Pak. J. Agri. Sci., Vol. 49(1), 95-100; 2012
ISSN (Print) 0552-9034, ISSN (Online) 2076-0906
http://www.pakjas.com.pk

APPLICATION OF REGRESSION MODEL ON STREAM WATER QUALITY PARAMETERS

Farhana Maqbool[1,*], Amir H. Malik[2], Zulfiqar A. Bhatti[2], Arshid Pervez[2] and M. Suleman[1]

[1]Department of Microbiology, Hazara University Mansehra, Pakistan; [2]Department of Environmental Sciences, COMSATS Institute of Information Technology Abbottabad, Pakistan
[*]Corresponding author's e.mail: fairy_es11@yahoo.com

Statistical analysis was conducted to evaluate the effect of solid waste leachate from the open solid waste dumping site of Salhad on the stream water quality. Five sites were selected along the stream. Two sites were selected prior to mixing of leachate with the surface water. One was of leachate and other two sites were affected with leachate. Samples were analyzed for pH, water temperature, electrical conductivity (EC), total dissolved solids (TDS), Biological oxygen demand (BOD), chemical oxygen demand (COD), dissolved oxygen (DO) and total bacterial load (TBL). In this study correlation coefficient *r* among different water quality parameters of various sites were calculated by using Pearson model and then average of each correlation between two parameters were also calculated, which shows TDS & EC and pH & BOD have significantly increasing *r* value, while temperature & TDS, temp & EC, DO & BL, DO & COD have decreasing *r* value. Single factor ANOVA at 5% level of significance was used which shows EC, TDS, TCL and COD were significantly differ among various sites. By the application of these two statistical approaches TDS and EC shows strongly positive correlation because the ions from the dissolved solids in water influence the ability of that water to conduct an electrical current. These two parameters significantly vary among 5 sites which are further confirmed by using linear regression.
Keywords: Single factor ANOVA, correlation coefficient *r*, leachate, TDS, linear regression

INTRODUCTION

It is unfortunate that most of our natural water bodies are gradually becoming degraded to a great extent due to rapid progress of industrialization and population explosion. There is no doubt that the physical and chemical characteristics of water can adversely affect the aquatic environment and ecological balance in many different ways. Stream water chemistry is controlled by numerous natural and anthropogenic factors (Ahearn *et al.*, 2005). Their effects on hydrochemistry can either be diffuses (e.g., runoff from urban and crop cultivation, interflow through organic rich soils) or point pollutants e.g., industrial effluents (Sliva and Williams, 2001; Li *et al.*, 2008). Also, watershed characteristics including topography and geology can influence surface water quality (Sliva and Williams, 2001). The variation of water quality parameters and stream flow is affected by many factors. These factors may be formed by the different substances in the surface water due to activities in the catchment and may be further obscured by random events (Antonopoulos and Papamichail, 1991). The objectives of the present study was to assess the magnitude of health hazards posed by the existing waters supplies, monthly monitoring of the impact to check the seasonal and physicochemical variations, and with the application of statistical model, data must be converted into easy understandable form.

The Pearson correlation values and GIS techniques were used to assess the degree of heavy metal contamination in the soils of the urban areas of vehicles (Altan *et al.*, 2011).
The hypothesis to conduct ANOVA was most of the parameters was significantly different for various sites due to different contamination level. But the results showed some parameters have non-significant variation. The analysis of variance (ANOVA) is generally used to confirm that there are not significant differences in sample population means. ANOVA determines whether there is greater variability among sample populations or within population groups (NCTGM, 2000).
Regression analyses establish a relationship between two variables that allows prediction of the dependent variable (predicted variable) for a given value of an independent variable (predictor variable). However, scientists (other than statisticians) apply regression analyses to field data when a relationship is known to exist, even when there is no true independent variable (e.g., cell counts of algae and chlorophyll concentration; nutrient concentrations and chlorophyll concentration) (Ott, 1988). According to Shyamala *et al.* (2008) interrelationship between different variables are very helpful tools in promoting research and opening new frontiers of knowledge. The study of correlation reduces the range of uncertainty associated with decision making. Carlson and Ecker (2002) statistically compared the water quality in the two lakes in each year and

examine whether or not each lake has changed, in terms of water quality variables.

MATERIALS AND METHODS

The study area is located down stream from the open dumping site of Salhad towards the Salhad stream. It lies in the Southern part of Abbottabad. Salhad stream is being already polluted by the community waste water coming from the Abbottabad Cantonment and City urban areas. After passing near the Salhad dumping site, this water is again being polluted with chemical and biological pollutants. In this study, 5 water sampling sites were selected and the co-ordinates of these sites are as follows:

1. Dabba (spring), 34° 07.567'N and 73° 11.446'E, elevation = 3822 feet.
2. Surface water (before mixing with solid waste leachate) 34° 07.567'N and 73° 11.446'E, elevation = 3822 feet.
3. Leachate, 34° 07.418'N and 73° 11.449'E, elevation = 3793 feet
4. Salhad stream (after mixing with the solid waste leachate), 34° 07.418'N and 73° 11.449'E, elevation = 3793 feet
5. Spring water of Malauchar, 34° 07.500'N and 73° 11.274'E, elevation = 3633 feet.

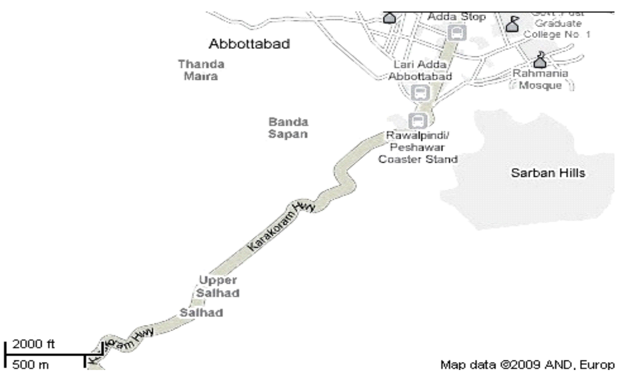

Figure 1. Study area map (source: Google earth)

Water samples from all sites were collected during rainy and dry seasons between August 2007 and April 2008 at interval of about two months. Sampling was according to the recommended procedure and accordingly all the samples for the microbiological analysis were collected in small sterilized sample bottles of 250 ml and transported in ice pack within 6 hr to the laboratory. Additionally, about 1.5 liter of water was also collected in a clean plastic can for physicochemical analysis. The bottles used were first washed with Vim powder then with hot water and finally rinsed with distilled water. All the samples were analyzed immediately after their collection for microbiological and BOD test. On reaching to the laboratory the samples were filtered through Whatman filter paper. All the samples were kept cold 4'C in refrigerator. Membrane filter technique (MF) was used for bacteriological analysis. For the total bacterial count, nutrient agar media was used (Karahan *et al.*, 2011). The physical parameters, like electrical conductivity, total dissolved solids and NaCl concentration were determined by using Conductivity/TDS meter, Microprocessor HI 9835. Dissolved oxygen was calculated by using Dissolved oxygen meter, Microprocessor Auto Cal HI 9145. Temperature and pH were calculated by using HANNA sensor checker. Biochemical oxygen demand (BOD) was determined by using BOD 5 days test. COD analyzed by using dichromate reduction method, apparatus used for the digestion of samples were Wagtech DRB200 preheated to 150^{o}C for 2 hours and measurement of dichromate reduction was done by using apparatus Lovibond PC Chekit, Photometer COD Vario (Maqbool *et al.*, 2011).

Statistical analysis: Statistical analysis was conducted by MS Excel 2007 version. In the present study Pearson coefficient correlation model was applied and *r* values were computed to find significant correlation among different parameters of various sites during August 2007 to April 2008. A correlation coefficient, usually identified as *r*, can be calculated to quantitatively express the relationship between two variables (NCTGM, analyze data). ANOVA is applied and resulted F values show the significance of our observed seasonal variation of each parameter at 5% level of significance. In the present study besides other parameters TDS & EC shows strongly positive correlating between them and this estimation is further confirmed by using regression analysis, which shows perfect value of r^2=0.961 average coefficient of determination.

RESULTS AND DISCUSSION

Pearson model of five sites were shown from Table 1-5. Site 1 & 2 shows stream water before leachate mixing. In table 1 & 2 temperature of both sites shows strongly negative correlation with TDS and EC consistent with Jothivenkatachalam *et al.* (2010), positive with BOD, and significantly positive with TBL (*r*=0.9 at site 2). The reason for positive correlation of temperature with BOD and TBL was that, as the stream temperature increases during summer the growth of mesophiles and thermophiles were higher and subsequently BOD was higher during that time. TDS shows significantly positive correlation with EC and negative correlation with DO, BOD and TBL. EC shows positive correlation with pH, negative with DO and TBL. pH shows negative correlation with COD and positive correlation with BOD. It may be due to the reason that chemical reactions in the polluted stream water occur firstly in the acidic pH, therefore COD was greater. As the bacterial growth was higher in neutral or slightly alkaline pH so the BOD was higher, DO show negative correlation with COD because of utilization of O_2 in redox reactions which increase COD and

Table1. Pearson model for Site 1 values

	Temp	TDS	EC	pH	DO	BOD	TBL	COD
Temp	1							
TDS	-0.9737	1						
EC	-0.9359	0.9867	1					
Ph	-0.4290	0.5709	0.6731	1				
DO	0.1614	-0.1796	-0.2588	-0.0589	1			
BOD	0.4123	-0.4102	-0.2906	0.1288	-0.5658	1		
TBL	0.5865	-0.5148	-0.4491	-0.3594	-0.5944	0.4346	1	
COD	0.0073	-0.1156	-0.1057	-0.4434	-0.8110	0.5330	0.5110	1

Table2. Pearson model for Site 2 values

	Temp	TDS	EC	pH	DO	BOD	TBL	COD
Temp	1							
TDS	-0.7500	1						
EC	-0.6778	0.9867	1					
pH	0.0627	0.2914	0.4269	1				
DO	-0.1088	-0.3105	-0.3827	-0.1537	1			
BOD	0.5332	-0.7143	-0.6133	0.2965	0.0410	1		
TBL	0.9167	-0.5916	-0.5225	-0.0535	-0.4603	0.4053	1	
COD	-0.1648	0.6000	0.6036	-0.0675	-0.8533	-0.5537	0.1987	1

Table3. Pearson model for Site 3 values

	Temp	TDS	EC	pH	DO	BOD	TBL	COD
Temp	1							
TDS	-0.0814	1						
EC	-0.0582	0.9990	1					
pH	-0.1851	-0.4211	-0.4358	1				
DO	-0.2577	-0.0518	-0.0876	-0.1843	1			
BOD	-0.0472	-0.3383	-0.3179	0.6296	-0.8287	1		
TBL	0.1854	0.1117	0.1358	-0.9194	-0.0001	-0.3664	1	
COD	0.0406	0.5228	0.5484	0.0643	-0.8622	0.5983	-0.0584	1

Table4. Pearson model for Site 4 values

	Temp	TDS	EC	pH	DO	BOD	TBL	COD
Temp	1							
TDS	-0.8400	1						
EC	-0.8455	0.9559	1					
pH	-0.2005	0.2604	0.5038	1				
DO	-0.1160	-0.3034	-0.2156	0.1756	1			
BOD	-0.0430	-0.2746	0.0084	0.6433	0.4624	1		
TBL	-0.7265	0.8640	0.8252	0.0614	-0.5934	-0.2192	1	
COD	-0.0150	0.5472	0.5010	0.2798	-0.7695	-0.4383	0.5125	1

Table5. Pearson model for Site 5 values

	Temp	TDS	EC	pH	DO	BOD	TBL	COD
Temp	1							
TDS	-0.4114	1						
EC	-0.4315	0.9696	1					
pH	-0.2003	0.4364	0.6207	1				
DO	0.0351	-0.4393	-0.2291	0.3012	1			
BOD	-0.0922	0.5277	0.7086	0.8744	0.4373	1		
TBL	-0.4324	0.2963	0.1542	-0.0973	-0.8121	-0.4447	1	
COD	-0.2152	0.5666	0.3494	-0.4311	-0.8936	-0.3835	0.6411	1

decrease DO.

In the Table 3 of leachate site values most of parameters shows neutral or non significant negative correlation like temperature had very weak correlations with other water quality parameters such as pH, DO and BOD, which were similar with Ouyang *et al.* (2006). Only pH shows strongly negative correlation with TBL and DO with BOD and COD (r= -0.8). Significantly positive correlation exists between TDS & EC, TDS & COD, EC & COD, pH & BOD and BOD & COD.

In the Tables 4 & 5 as the pollutant concentration decreases downward the stream due to self purification of flowing water so the correlation values vary. Temperature varies from significantly negative correlation to non-significant correlation with TDS, EC and TBL. TDS shows strongly positive correlation with EC, BOD and COD. EC shows positive correlation with pH & BOD. pH shows positive correlation with BOD.

Table 6 shows the averages values of r among various parameters of 5 different sites. It is interpreted from the above table that TDS & EC and pH & BOD have significantly increasing (+ve) r value. Positive correlation shows both parameters run parallel, if there is increase in one the other will also be increases. TDS shows the concentration of dissolved solids, it is obvious that when TDS increases EC will also increases because the charge bearing ions will increase which help to conduct electrons. When the pH of the influent sewage was above 9 the reduction of BOD was poor. If the pH of stabilization pond is above 9 the oxidation of organic matter by the heterotrophic organisms is severely inhibited (Pipes, 1962). Usually heterotrophic bacteria grow well at neutral pH range (7) so higher or lower pH can inhibit their growth and BOD will be lower at that stage.

Table 6 also shows that temperature & TDS, temp & EC, DO & BL, DO & COD have significantly decreasing (-ve) r value. Negative correlation shows that both parameters are inversely proportional to each other, if one increases the other will decreases. DO have significant negative relation with BL and COD. According to UNESCO, 2005, Natural and human-induced changes to the aquatic environment can affect the availability of dissolved oxygen. Cold water can hold more oxygen than warm water, and high levels of bacteria from sewage pollution can cause the percentage saturation to decrease.

Human activity can also affect DO levels in the Red river. Summer increases in the amount of nutrients (phosphorus, nitrogen [N] as ammonia, nitrite, and nitrate) from lawn and farm fertilizers in runoff, runoff from feedlots, storm water, and other discharges. This can result in the increased growth of plants and algae. Bacteria take up oxygen and reduce DO as they decompose this excess organic matter. Very high levels of fecal bacteria can give water a cloudy appearance, cause unpleasant odors, and increase oxygen demand (Anonymous, 2009).

Table 6. Average Coefficient of correlation r among different parameters of 5 different sites

	Temp	TDS	EC	pH	DO	BOD	BL	COD
Temp	1							
TDS	-0.6	1						
EC	-0.56	0.96	1					
pH	-0.18	0.2	0.35	1				
DO	-0.05	-0.24	-0.22	0.02	1			
BOD	0.18	0.1	-0.1	0.49	-0.09	1		
BL	0.11	0	0.03	-0.26	-0.48	-0.02	1	
COD	-0.06	0.42	0.36	-0.1	-0.8	-0.03	0.35	1

Table 7. Single Factor ANOVA (5% level):

Parameters	**F Observed value**	**F Critical value**	**Interpretation**
Temp	0.1216	3.055	Non significant
TDS	8.860	3.0555	Significant
EC	18.999	2.866	Significant
Ph	1.4069	2.866	Non significant
DO	0.249	3.0555	Non significant
BOD	2.409	2.866	Non significant
TBL	1.727	2.866	Non significant
TCL	8.395	2.866	Significant
COD	4.244	2.866	Significant

Linear regression analysis: TDS and the electrical conductivity are in a close connection. The more salts are dissolved in the water; the higher is the value of the electric conductivity, so their relationship is best defined by linear regression model. The water temperature affects the electric conductivity so that its value increases from 2 up to 3 % per 1 degree Celsius (WTS).

The electrical conductivity of water is directly related to the concentration of dissolved solids in the water. Ions from the dissolved solids in water influence the ability of that water to conduct an electrical current. When correlated with laboratory TDS measurements, electrical conductivity can provide an accurate estimate of the TDS concentration. Conductivity measurements for the Red River correlate very closely to the TDS determinations.

In addition to Pearson's correlation model application single factor ANOVA at 0.05 level of significance was applied in this study to examine whether or not stream water quality parameters were seasonally variable among five different sites (table 5). Results show that among 9 parameters 4 (TDS, EC, TCL & COD) changed significantly among various site with seasonal variation. Whereas temperature, pH, DO, BOD & TBL was not significantly changed throughout the year among the five sites.

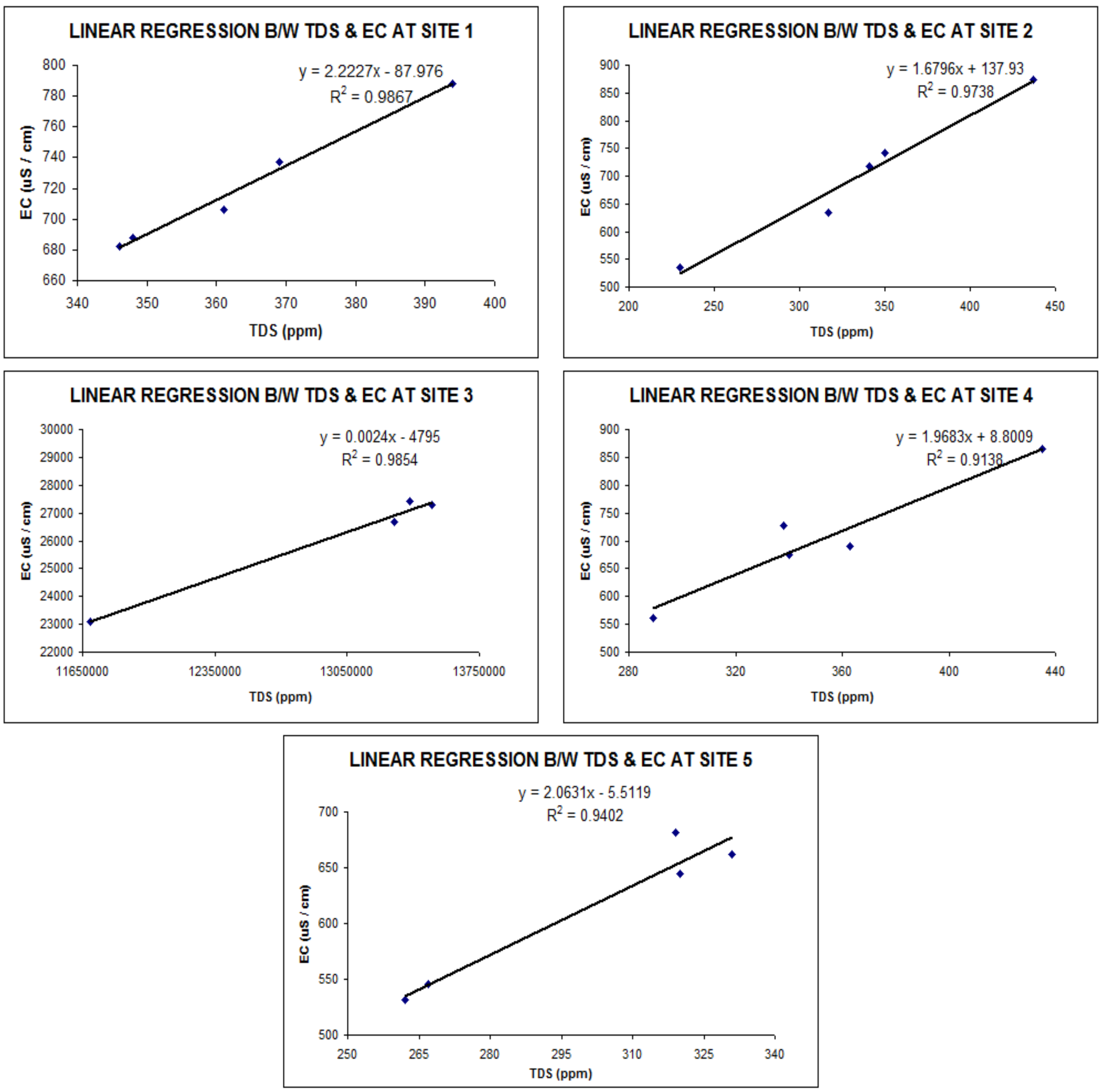

Figure 2. Linear regression between TDS & EC of different sites

CONCLUSION

It is concluded that for long water quality monitoring data statistical approaches are more convenient to understand the effect of seasonal variation on water quality and correlation coefficient model can be use to predict the other parameter (dependent) if one parameter (independent) is known, which safe time and money.

It is also concluded that from the analysis of correlation coefficient and the level of significance of various parameters will help to select the treatment strategy to minimize the contamination of stream.

REFERENCES

Ahearn, D.S., R.S. Sheibley, R.A. Dahlgren, M. Anderson, J. Jonson and K.W. Tate. 2005. Land use and land cover influence on water quality in the last free-flowing river draining the western Sierra Nevada, California. J. Hydrol. 313:234-247.

Altan, M., Ö. Ayyildiz, S. Malkoç, B. Yazici and S. Koparal. 2011. Heavy metal distribution map in soil by using GIS techniques. J. Environ. Sci. Eng. 5:15-20.

Antonopoulos, V. and D. Papamichail. 1991. Stochastic analysis of water quality parameters in streams. In: G. Tsakiris (ed.). Water resources, management and technology. Balcema 369-376.

Carlson, E. and M.D. Ecker. 2002. A Statistical examination of water quality in two Iowa lakes. Amer. J. Undergrad. Res. 1:1-16.

Jothivenkatachalam, K., A. Nithya and M.S. Chandra. 2010. Correlation analysis of drinking water quality in and around perur block of Coimbatore district, Tamilnadu, India. Rasanyan J. Chem. 3:649-654.

Karahan, M., M.N. Acik and B. Cetinkaya. 2011. Investigation of virulence genes by PCR in *Stapylococcus aureus* isolates originated from subclinical bovine mastitis in Turkey. Pak. Vet. J. 31:249-253.

Li, S., W. Liu, S. Gu, X. Cheng, Z. Xu and Q. Zhang. 2009. Spatio-temporal dynamics of nutrients in the upper Han River basin, China. J. Haza. Mater. 162:1340-1346.

Maqbool, F., Z.A. Bhatti, A.H. Malik and Q. Mahmood. 2011. Effect of landfill leachate on the water quality of the Salhad stream in Abbottabad, Pakistan. Int. J. Environ. Res. 2:491-500.

NCTGM. 2000. Nutrient Criteria Technical Guidance Manual, Rivers and Streams, APPENDIX C: Statistical tests and modeling tools, Pg A-75, United States Office of Water EPA, 2000, Washington, DC.

Ott, L. 1988. An introduction to statistical methods and data analysis. 3rd Ed. PWS Publishing Company, Boston, MA.

Ouyang, Y., P. Nkedi-Kizza, Q.T. Wu, and C.H. ShindeHuang. 2006. Assessment of seasonal variations in surface water quality. W. Res. J. 40:3800-3810.

Pipes, W.O. 1962. pH variation and BOD removal in stabilization ponds. J. Water Pollution Cont. Fed., Pub. WEF. 34:1140-1150.

Anonymous. 2009. Red river basin water quality monitoring volunteer manual. (Available online with updates at http://www.internationalwaterinstitute.org/forms/water_quality_manual_part1.pdf).

Shyamala, R., M. Shanthi and P. Lalitha. 2008. Physicochemical analysis of bore well water samples of Telungupalayam area in Coimbatore District, Tamilnadu, India. E-J. Chem. 5:924-929.

Sliva, L. and D. Williams. 2001. Buffer zone versus whole catchment approaches to studying land use impact on river water quality. Water Res. 35:3462-3472.

UNESCO. 2005. Introduction to sandwatch: an educational tool for sustainable development. Coastal region and small island papers 19, UNESCO, Paris. pp. 91.

Anonymous. 1998-2011. Water treatment solutions. (Available online with updates at http://www.lenntech.com/calculators/tds/tds-ec_engels.htm).