*Asian Journal of Chemistry* Vol. 19, No. 4 (2007), 2679-2682

# Correlation and Regression Studies of Water Quality Parameters: A Case Study of Water from the Bhavani River

Y. DOMINIC RAVICHANDRAN* and K. RAMAKRISHNAN†
*Department of Science and Humanities, Sri Krishna College of Engineering and Technology, Coimbatore-641 008, India*

In this paper, a calculation of correlation coefficient between various water quality parameters of Bhavani river is carried out. The analysis show significant linear relationship between pH, alkalinity, hardness, $Ca^{2+}$, $Mg^{2+}$, concentration of $F^-$ and total solids. Finally, multiple regression equation relating hardness with other paramters is developed. The usefulness of correlation and implication of regression equations are discussed.

**Key Words: Bhavani river, Correlation, Water quality parameters, Multiple regression.**

## INTRODUCTION

In order to keep the quality of water at an optimal level, certain water quality parameters must be monitored and controlled. Some of these parameters that most directly affect the health of the system are pH and alkalinity, hardness, temperature, dissolved oxygen and nutrients (such as nitrogen and phosphorus). This paper presents the evaluation of the available techniques for estimating missing parameter on a temporal basis. The techniques investigated include: correlation and multiple regression analysis.

The missing water quality and hydrological data from a time series or a spatial field of observations can present a serious obstacle to data analysis, modelling studies, ensuring compliance by Government agencies in terms of pollutant loadings in streams and operational forecasting purposes *etc*. Therefore, it is essential to estimate the missing observations to be able to develop reliable models of the hydrological and environmental systems based on the complete and accurate data sets. There are several techniques available for estimating the missing observations in space and time that are described in literature[1,2]. Linear interpolation for estimate of missing phosphorus concentrations in time[3] bias and variance associated with the three log-linear regression models in calculating pollutant load-

†Department of Mathematics, Sri Krishna College of Engineering and Technology, Coimbatore-641 008, India; e-mail: murali_19722002@yahoo.co.in

ings[4] regression analysis to develop a highly efficient technique to obtain the best least squares approximations of missing hydrological data[5]. The variances of estimated missing water quality observations to assess their reliability[6] were also reported. A systematic calculation of correlation coefficients between various ground water paramters and multiple regression equation relating with other parameters is reported[7]. In this paper, a calculation of correlation coefficient between various water quality parameters of Bhavani river is carried out. The analysis show significant linear relationship between pH, alkalinity, hardness, $Ca^{2+}$, $Mg^{2+}$, concentration of $F^-$ and total solids. Finally multiple regression euqation relating hardness with other parameters is developed. The usefulness of correlation and implication of regression equations are discussed.

## EXPERIMENTAL

All chemical parameters of the water samples from various parts of Bhavani river were determined using standard methods of water analysis. pH and total solids were determined using pH meter systronics MK VI and given in the Table-1.

TABLE-1
WATER QUALITY PARAMETERS OF BHAVANI WATER

| Sample No. | pH | Alkalinity | Hardness | $Ca^{2+}$ | $Mg^{2+}$ | $F^-$ | Total solids |
|---|---|---|---|---|---|---|---|
| 1 | 8.33 | 74.5 | 24.05 | 12.83 | 11.20 | 0.442 | 78.0 |
| 2 | 7.88 | 47.5 | 16.03 | 10.83 | 5.21 | 0.221 | 146.0 |
| 3 | 7.66 | 53.0 | 17.64 | 10.42 | 7.21 | 0.884 | 57.0 |
| 4 | 7.92 | 83.5 | 26.85 | 16.43 | 10.42 | 1.326 | 156.0 |
| 5 | 8.30 | 122.5 | 40.48 | 21.24 | 19.24 | 1.326 | 192.0 |
| 6 | 8.36 | 140.5 | 47.69 | 26.45 | 21.24 | 0.884 | 251.0 |
| 7 | 8.39 | 158.5 | 56.11 | 30.06 | 26.05 | 0.884 | 316.0 |
| 8 | 8.41 | 168.5 | 55.31 | 26.85 | 28.45 | 0.884 | 321.0 |
| 9 | 8.43 | 178.0 | 62.12 | 34.87 | 27.25 | 1.105 | 335.5 |
| 10 | 8.58 | 257.0 | 88.98 | 45.29 | 43.69 | 1.326 | 435.0 |

### Statistical analysis

**Deterministic model development:** The simplest approach to estimate missing data between two adjacent observations is the use of some deterministic models in the form of interpolation. The interpolation models can be either a linear or non-linear combination of the two adjacent observations.

**Regression model development:** The deterministic models do not make use of the historical data available, which may contain vital information about the relationship among various input and output variables of a system. A regression technique can be employed to develop such a relationship by first assuming the functional form of the relationship and then calibrating the model parameters through the use of the historical data. Regression model structure was investigated in the present study to estimate the missing water quality parameters data on a temporal basis. This was a multiple regression model structure. The Pearson's correlation coefficient was calculated for the pair of variables, which showed significant correlation, as shown in Table-2.

TABLE-2
CORRELATION COEFFICIENTS VALUES AMONG THE PARAMETERS FOR BHAVANI RIVER

| | pH | Alkalinity | Hardness | $Ca^{2+}$ | $Mg^{2+}$ | $F^-$ | Total solids |
|---|---|---|---|---|---|---|---|
| pH | **1.0000** | 0.8415 | 0.8348 | 0.8135 | 0.8443 | 0.2832 | 0.7859 |
| Alkalinity | 0.8415 | **1.0000** | 0.9985 | 0.9876 | 0.9962 | 0.5574 | 0.9576 |
| Hardness | 0.8348 | 0.9985 | **1.0000** | 0.9931 | 0.9937 | 0.5452 | 0.9607 |
| $Ca^{2+}$ | 0.8135 | 0.9876 | 0.9931 | **1.0000** | 0.9738 | 0.5529 | 0.9637 |
| $Mg^{2+}$ | 0.8443 | 0.9962 | 0.9937 | 0.9738 | **1.0000** | 0.5309 | 0.9457 |
| $F^-$ | 0.2832 | 0.5574 | 0.5452 | 0.5529 | 0.5309 | **1.0000** | 0.4551 |
| Total solids | 0.7859 | 0.9576 | 0.9607 | 0.9637 | 0.9457 | 0.4551 | **1.0000** |

A linear relationship was worked out by the method of least squares. Inter parameter correlation are obtained by computing the correlated matrices for the 2 blocks. The parameters considered are pH, alkalinity, hardness, $Ca^{2+}$, $Mg^{2+}$, concentration of $F^-$ and total solids. Further multiple regression analysis was done using the following relationship and the results are given in the Table-3.

$$Y = C + a_1X_1 + a_2X_2 + a_3X_3 + ...... + a_nX_n$$

## RESULTS AND DISCUSSION

Various water quality parameters of the water samples collected from different parts of Bhavani river were checked for correlation and was found to be highly correlated. The greater the numerical value of correlation coefficient, the greater is the linear relationship between the two variables. Thus, the relationship between variables with the correlation coefficient more than 0.54 is utilized to develop multiple regression equations.

Total hardness is showing high positive correlation with pH, alkalinity, $Ca^{2+}$, $Mg^{2+}$, total solids and moderate correlation with concentration of $F^-$. Thus, it can be inferred that total hardness can be used to estimate pH, alkalinity, $Ca^{2+}$, $Mg^{2+}$, concentration of $F^-$ and total solids. The multiple regression studies between total hardness and strongly correlated pH, alkalinity, $Ca^{2+}$, $Mg^{2+}$, concentration of $F^-$ and total solids, showed that these contents can be estimated from total hardness values. The maximum error in predication is ± 9.5 %. The corresponding equations are shown in the Table-3.

TABLE-3

| Relation | Maximum error (%) |
|---|---|
| TA = 4.9561 + 2.8349 TH | 6.1059 |
| $Ca^{2+}$ = 786.7645 + 449.4885TA - 158.38 TH | 9.4399 |
| $Mg^{2+}$ = -150.1225 - 84.7643 $Ca^{2+}$ + 30.2187 TA - 0.8063 TH | 0.0247 |
| pH = -9.6313*1.0e + 003 - 5.4948*1.0e + 003 $Mg^{2+}$ + 1.9404*1.0e + 003 $Ca^{2+}$ -0.0121*1.0e + 003 TH | 2.8037 |

## Conclusion

The following conclusion are drawn from the results of multiple regression analysis: i) the positive sign of the coefficient of TH indicates that it has direct impact on total alkalinity, ii) the negative sign of the coefficient of TH indicates that it has indirect impact on $Ca^{2+}$, $Mg^{2+}$ and pH, iii) the positive sign of the coefficient of TA indicates that it has direct impact on $Ca^{2+}$ and $Mg^{2+}$, iv) the negative sign of the coefficient of $Ca^{2+}$ indicates that it has indirect impact on $Mg^{2+}$ and v) the positive sign of the coefficient of $Ca^{2+}$ indicates that it has direct impact on pH and negative sign of the coefficient of $Mg^{2+}$ indicates that it has indirect impact on pH.

## ACKNOWLEDGEMENTS

The authors are thankful to the Management and the Principal of Sri Krishna College of Engineering and Technology, for their constant support and encouragement. One of the authors, (KR), wish to thank Dr. R. Balasubramanian, Head, Department of Mathematics, SKCET for his encouragement.

## REFERENCES

1. J.D. Creutin and C. Obled, *Water Resour. Res.*, **18**, 413 (1982).
2. N.S.N. Lam, *Am. Cartographer*, **10**, 129 (1983).
3. W.A. Scheider, J.J. Moss and P.J. Dillon, Measurements and Uses of Hydraulic and Nutrients Budgets, Lake Restoration Proc. National Conference, U.S. Environmental Protection Agency, Washington, DC (1978).
4. T.A. Cohn, L.L. DeLong, E. Gilroy, R.M. Hirsch and O.K. Wells, *Water Resour. Res.*, **25**, 937 (1989).
5. S. Bennis, F. Berrada and N. Kang, *J. Hydrol.*, **191**, 87 (1997).
6. G. Shih, X. Wang, H.J. Grimshaw and J. Van Arman, *J. Environ. Eng.*, **124**, 1114 (1998).
7. G.L. Verma, H.S. Bhatia and D. Katyal, *Indian J. Environ. Protect.*, **23**, 601 (2003).

(*Received*: 13 February 2006; *Accepted*: 27 December 2006) AJC-5306