Volume : 2 | Issue : 2 | Feb 2013 • ISSN No 2277 - 8160

Research Paper | Engineering

Groundwater Quality Monitoring-Statistical Modeling

Dr. S. Siddi Raju Professor in civil engineering, Siddhartha Institute of Engineering & Technology, Puttur – 517 583, Chittoor District, A.P.

ABSTRACT *Water quality data obtained from Pulang river basin, Cuddapah District which is a hard rock terrain was subjected to multiple regression analysis. Ninety two wells were monitored; each sample was analyzed for 10 parameters. Multiple regression analysis model was be used as a positive predictive tool in finding out the Chemistry of groundwater if the dependent variable/SEC is measured at any location. One important application of this analysis is that the entire water quality can be predicted through a single simple test like SEC which in turn clearly indicating the total dissolved solids.*

KEYWORDS: Ground Water, Statistical Model, Geology, SEC

INTRODUCTION:

Water is the most valuable and vital resource for sustenance of life and also for any developmental activity. With the surface water sources dwindling to meet the various demands, groundwater has become the only reliable resource. Groundwater is a natural resource for the economic and secure provision of drinking water, which plays a fundamental role in human well being.

.

The Pulang River originates from the seshachalam hill ranges in rajampet taluk of cuddapah district, A.P.., Flows in a general northerly direction and joins Cheyyeru River. The area of the pulang river basin accounts for 757 km^2 with its six sub basins with a course length of 74 km, the basin is located in between the north latitude 13°5' and 14°15' and the east longitude 70°2'2" and 79°18'50". Generally the climate of the river basin is semi arid. The basin receives an annual average rainfall of 795 mm, out of which 45 to 65 % of the total rainfall is receives during the north east monsoon period.

GEOLOGY:

The rock formations of te basin area represents a suite of sedimentary and metamorphic rocks formed during Precambrian times. Lithologically the cuddapah formations are predominantly argillaceous sequence with a subordinate calcareous sediments. Characteristically each group starts with quartzite and ends with dolomite or shale/phyllite. Structurally, the rocks have a general trend of NNW-SSE with dips of formations generally varying from 10 to 40°.

The Nagari Quarzite is exposed mainly in the southern part of the basin. This is dominantly an arenaceous consisting of conglomerate quartzite, quartzite with shale formations. The Pullampet formation rests over the Nagari quartzite conformably in the southern part of the basin with purple shale, carbonaceous shale and calcareous shale with prominent intercalations of dolamitic lime stones. The basal part of the Pullampet is marked by ferruginous chert and jasper with lensoid dolomite patches. Large outcrops of quartzite are seen in the southern and western portion of the area as hills and ridges.

Dolamitic limestone occurs at places as discontinuous interbands and lenses. The shale occurs mostly in low lying bands and strike in a NNW direction with variable dips. The shale is intersperse by bands with quartzite , which sometimes occurs as low lying elongated hillocks. Alluvium of recent age is composed mostly of sand and subordinately of silty clayey sand and is confined all along on either side of the purling river.

METHODOLOGY:

sampling of ground water has been carried out as per the established norms. The mode of collection of 92 ground water samples collected from different wells and bore wells are analysed systematically as given by Rainwater and Thatcher (1960) , E.Brown et.al (1970) and Hem(1970) . The constituents include silica , calcium chloride , dissolved solids , hardness as $CaCo_{3,}$ alkalinity as $CaCo_{3,}$ non carbon hardness, hydrogen ion concentration , sodium adsorption ratio, percent sodium potential salinity and residual sodium carbonate.

MULTIPLE LINEAR REGRESSIONS

The 'regression' model assumes that some variable 'Y' responds to changes in other 'X' variables. The 'Y' variable is the quantity under study and is known is 'response' or 'dependent' variable. The X variables are those which exhibit a casual effect on the value of the Y and are known as the 'explanatory' or 'independent' variables. The Model is expressed by

$Y = bo + b_1 x 1 + \ldots\ldots b_k X_k$

Where b_o, b_1, ...b_k are the least squares estimate of the unknown parameters, bo is the intercept, while b_1... b_k are regression coefficients. They are chosen in such a way as to minimize the squared sum of the residuals or deviations from the estimated regression plane.

The major issues in the development of this model are;

a. The identification of those variables which have significant and separate effects on the dependent variables.
b. The model must not only provide a good statistical fit to the present day data but must also be a logical and meaningful one.
c. The variables must be a meaningful one in explaining the dependent variable behaviour.

With such an equation, it is possible to develop future levels of the dependent variable given future predictor indicators. The adequacy of the model can be tested by the analysis of variance approach.

The total sum of square (SST) is decomposed into regression sum of squares (SSR) and error sum of squares (SSE).

SST = SSR + SSE

The multiple correlation coefficient R^2

$R^2 = SSR / SST$

This indicates the degree of association between independent variable and the dependent variable. It varies between 0 and 1 , closer to 0 is worse but closer to 1 is better. The significance of R is that its square R^2 is approximately the decimal fraction of the variation of 'Y' accounted by independent variables xi, i.e., if R^2=0.981 then 98 percent of the total variance in the data is explained by the model.

T TEST:

The't' statistic indicates the significance of the regression coefficient of each independent variable. Independent variables which have a 't' value of less than the table 't' value at the degrees of freedom do not have a significant relationship with the dependent variable and therefore, contribute nothing to the equation . If t3 calculated for a parameter 3 is 2.543 and t at 90 percent level at (n-k-l) degrees of freedom is 3.36 from tables i.e. t^3 calculated , the coefficient does not significantly differ from Zero.

F TEST:

The regression sum of squares can be used to give some indication concerning whether or not the model is an adequate explanation of the true situation. One test is the F ratio, Where , F follows F-distribution With

F = SSR/K SSE (n-k-1)

Where K,n-k-1 are degrees of freedom

From F tables at prescribed confidence level, the value of F can be known. F calculated must be more than F tables in which case the variation in Y is explained and is not by chance

DEVELOPMENT OF LINEAR REGRESSION MODELS

In water samples, EC (in micro mhos) at 25°C is mainly dependent as the concentrations of major ions like SiO_2, Ca, Mg, Na, K, HCO_3, CO_3, Cl, SO_4 and pH. The functional relationship could be taken as EC=F (Ca, Mg, Na, K, HCO_3, CO_3, Cl, SO_4, SiO_2, and pH) and the regulating regression coefficients of a data matrix could be used for the development of a model. A simple conversion steps has been adopted in this study to compute parameters of a known value of the dependent variable i.e., E.C.

The dominant factors and principles components of the Pulang river basin extracted from the loading matrix are.

Basin	Extracted Parameters	Chemical Parameters	Cumulative % of variance
Pulang	FACTOR I	Cl, Mg, HCO_3, Na	34.3
	FACTOR II	CO_3, pH, Na, Cl	54.5
	FACTOR III	K, Ca, Mg	65.1
	FACTOR IV	SO_4, SiO_2, Mg	75.4

In the present investigation the water samples are analyzed to determine 10 different chemical properties like Silica, Calcium, Magnesium, Sodium, Potassium, Chloride, Bicarbonate, Carbonate, Sulfate, Total dissolved solids (TDS), and hydrogen ion-concentration.

Since TDS is the single parameter which could reflect the influences of all the dissolved constituents, it is desirable to develop a model by means of which TDS could be predicted or computed if give all or a few of the ten independent chemical constituents. In other words it is hypothesizes that

TDS = fe'(SiO2, Na, Ca, K, Mg, Cl, (HCO3 + CO3) SO4, pH)

There could be several types of functional forms but multiple linear regression is the most powerful and easily explainable model available in the literature. For successful development of such a model, the causal variable must be truly independent of each other as explained earlier. For this purpose factor analysis was performed on the test battery and it was found that there are four basic dimensions which are truly uncorrelated with a view to develop a number of regression models for possible selection. All the possible combinations are tried in this case.

The first four factors show verification in the dependent variable. The estimates of coefficients of the factors are all significant at 1 percent level. Factor four appears to be an most important one to account for variation in dependent variable. Whenever this factor increases the second and third factors are more important than the first one in influencing the dependent variable. All the four factors and the dependent variable move in the same direction.

The first and eight original variables operate through the factor four and generate a significant positive impact on dependent variable.

The fourth, sixth and tenth original variables operate through the factor two and generate positive impact on dependent variable.

The second and fifth original variables operate through the factor three and generate a significant positive impact and dependent variable.

The third and nineth original variables through the factor one and generate a significant negative impact on dependent variable.

The regression estimates are statistically significant as shown in Table 1. The equation has a good fit with a very high R^2 values and highly significant F value. Durbin Watson (Table 2) is within the tolerable limit and does not indicate the presence of the autocorrelation. The variance-covariance, Matrix for coefficience also confirm the reliability of results (Table 3). The regression estimates have been made for standardized data using the same variables. One important application of this analysis is that the entire water quality can be predicted through a single simple test like SEC which in turn clearly indicating the total dissolved solids.

TABLE 1 : Unweighted least squares linear regression of TDS

Predictor variables	Coefficient	STD error	Student's T	P
CONSTANT	571.37	3.7210	153.55	0.0000
P1	-128.39	2.0206	-63.54	0.0000
P2	90.64	2.6298	34.46	0.0000
P3	-14.78	3.6442	-4.06	0.0002
P4	-23.52	3.6719	-6.40	0.0000

Cases included	92
Missing cases	0
Degrees of freedom	87
Overall F	1.321E+03
P value	0.0000
Adjusted R squared	0.9831
R Squared	0.9838
Resid. Mean square	1.274E+03

TABLE 2 : Durbin – Watson test for Autocorelation

DURBIN-WATSON STATISTIC	1.8906
P VALUES, USING DURBIN – WATSON'S BETA APPROXIMATION:	
P (POSITIVE CORR) =	0.2681,
P (NEGATIVE CORR) =	0.7319
EXPECTED VALUE OF DURBIN – WATSON STATISTIC	2.0188
EXACT VARIANCE OF DURBIN –WATSON STATISTIC	4.2481E-02
CASES INCLUDED	92
MISSING CASES	0

TABLE 3 : Variance – covariance matrix for coefficients

CONSTANT	P1	P2	P3	P4	
CONSTANT	13.85				
P1	-5.503E-10	4.083			
P2	-5.648E-09	3.387E-10	6.916		
P3	-9.166E-09	5.950E-10	-3.273E-10	13.28	
P4	5.562E-10	-6.941E-11	-4.576E-10	-1.076E-09	13.48

REFERENCES

Davis John C., 1973, Statistics and data analysis in Geology John Willey and Song, New York, pp. 412-536. | Herman, H.H., 1976. Modern factor analysis, 3rd Ed. The University of Chicago Press, 487. | Lawrence, F.W., and Upchursh S.B., 1982. Identification of recharge areas using geochemical factor analysis. Groundwater, V. 25, pp. 914-923. | Ahmed E.Hassan, Hesham M.Bekhit, Jenny B.Chapman (2008) 'Uncertainty assessment of a stochastic groundwater flow model using GLUE analysis' Journal of Hydrology, No. 362 pp. 89109. | Alexander Y.Sun (2008) 'CONSID: A Toolbox for contaminant Source dentification' Ground Water, Vol. 46 No. 4 pp. 638641. | 3 Bhasker K. Rao and Deborah L. Hathaway (1989) 'A three dimensional mixing cell solute transport and its application' Ground Water, Vol. 27 No. 4 pp. 509516.