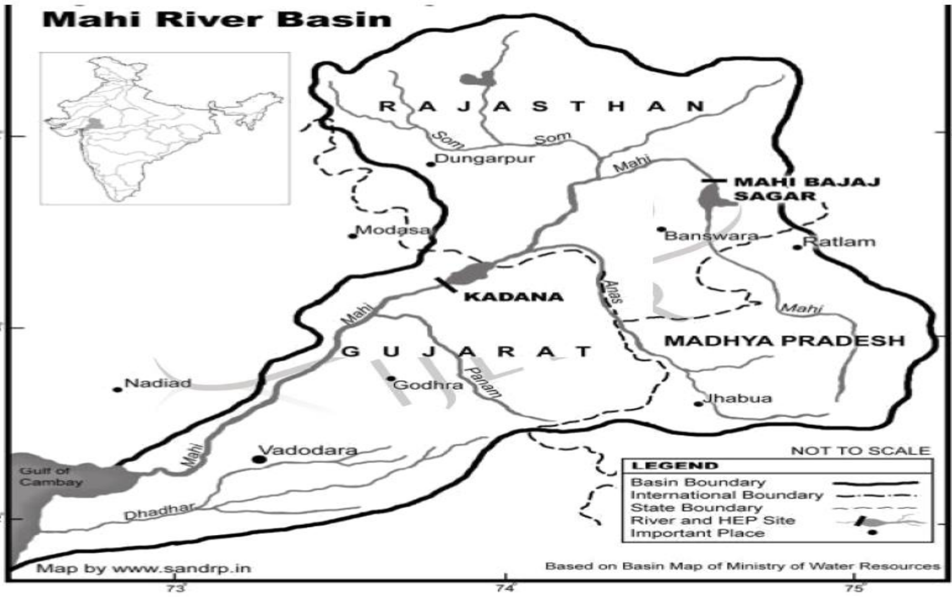

Figure 1: Study Area of Mahi River *(Source: Gor et al., 2014)*

Sampling Sites

Water samples were collected from seven different locations namely Kadana Dam (M1), Santramoura (M2), Limbodra (M3), Halol Bridge (M4), Panam Bridge (M5), TriveniSangam, Mahisagar Verama Village (M6), Sevaliya Godhra Road-1 (M7), Sevaliya Godhra Road-2(M-8), Kun River (M9), Wanakbori Dam (M-10) to study the physicochemical parameters of Mahi River. Samples were analysed as per APHA methods. Map of the sampling site is given in Fig. 1.

Figure 1: Map of Mahi Rive River Locations

Table 1: Water Quality Parameters of Mahi River at location - 1

pH	TS mg/L	TDS mg/L	TSS mg/L	NH_3N mg/L	Cl mg/L	TA mg/L	TH mg/L	Ca+2 mg/L	Mg+2 mg/L	COD mg/L	BOD mg/L	DO mg/L	Tempe-rature
8.2	270	256	4	1	40	160	160	70	90	25	3	11	24
7.63	252	226	18	0.5	40	22	130	50	80	17	1	3	33
8.60	326	306	-	0.5	24	20	124.26	74.19	50.07	-	2	7	30
8.12	288	278	8	5	22	154	118.8	69.3	49.5	10	3	8	21
8.81	178	156	12	1	25	167	110	50	60	9	-	4	33
8.47	220	194	2	2	25	156	118.8	59.4	59.4	17	3	7	27
8.95	252	228	14	3	55	114	120	70	50	19	2	8	33.3

Statistical Analysis:
The most commonly used measure of correlation is Pearson's r. It is also called the linear correlation coefficient because 'r' measures the linear association between two variables. If the data lie exactly along a straight linear with positive slope, then r = 1. Correlation coefficient (Pearson 'r') has been calculated between each pair of water quality parameters by using Excel spread sheet for the experimental data. Let X and Y are the two variables, then the correlation 'r' between the variable X and Y is given by:

$$Correlation\ (r) = \frac{n\,\Sigma(x_i y_i) - (\Sigma\, x_i y_i\,)}{\sqrt{(n\,\Sigma(x_i)2 - (\Sigma\, x_i\,)2)(\Sigma(y_i)2 - ((\Sigma\, y_i\,)2)}} \quad \text{-------------------- (2.1)}$$

Where, x and y represents two different water parameter, n = number of samples or data
If the values of correlation coefficient 'r' between two variables X and Y are fairly large, it implies that these two variables are highly correlated.

Linear Regression Analysis:
Linear Regression analysis is a statistical tool for the investigation of relationships between variables. Usually, the investigator seeks to ascertain the causal effect of one variable upon another—the effect of a price increase upon demand, for example, or the effect of changes in the money supply upon the inflation rate. To explore such issues, the investigator assembles data on the underlying variables of interest and employs regression to estimate the quantitative effect of the causal variables upon the variable that they influence. The investigator also typically assesses the "statistical significance" of the estimated relationships, that is, the degree of confidence that the true relationship is close to the estimated relationship.

Predict the value of a dependent variable based on the value of at least one independent variable. The process of predicting variable Y using variable X. Tells you how values in y change as a function of changes in values of x.

Regression tells us how to draw the straight line described by the correlation. The modified model may be written:

$$y = \beta_0 + \beta_1 x \quad \text{--------------- (2.2)}$$

Where, x = Independent variable
y = Dependent Variable
β_0 = Intercept
β_1 = Slope

The Slope will be defined:

$$\beta_1 = \frac{n\,\Sigma(x_i - y_i) - (\Sigma\, x_i\,\Sigma\, y_i)}{n(\Sigma\, x^2) - (\Sigma\, x)^2} \quad \text{-------------- (2.3)}$$

Where, n = no. of observation
X_i = Independent Variable
Y_i = Dependent Variable

The intercept will be defined:

$$\beta_0 = \frac{\sum y - \beta_1 \sum x}{n} \quad \text{-------------- (2.4)}$$

Where, y = Dependent Variable
x = Independent Variable
β_0 = Slope
n = no. of observation

Standard Error: To check the validity of the regression equation results the measured and estimated values of parameters compared by using Standard Error value.

$$\text{Standard Error \%} = \frac{\text{Y predicted} - \text{Y measured}}{\text{Y measured}} * 100 \quad \text{-------------- (2.5)}$$

III. RESULT AND DISCUSSION

The correlation coefficient shows relation between two parameters. Correlation measure the degree association between two parameters. Table 2 shows Correlation Coefficient of physicochemical Parameters of Mahi River at M-1 location.

Table 2: Correlation Coefficient of physicochemical Parameters of Mahi River at M-1 location

Parameters	pH	TS mg/L	TDS mg/L	TSS mg/L	NH_3N mg/L	Cl^- mg/L	TA mg/L
pH	1.00						
TS mg/L	-0.2	1.00					
TDS mg/L	-0.24	0.99	1.00				
TSS mg/L	-0.13	-0.07	-0.12	1.00			
NH_3N mg/L	-0.19	0.21	0.29	-0.21	1.00		
Cl^- mg/L	0.08	-0.02	-0.05	0.44	-0.36	1.00	
TA mg/L	0.31	-0.53	-0.45	-0.76	0.40	-0.13	1.00
TH mg/L	-0.42	0.35	0.37	-0.28	-0.19	0.34	-0.01
Ca^{+2} Hardness mg/L	0.26	0.78	0.79	-0.43	0.26	0.10	-0.03
Mg^{+2} Hardness mg/L	-0.59	-0.13	-0.12	-0.04	-0.35	0.28	0.01
Parameters	TH mg/L	Ca^{+2} Hardness mg/L	Mg^{+2} Hardness mg/L	COD mg/L	BOD mg/L	DO mg/L	Temperature
TH mg/L	1.00						
Ca^{+2} Hardness mg/L	0.31	1.00					
Mg^{+2} Hardness mg/L	0.80	-0.30	1.00				
COD mg/L	0.83	0.42	0.64	1.00			
BOD mg/L	-0.02	-0.06	0.01	-0.06	1.00		
DO mg/L	0.59	0.81	0.08	0.57	0.37	1.00	
Temperature	-0.36	-0.47	-0.06	-0.06	-0.33	-0.66	1.00

Location	Dependent Variable	Independent Variable	Regression Equation	r	S.E %
M-1	Mg+^{2}	TH	**Mg^{+2} = 0.80*TH - 38.155**	0.8040	10.496
	TDS	TS	**TDS = 1.060*TS - 35.644**	0.9942	5.973
	TSS	Temperature	**TSS = 0.910*Temperature - 15.032**	0.7246	5.647
M-2	Ca^{+2}	TH	**Ca^{+2} = 0.7392*TH - 25.3281**	0.8385	10.809
	TDS	TS	**TDS = 0.711*TS + 41.717**	0.8847	27.488
	NH_3N	pH	**NH_3N = 2.8044*pH - 21.118**	0.6259	1.329
M-3	TDS	TS	**TDS = 1.0924*TS - 38.5344**	0.9974	3.510
	Mg^{+2}	TH	**Mg^{+2} = 0.7329*TH - 40.9464**	0.8542	9.384
M-4	TDS	TS	**TDS = 0.889*TS + 11.155**	0.9641	13.312
	Mg^{+2}	TH	**Mg^{+2} = 0.796*TH - 47.037**	0.6411	15.392
	BOD	DO	**BOD = 0.588*DO - 1.466**	0.7602	0.806
	COD	pH	**COD = 13.86*pH - 100.63**	0.7368	5.504
M-5	TDS	TS	**TDS = 1.045*TS - 27.143**	0.9807	9.861
	pH	TA	**pH = 0.0026*TA + 7.644**	0.6711	0.228
	Mg^{+2}	TH	**Mg^{+2} = 0.404*TH + 1.499**	0.7287	11.536
M-6	TDS	TS	**TDS = 0.888*TS + 13.333**	0.9861	4.521
	Mg^{+2}	TH	**Mg^{+2} = 0.6721*TH - 27.976**	0.6456	11.062
	BOD	DO	**BOD =0.445*DO - 0.075**	0.8932	0.441
M-7	TDS	TS	**TDS = 0.7552*TS + 40**	0.9879	17.126
	Ca^{+2}	TH	**Ca^{+2} = 0.59*TH + 0.815**	0.7334	12.219
	TH	TA	**TH = 0.22*TA + 106.97**	0.7174	15.562

Table 3: Linear correlation coefficient r and Regression equation for some pairs of parameters which have significant positive value of Correlation.

Correlation coefficients of TDS - TS (r = 0.9942), Mg^{+2} - Total Hardness (r = 0.8040), Total Suspended Solids-Temperature (r = 0.7246) are obtained at location 1. Correlation coefficients of TDS - TS (r = 0.8847), Ca^{+2} - Total Hardness (r = 0.8385), NH_3N - pH (r = 0.6259) are obtained at location - 2. Correlation coefficients of TDS - TS (r = 0.9641), Mg^{+2} - Total Hardness (r = 0.6411) are obtained at location - 3. Correlation coefficients of TDS - TS (r = 0.9974), Mg^{+2} - Total Hardness (r = 0.8542), Biochemical Oxygen Demand - Dissolved Oxygen (r = 0.7602), pH - Chemical Oxygen Demand (r = 0.7368) are obtained at location 4. Correlation coefficients of TDS - TS (r = 0.9807), Mg^{+2} - Total Hardness (r = 0.7287), pH - Total Alkalinity (r = 0.6711) are obtained at location - 5. Correlation coefficients of TDS - TS (r = 0.9861), Mg^{+2} - Total Hardness (r = 0.6456), Biochemical Oxygen Demand - Dissolved Oxygen (r = 0.8932) are obtained at location - 6. Correlation coefficients of TDS - TS (r = 0.9879), Ca^{+2} - Total Hardness (r = 0.7334), Total Hardness - Total Alkalinity (r = 0.7174) are obtained at location 7.

The estimated values of TDS, TSS and Mg^{+2} compared with the observed values at M-1 location and the standard error values obtained are respectively 10.496%, 5.973% and 5.647%. The estimated values of Ca^{+2}, NH_3N, TDS compared with the observed values at M-2 location and the standard error values obtained are respectively 10.809%, 1.329% and 27.488%. The estimated values of TDS and Mg^{+2} compared with the observed values at M-3 location and the standard error values obtained are respectively 3.510% and 9.384%. The estimated values of TDS, Mg^{+2}, BOD, pH compared with the observed values at M-4 location and the standard error values obtained are respectively 13.312%, 15.392%, 0.806% and 0.292%. The estimated values of TDA, pH, Mg^{+2} compared with the observed values at M-5 location and the standard error values obtained are respectively 9.861%, 5.504% and 11.536%. The estimated values of TDS, Mg^{+2}, BOD compared with the observed values at M-6 location and the standard error values obtained respectively 4.521%, 11.062% and 0.441%. The estimated values of TDS, Ca^{+2} and TH compared with the observed values at M-7 location and the standard error values obtained are respectively 17.126%, 12.219% and 15.562%. Standard error value 0.228% at M-5 site for pH indicate a very less difference in observed and estimated value. Standard error value 27.488% at M-3 site for TDS indicate a major difference in observed and estimated value.

IV. CONCLUSION

The correlation and regression of physico - chemical parameters of river water revealed that all the parameters were more or less correlated with one another. A linear regression analysis technique has been proven to be a very useful tool for monitoring surface water and has a good accuracy. The validity of the equations are tested with the standard error results analyzed in this work and results obtained from standard errors indicate that differences in observed and estimated values. This indicates the reliability of the relationships which suggests that it can be used to predict the levels of pollution by the parameters investigated and possibly proffering a preventive measure prior to detailed investigation of the Mahi River or in pollution monitoring. Strong correlations observed between Total Dissolved Solids and Total Solids, Mg^{+2} and Total Hardness, Ca^{+2} and Total Hardness, Total Hardness and Total Alkalinity. For the proper treatment to minimize the contamination of the river water, the correlation coefficient value helps. There is also need of increasing awareness among the people to maintain the river water at their highest quality and purity level. To improve the quality of water there should be continuous monitoring of pollution level and methods should be applied for removing water pollution.

ACKNOWLEDGMENT

I thank Prof. Mrs. N.R. Patel for guiding me during the course of this study. I would also like to show our gratitude to our parents for their support and encouragement. I am also thankful to my friends Jay Patel and Pranali Shah for their support in this work.

REFERENCES

[1] Aziz-ur-Rahman and Muhammad Owais Chughtai "Reginol interpretation of river Indus water quality data using regression model", Journal of Water Resource Manage, January 2014, Vol. 8(1).

[2] Ami Gor, Arvind Shah, "Water Quality Index of Mahi River, Vadodara, Gujarat." IJEDR, 2014, Volume 2, Issue 3.

[3] Beenu Tripathi, Ruby Pandey, Divya Raghuvanshi, Harendra Singh, Vikash Pandey and D.N.Shukla, "Studies on the Physico-chemical Parameters and Correlation Coefficient of the River Ganga at Holy Place Shringverpur, Allahabad" IOSR Journal of Environmental Science, Toxicology and Food Technology, 2014.

[4] Hemant Pathak, "Interdependency between Physicochemical Water Pollution Indicators: A Case Study of River Babus, Sagar, M.P., India." Pollution Research, 2011.

[5] Kalyanaraman,S.B. & Geetha,G., "Correlation analysis and prediction of characteristics parameters and water quality index of ground water" Poll.Res., 2005, 24 (1), 197-200.

[6] Muthulakshmi L, Ramu A, Kannan.N., Murugan A., "Application of Correlation And Regression Analysis In Assessing Ground Water Quality" International Journal of ChemTech Research,2013, ISSN : 0974-4290 Vol.5, No.1, pp 353-361.

[7] Mahajan, S. V., Savita Khare & Shrivastava, V.S., "A correlation and Regression Study, Indian." J. Env. Prot., 2005, 25 (3), 254-259.

[8] Mohammad Mehdi Heydari, Ali Abasi, Seyed Mohammad Rohani and Seyed Mohammad Ali Hosseini "Correlation Study and Regression Analysis of Drinking Water Quality in Kashan City, Iran" Journal of Scientific Research, 2013.

[9] Megha Agarwal, Animesh Agarwal "Linear Regression and Correlation Analysis of Water Quality Parameters: A Case Study of River Kosi at District Rampur, India" International Journal of Innovative Research in Science, Engineering and Technology, December 2013, Vol. 2, Issue 12.

[10] Navneet Kumar, D.K. Sinha "Drinking water quality management through correlation studies among various physicochemical parameters: A case study" International Journal of Environmental Sciences 2010, Volume 1, No 2.

[11] Naseema Khatoon, Altaf Husain Khan, Masihur Rehman and Vinay Pathak, "Correlation Study for the Assessment of Water Quality and Its Parameters of Ganga River, Kanpur, Uttar Pradesh, India." IOSR Journal of Applied Chemistry, 2013.