Phosphorous Modelling of Renuka Lake Using Statistical Technique

M. K. Sharma, Omkar Singh and V. K. Choubey

National Institute of Hydrology, Roorkee-246 67, Uttarakhand
e-mail: mks@nih.ernet.in

ABSTRACT

The Renuka lake is a very important tourist & religious place of the HP. The National Wetland Management Committee of the Ministry of Environment & Forests (Govt. of India) has given recognition of the Renuka lake as one of important wetlands of the country, which requires conservation and management on priority basis (State Council for Science, Technology & Environment, Govt. of HP). The Major Problems in the Renuka lake are reported as soil erosion, siltation of the lake, eutrophication and enrichment of pollutants from catchment. Phosphorous is the limiting nutrient responsible for eutrophication of the lake. In the present study, the statistical technique based on best subset procedure (R^2 and F values) is used in model dissemination for predicting the phosphorous concentration. Regression models are developed in explaining variation in phosphorous concentration using routinely measured parameters. The predicted values of phosphorous are compared with observed values and reasonably good matching is obtained. The result indicates that the developed model may be used for prediction of missing observed value of phosphorous concentration and thereby the status of the eutrophication of the lake.

INTRODUCTION

Lakes are the important water resources and are being used for various purposes viz; drinking, irrigation and recreation. One of the most severe problems faced, is the eutrophication. Eutrophication refers to the excessive rate of addition of nutrients, usually in reference to anthropogenic activities and the addition of phosphorous and nitrogen to natural waters. Nutrient additions result in the excessive growth of plants including phytoplaknton (free-floating algae), periphyton (attached or benthic algae), and macrophytes (rooted, vascular aquatic plants). The process has some undesirable effects on water quality viz; excessive plant growth (green color, decreased transparency, excessive weeds), hypolimnetic loss of dissolved oxygen (anoxic conditions), loss of species diversity (loss of fishery), taste and odor problems.

Eutrophication is a natural ageing process in which the water content becomes organically enriched, leading to the domination of undesirable aquatic growth, such as algae, water hyacinth and so on. The eutrophic process tends to decrease aquatic life and has detrimental dissolved oxygen effects. The eutrophication is caused mainly due to the excess amount of the phosphorous input to the lake and changes in depth of a lake. Thus the concentration of phosphorous is most important controlling parameter of the eutrophication of the lake.

Both modeling and monitoring are useful in attempting to ascertain a lake's response to phosphorous loading from development. Monitoring provide present conditions but is expensive, time consuming as lakes exhibit high variability in terms of algal response and nutrient concentrations on both an annual and seasonal basis. Therefore there is a need to provide a cost-effective alternative which can help to protect the water quality of lakes. Modelling provides such an alternative to interpretate data with prediction for the future.

However, vast amount of data requirement (rainfall, soil type, land use etc.) places several constraints on the application of these models in Indian conditions. The statistical approach provides a viable alternative over other conventional models and gives an account of the advantage to predict the change in future based on past information. An additional advantage of this approach is the substantial reduction in the cost involved in the analysis of large number of water quality parameters.

A number of regression equations have been developed by different workers for estimation of concentration of different water quality constituents using routinely-monitored water quality parameters (Kumar et al., 1994; Rao and Rao, 1994; Krishna et al., 1995; Singanan et al., 1995; Bhatia et al., 1997; Jain and Sharma, 1997, 2000, 2002; Jain et al., 1998, 1999; Balasankar and Nagarajan, 2000). The best subset modelling procedure enables comparison between full model (containing all the independent variables) to subset models (containing subset of independent variables). Best subset procedure based on R^2 and F values can be used in model dissemination (Sharma and Jain, 2005).

Sharma and Jain (2005) developed statistical models to estimate fluoride concentration in arid region of Rajasthan, district Jodhpur. Best subset procedure based on R^2 and F values was used in model formulation. It was found that two separate models are required for shallow and deep aquifer to predict the concentration of fluoride.

In our earlier paper, a statistical model was developed for prediction of phosphorous concentration in the Upper Bhopal Lake and the predicted model was successful in estimation of the phosphorous concentration (Sharma et al., 2008). In the present investigation, an attempt has been made to develop statistical model to predict phosphorous concentration using routinely monitored water quality parameters for the Renuka Lake (Himachal Pradesh).

STUDY AREA

Renuka lake is one of the natural wetlands located in the Sirmour district of the Himachal Pradesh. It falls under Latitude 30^0 36' N and Longitude 77^0 27' E at an elevation of 645 m above mean sea level. The surface area of the lake is 75 ha, shaped like the profile of a reclining woman, which is regarded as the embodiment of the goddess Renuka (Mother of Lord Parshurama). The lake rests in a long valley and the surrounding slopes are covered with a variety of vegetation and thick woods. The supply of the lake is through nallahs draining the catchement and probably numerous underlying springs. The catchment has an area of 254.3 hectares of which 0.5% urban, 9.5% agriculture and 90% sub-tropical deciduous forest (Singh et al., 1987). The annual rainfall of the study area is 1500 mm and mean maximum temperature reaches upto 38^{0}C. The bathymetric map of the Renuka lake was obtained from Himachal Pradesh State Council for Science & Technology, Shimla (Fig. 1).

The Renuka lake is a very important tourist & religious place of the HP. Being situated in the main range of lesser Himalayas, this wetland is of special importance from bio-diversity point of view. The National Wetland Management Committee of the Ministry of Environment & Forests (Govt. of India) has given recognition of the Renuka lake as one of important wetlands of the country, which requires conservation and management on priority basis (State Council for Science, Technology & Environment, Govt. of HP). This lake also finds placed in the priority list lakes, which required immediate attention for restoration under the title "Management of Lakes in India" (Reddy and Char, 2004). Recently, the Renuka lake has also been declared as Ramsar site in the HP.

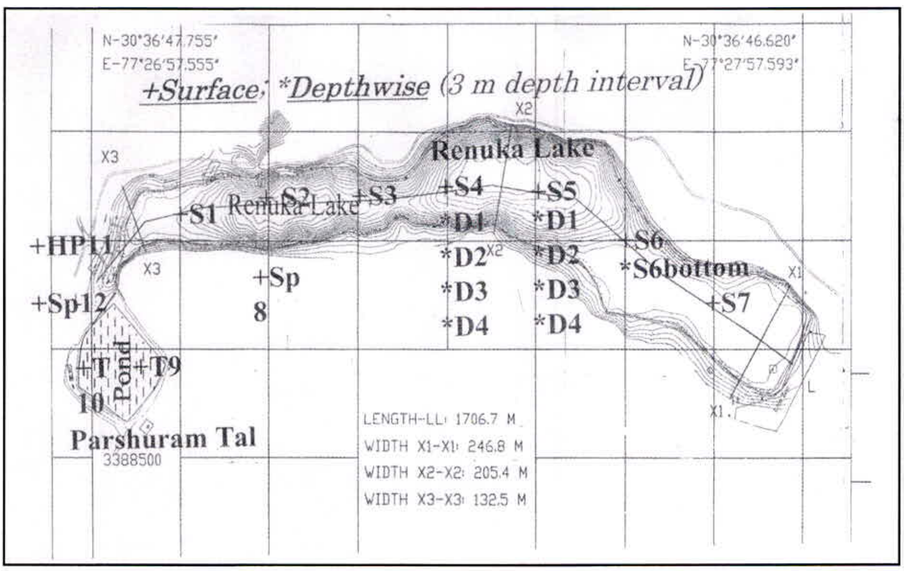

Fig. 1. Map showing the study area

WATER QUALITY SAMPLING

For the evaluation of water quality of the lake, water samples of lake were collected from various locations in the month of October 2006, January, 2007 and 2008 by dip (or grab) sampling method (Fig. 1). The water samples were collected in polyethylene bottles. All the samples were collected at a depth of 15 cm to avoid introduction of floating particles, using standard water sampler (Hydro Bios, Germany). Some parameters like pH, air temperature, water temperature and electrical conductance were measured on the spot by means of portable meters (HACH, USA). For other parameters, samples were preserved by adding an appropriate reagent and brought to the laboratory in sampling kits maintained at 4°C for chemical analysis. All chemicals used in the study were obtained from Merck, India/Germany and were of analytical grade. Deionised water was used throughout the study. Physico-chemical analysis was conducted following standard methods (APHA, 1992; Jain and Bhatia, 1987).

DATA USED

The physico chemical analysis data of water samples collected in the months of October 2006 and January 2007 and 2008 from Renuka Lake was used for the statistical analysis. The Best subset analysis was carried out for full data sets. Systat 11 was used for carrying out the analysis.

FORMULATION OF MODEL

The general representation of statistical models is given by

$$Yi = \sum_{j=0}^{k} \beta_j x_{ij} + e \qquad (1)$$

with $x_{i0} = 1$

where x_{ij} is the independent variable for the i^th^ observation (various water quality constituents),

Y_i is the dependent variable for the i^th^ observation, b are the unknown coefficients to be estimated, k+1 are the number of coefficients (to be estimated) in the model and e is the error in the determination of Y_i which is generally assumed as having zero mean and constant standard deviation (s).

The method of ordinary least squares is the most widely used method for assigning because of its simple concept and no assumption is necessary on the probability distribution of data. This method will be used for estimating the coefficients associated with various water quality constituents used for prediction of phosphorous concentration.

Before carrying out a statistical regression of the data set, preliminary analysis of data was done. This consists of initial filtration of data, partial visual inspection of the data files and the creation of the scatter plots. Once the identified input errors were removed, a general regression analysis assuming all water quality parameters as independent variables and phosphate as dependent variable was made to identify any outliers on the basis of leverage value and studentized residual stastistics. Using the filtered data, correlation matrices predicting the correlation of each water quality constituents with phosphorous were obtained. Table 1 represents the correlation coefficient, square of which indicates the contribution of individual water quality parameters in explaining the variation in the dependent variable.

Table 1 : Correlation coefficient between water quality parameters of Renuka Lake

	pH	EC	DO	BOD	Alk	Hard	Cl	SO_4	PO_4	NO_3	Na	K	Ca	Mg
pH	1.000													
EC	-0.439	1.000												
DO	-0.343	0.061	1.000											
BOD	-0.392	0.108	-0.004	1.000										
Alk	-0.012	0.020	-0.523	0.125	1.000									
Hard	0.195	0.513	-0.468	0.002	0.574	1.000								
Cl	0.013	-0.228	-0.571	0.083	0.644	0.297	1.000							
SO_4	0.038	0.577	0.252	-0.124	-0.494	0.100	-0.582	1.000						
PO_4	-0.502	0.286	-0.240	0.385	0.606	0.347	0.545	-0.241	1.000					
NO_3	-0.521	0.247	-0.214	0.382	0.465	0.088	0.395	-0.140	0.730	1.000				
Na	-0.424	0.438	0.167	0.033	-0.103	0.063	-0.077	0.293	0.427	0.449	1.000			
K	-0.066	0.231	0.484	-0.078	-0.751	-0.338	-0.529	0.676	-0.295	-0.125	0.385	1.000		
Ca	-0.230	0.588	-0.247	0.099	0.427	0.699	0.345	0.144	0.724	0.335	0.428	-0.074	1.000	
Mg	0.235	0.049	-0.456	0.116	0.579	0.653	0.368	-0.199	0.224	0.078	-0.181	-0.529	0.284	1.000

The model is useful for predictive purposes if it includes as many independent variables as possible, so that reliable fitted values can be determined. One obviously wants R^2 to be large, since R^2 gives the proportion of variation in the dependent variables, that is explained by the fitted regression model. On the other hand, because of the effort involved in the monitoring of a large number of independent variables, there is interest in including as few independent variables as possible. One has to make compromise between these extremes which is usually called selecting the best regression variables and consequently the best model. There is no unique statistical procedure for doing this (Draper and Smith, 1981). However, many workers has suggested different statistical approaches, such as all possible regression, backward elimination, forward elimination in stepwise regression, ridge regression, principal component regression and stagewise regression which may help in optimum model formulation (Draper and Smith, 1981; Montgomery and Peck, 1982; Weisberg, 1980). In the present case, the best subset regression approach has been used to select the best set of independent variables.

Best subset regression

Different best subsets of independent variables can be selected using the proportion of variation explained in the dependent variable (that is R^2 information). Assessment of each subset was made on the priority of the value of R^2 achieved, F value and the number of observations used in developing the model. The model obtained from the large data set and achieving higher values of R^2 and F value will always be preferred. The above two criteria (R^2 and F values) which will used in model selection are briefly described by Weisberg (1980).

RESULTS AND DISCUSSION

Selection of independent variables

It can be seen from the Table 1 that nitrate is the best single variable explaining 53% variation in the concentration of phosphorous. The other water quality constituents namely Ca, Alkalinity (Alk), Cl, pH, Na, BOD, Hardness (Hard), K, Electrical Conductivity (EC), SO_4, DO and Mg if taken alone as independent variable explain approximately 52%, 37%, 30%, 25%, 18%, 15%, 12%, 9%, 8%, 6%, 6% and 5% variation in the concentration of phosphorous respectively (Table 2). In order to increase the R^2, various combinations of water quality parameters with NO_3 are attempted. It is found that the combination consisting of NO_3 and Ca having larger R^2 and F values, is the best model among the other combinations. Further to increase the R^2 value, other water quality parameters were included in the combination of NO_3 and Ca. NO_3+Ca+SO_4 is selected as the best 3-parameter model. Similarly the 4-parameter model, 5-parameter model,

Table 2 : R^2 of water quality parameters with Phosphate

Parameter	R^2
pH	0.252
EC	0.082
DO	0.058
BOD	0.148
Alk	0.367
Hard	0.120
Cl	0.297
SO_4	0.058
NO_3	0.533
Na	0.182
K	0.087
Ca	0.524
Mg	0.05

6-parameter models, 7-parameter models and 8-parameter models were selected which are (NO_3+Ca+SO_4+Hard), (NO_3+Ca+SO_4+Hard+BOD), (NO_3+Ca+SO_4+Hard+BOD+Alk), (NO_3+Ca+SO_4+Hard+BOD+Alk+Na), (NO_3+Ca+SO_4+Hard+BOD+Alk+Na+K) respectively. Various selected models are given in Table 3. Now to choose the best model among the various models, one has to keep both things in mind that the selected model should have minimum number of explaining variables and maximum R^2 value. There is a compromise between the two criteria. The decision can be made on the basis of the value of F statistics as explained by Weisberg (1980). The selection procedure is explained in Table 4. It is found that the set of variable (NO_3+Ca+SO_4) is the best subset of water quality parameters that can be used for explaining the variability in the phosphorous concentration of Renuka Lake.

Model performance

The models for concentration of phosphorous is given by

Phosphorous = 0.059 + 0.028 NO_3 + 0.002 Ca – 0.001 SO_4(R^2 = 85%, F = 83.745) (2)

The first parameter is of great significance and explains the variation of phosphorous concentrations in the study area by about 53%. Just by addition of calcium, proportion of explained variation becomes 79% suggesting calcium plays an important role in explaining variation of concentration of phosphorous of the lake. Sulphate values further, enhances the capability of model by increasing the proportion of explained variation to 85% indicating the important role of sulphate in explaining the variation in phosphorous concentration. As calcium is abundant in mineral present in the study area and thereby resulting in eutrophication of the lake.

It can be seen that the statistical models developed in this study perform well in computing the phosphorous concentrations for the lake. The phosphorus values obtained by the developed regression model indicate statistically significant regressions. Fig. 2 presents the comparison of observed and model computed phosphorous concentrations for the lake which suggests good agreement in observed and model computed phosphorous concentrations with r^2 equal to 0.81.

Thus the water quality constituents of lake play a significant role in controlling the concentration of phosphorous and thereby the eutrophication of the lake and the statistical model developed in this study performs satisfactory in computing the phosphorous concentrations for the lake.

CONCLUSION

The developed statistical model in this study is successful in explaining 85% variation in phosphorous concentration. The developed model may be used for prediction of missing observed value of phosphorous concentration in Renuka Lake.

Table 3 : Selected sets/subsets, candidate for possible model independent variables for Renuka Lake

Number of Variables	Set of independent	N	R^2	F-value	SSE
8	NO_3+Ca+SO_4+Hard+BOD+Alk+Na+K	47	0.904	44.918	0.024
7	NO_3+Ca+SO_4+Hard+BOD+Alk+Na	47	0.900	50.132	0.025
6	NO_3+Ca+SO_4+Hard+BOD+Alk	47	0.896	57.567	0.026
5	NO_3+Ca+SO_4+Hard+BOD	47	0.882	61.257	0.029
4	NO_3+Ca+SO_4+Hard	47	0.869	69.830	0.032
3	NO_3+Ca+SO_4	47	0.854	83.745	0.036
2	NO_3+Ca	47	0.792	83.719	0.052
1	NO_3	47	0.534	51.467	0.116

Table 4 : Selection of model variables on the basis of F-statistics for Renuka Lake

Full model with k-parameters			Reduced model with (k-m) Parameters		k-m	N-k-1	$F_{k-m, N-k-1}$	F* (α=0.05)	Preferred model
Model	N	SSE	Model	SSE					
NO_3+Ca+SO_4+Hard+BOD+Alk+Na+K	47	0.024	NO_3+Ca+SO_4+Hard+BOD+Alk+Na	0.025	7	38	0.226	2.26	Reduced
NO_3+Ca+SO_4+Hard+BOD+Alk+Na	47	0.025	NO_3+Ca+SO_4+Hard+BOD+Alk	0.026	6	39	0.260	2.34	Reduced
NO_3+Ca+SO_4+Hard+BOD+Alk	47	0.026	NO_3+Ca+SO_4+Hard+BOD	0.029	5	40	0.923	2.45	Reduced
NO_3+Ca+SO_4+Hard+BOD	47	0.029	NO_3+Ca+SO_4+Hard	0.032	4	41	1.060	2.61	Reduced
NO_3+Ca+SO_4+Hard	47	0.032	NO_3+Ca+SO_4	0.036	3	42	1.750	2.84	Reduced
NO_3+Ca+SO_4	47	0.036	NO_3+Ca	0.052	2	43	9.556	3.23	Full
NO_3+Ca+SO_4	47	0.036	NO_3	0.116	1	43	95.556	4.08	Full

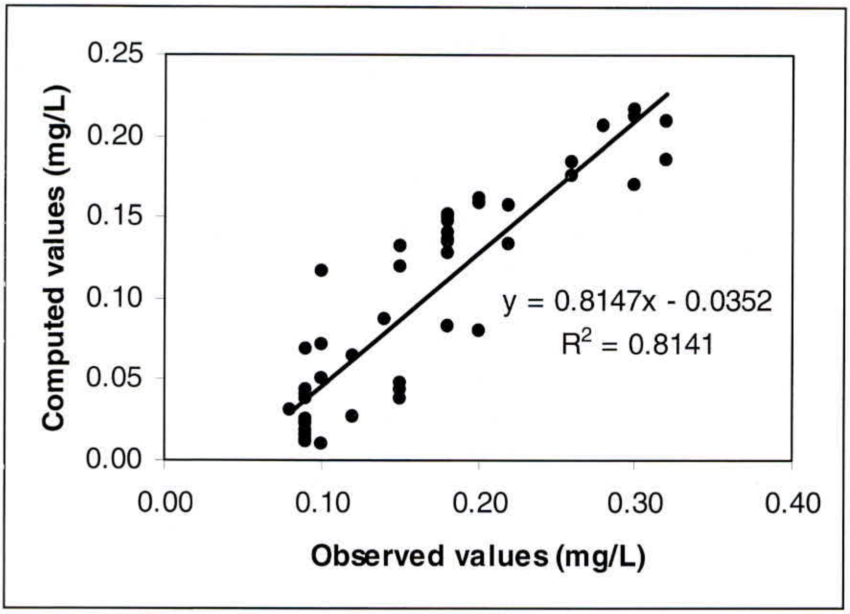

Fig. 2 : Comparison of Observed and Computed values of Phosphorous concentration

REFERENCES

1. APHA. 1992. Standard Methods for the Examination of Water and Waste water, American Public Health Association, New York.
2. Balasankar, T. and S. Nagarajan. 2000. A correlation study on physico-chemical characteristics of ground water in and around Cuddalore Sipcot, Tamil Nadu, Ind. J. Environ. Prot. 20(6), 427-429.
3. Bhatia, K.K.S., A. Tyagi and M.K. Sharma. 1997. Groundwater conductivity modelling using Best Subset Procedure, Proceeding of International Symposium on Emerging Trends in Hydrlogy, Department of Hydrology, University of Roorkee, Roorkee, Sept, 25-27, 1997, pp. 41-53.
4. Draper, N.R. and H. Smith. 1981. Applied regression analysis, Second edition, John Wiley and Sons, New York.
5. Jain, C.K. and K.K.S. Bhatia. 1987. Physico-chemical Analysis of Water and Wastewater, User's Manual UM-26, National Institute of Hydrology, Roorkee.
6. Jain, C.K. and M.K. Sharma. 1997. Relationship among water quality parameters of ground water of Jammu district, Poll. Res. 16(4), 241-246.
7. Jain, C.K. and M.K. Sharma. 2000. Regression analysis of water quality data of Sagar district, Madhya Pradesh, Ind. J. Environ. Health. 42(4), 159.
8. Jain, C.K. and M.K. Sharma. 2002. Regression analysis of ground water quality

data of Malprabha river basin, Karnataka, J. Indian Water Resources Society. 22(1), 30-35.

9. Jain, C.K., Ali Imran and M.K. Sharma. 1998. Salinity modelling of ground water of a coastal region using best subset procedure, Ind. J. Environ. Prot. 18(10), 762-768.
10. Jain, C.K., K.K.S. Bhatia, and M.K. Sharma. 1999. Statistical analysis of ground water quality data of Doon Valley, Dehradun, J. Applied Hydrology, XII (2&3), 27-33.
11. Krishna, J.S.R., K. Rambabu and C. Rambabu. 1995. Monitoring, correlations, and water quality index of well waters of Reddigudem Mandal, Ind. J. Environ. Prot. 15(12), 914-919.
12. Kumar, S., N. Garg and K. Gopal. 1994. Relationship between electrical conductivity and cationic and anionic contents in ground water, Ind. J. Environ. Prot. 14(8), 595-603.
13. Montgomery, D. and E. Peck. 1982. Linear regression analysis, John Wiley and Sons, New York, N Y.
14. Rao, K.S. and B.S. Rao. 1994. Correlations among water quality parameters of ground waters of Musunur Mandal, Krishna district, Ind. J. Environ. Prot. 14(7), 528-532.
15. Reddy, M.S. and N.V.V. Char. 2004. Management of Lakes in India (http://www.worldlakes.org/uploads/Management_of_lakes_in_India-10 Mar04.pdf)
16. Singh, R., S.H. Mishra, R.K. Shyamananda, G. Sharma, I. Mahajan and B.K. Aggarwal. 1987. Morphometry and catchment study of Renuka lake, Himachal Pradesh, India, with a note on its flora and fauna, In: Western Himalayas. Vol. II (Eds Pangtey, Y.P.S. and Joshi, S.C.), Gyanodaya Prakashan, Nainital, pp. 639-649.
17. Sharma, M.K. and C.K. Jain. 2005. Fluoride modeling using Best Subset Procedure, Journal of Applied Hydrology. XVIII, 12-22.
18. Sharma, M.K., V.K. Choubey and V.K. Dwivedi. 2008. Phosphorous modeling of Upper Bhopal Lake using Best Subset Procedure, 12th World Lake Conference Taal 2007, 28 Oct. – 2 Nov. 2007 Jaipur (Rajasthan), 2007.
19. Singanan, M., K.S. Rao and C. Rambabu. 1995. A correlation study on physico-chemical characteristics of ground water in Rameswaram Island, Ind. J. Environ. Prot. 15(3), 212-217.
20. Weisberg, S. 1980. Applied linear regression, John Wiley and Sons, New York, NY.