Estimation of Water Quality Parameters Using Artificial Neural Networks (ANN) & Geographic Information System (GIS) in Kulsi River Basin, Assam/Meghalaya

C. K. Jain*, S. K. Sharma** and R. D. Singh*
*National Institute of Hydrology, Roorkee – 247 667 (Uttarakhand)
**NIH-Centre for Flood Management Studies, Dispur, Guwahati – 781 006 (Assam)

Abstract:- The spatial distribution of quality of ground water is influenced by seasonal changes and is primarily governed by the extent and composition of its dissolved solids. The contamination of ground water quality is due to the adverse effects of man's activity at ground surface, unintentionally by agriculture, domestic and industrial effluents and unexpectedly by sub-surface or surface disposal of sewage and industrial wastes. The situation though not alarming at the moment in the North-Eastern Region of the country, requires systematic, scientific and sustainable approach in managing its ground water resources and its quality. The Kulsi River Basin, situated on the south bank of mighty River Brahmaputra has been selected to implement Integrated Water Resources Management (IWRM) studies. The sub-basin spreads in the Kamrup District of Assam as well as West Khasi Hills and Ribhoi District of Meghalaya. The River Kulsi drains out a total area of 2806 km2. In this paper, an attempt has been made to couple the processing capabilities of Geographic Information System (GIS) with modeling capabilities of Artificial Neural Network (ANN) to estimate the water quality parameters for Kulsi River Basin. The data used in the study comprise of historical monthly rainfall data for 21 years (1990-2010) procured from Indian Meteorological Department (IMD), Digital Elevation Model (DEM) downloaded from Shuttle Radar Topography Mission (SRTM) and ground water samples collected from fifty different locations comprising of various abstraction sources during pre- and post-monsoon seasons during 2012. The coordinates of rain gauges and sampling points were geo-referenced in the GIS platform and overlaid on the slope (in degrees) layer extracted from the DEM. Monsoon season rainfall corresponding to monthly means of June - September were computed for the rain gauges and interpolated using Inverse Distance Weighted (IDW) method of interpolation. The pixel values corresponding to slope, mean monthly rainfall (June, July, August and September) were tabulated for the sampling locations and constituted the input variables of ANN model. Laboratory analysis of the water samples for water quality constituents comprising of pH, TDS, Na, K, Ca, Mg, Cl, SO4, NO3 and F for pre- and post-monsoon were averaged and constituted the output variables of ANN model. The architecture of the feed forward back propagation ANN model used in the study comprised of five input nodes in the input layer, four nodes in the hidden layer and ten nodes in the output layer. Training and Validation sets in the ratio of 3:2 were obtained by random selection from the fifty samples. The performance criterion of the ANN model was evaluated by coefficient of determination (R2) and Root Mean Square Error (RMSE). The ANN model was found to perform satisfactorily for most of the output variables. The ANN model was used to predict the water quality constituents at Shillong, Meghalaya. The values were found to be within prescribed limits. The integration of GIS with ANN model show promising potential and can be used as a handy tool by policy planners to obtain meaningful and usable estimates of water quality parameters from inaccessible and remote areas.

I. INTRODUCTION

Integrated Water Resources Management (IWRM) calls for 'A blue revolution' to ensure more jobs and more crops per drop of water (GWP, 2000). Estimation, conservation and management of available water play a vital role in achieving higher productivity to sustain the increasing requirements. In recent years, an increasing threat to ground water quality due to human activities has become of great importance. A vast majority of ground water quality problems are caused by contamination, over-exploitation, or combination of the two. Most ground water quality problems are difficult to detect and hard to resolve. Ground water quality is slowly but surely declining everywhere. Ground water pollution is intrinsically difficult to detect, since problem may well be concealed below the surface and monitoring is costly, time consuming and somewhat hit-or-miss by nature.

The problem of ground water pollution in several parts of the country has become so acute that unless urgent steps for detailed identification and abatement are taken, extensive ground water resources may be damaged. The wide range of contamination sources is one of the many factors contributing to the complexity of ground water assessment. Pollutants move through several different hydrologic zones as they migrate through the soil to the water table. It is important to know chemical-soil-groundwater interactions in order to assess the fate and impact of pollutant discharged on to the ground. The serious implications of this problem necessitate an integrated approach in explicit terms to undertake ground water pollution monitoring and abatement programmes.

Pilot projects are being encouraged to identify water related issues and outcomes by consulting with different user groups in order to ensure wise water governance. Along these lines, National Institute of Hydrology (NIH) has undertaken a project on Integrated Water Resources Management (IWRM) under Pilot Basin Study (PBS) at each of its Regional Centres, during the XIIth plan period. The PBS programme involves identification of suitable basin in consultation with concerned State Govt. authorities, Hydrometerlogical data collection and storage, data processing and analysis using State of the Art

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ETWQQM -2014 Conference Proceedings

models, preparation of results and findings in a meaningful and usable form for the intended beneficiaries and different stake holders.

Since 1930s, numerous linear and non-linear hydrological models have been developed to simulate rainfall–runoff-ground water relationships. The last two decades have seen an increasing popularity of Artificial Neural Networks (ANNs) for estimating and forecasting water resources variables. The distinct advantage of an ANN is that it learns the previously unknown relationship existing between the input and the output hydrometerlogical variables through a process of training, without a priori knowledge of the catchment characteristics. Presently more and more researchers are utilizing ANNs because these models possess desirable attributes of universal approximation, and the ability to learn from examples without the need for explicit physical and chemical process. The flexibility of ANNs in inclusion of several parameters and its success in capturing the non-linearity of dynamic systems has made it an attractive tool for modeling the hydrological process (Hsu et al., 1995). Extensive reviews on ANN applications in hydrologic simulation and forecasting have been reported in ASCE (2000a, b), Dawson and Wilby (2001), Maier and Dandy (2000). These studies have recommended development of ANNs with relevant model input-outputs as key aspect requiring further attention.

Geographic Information Systems (GIS) provide a suitable platform to store, analyze and manage spatial data. Lately, topographical attributes are widely available with the advent of digital elevation models (DEMs) and digital terrain analysis techniques. The fine resolution of DEMs could help in better capturing of flow patterns over the catchments. These models provide more precise and reproducible estimates than the tedious and time consuming manual techniques applied to topographic maps such as toposheets. Knowledge of landscape morphology along with the hydrologic processes is required to conceptualize the hydrological cycle within a catchment. The combination of soil properties, topographic features and vegetation attributes act as a template for the hydrologic phenomena of the region. Ground water quality of a catchment is a result of interplay between the basins environmental factors such as soil, topography, drainage, and rainfall and land use pattern. Only a handful of studies have investigated the potential of including these attributes in ANN modeling for the prediction of ground water quality.

Sharma et al. (2006) reported that certain combinations of topography (DEM) and vegetation attributes (NDVI) performed better compared to using basic soil properties alone as inputs of ANN for prediction of soil hydraulic properties. Sharma and Tiwari (2009) found selected combinations of monthly rainfall, soil, topography and vegetation performed better than using rainfall alone for prediction of monthly runoff for Damodar Valley Catchments (DVC) using ANN. This study attempts to estimate the water quality parameters of a basin using ANN by using monthly rainfall and slope as inputs. The results of the study will be of immense use to the planners, administrators, scientists and engineers concerned with the management and protection of ground water quality in River Basins.

II. STUDY AREA

The Kulsi River Basin, selected by the Centre for Flood Management Studies (CFMS), NIH, Guwahati for PBS/IWRM studies is a part of the Brahmaputra sub-basin is situated on the south bank of the mighty River Brahmaputra. The basin spreads in the Kamrup District of Assam as well as west Khasi Hills and Ri-bhoi district of Meghalaya. It is located between latitude 25o30'N to 26o10'N and longitude 89o50'E to 91o50'E with an altitude between 100 to 1900 m above msl. The River Kulsi drains out a total area of 2806 km2 within the Kamrup District of Assam as well as west Khasi Hills and Ri-bhoi district of Meghalaya (Fig. 1).

The climate of the basin, excluding the upper most reach is similar to that of the other districts in Central Assam. The winter is cold and foggy, while the summer is hot and humid. There is no meteorological centre within the catchment for observation of temperature and humidity data. However the nearest observatories for the basin are at Guwahati, Umiam and Shillong. Based on long term data from these stations it has been observed that the average maximum temperature in this basin varies between 15 to 330C and average minimum temperature varies from 3 to 120C. The main occupations of the people of the sub-basin are agriculture and cottage industry. There is no big industry and forest wealth is also meagre. The local plain people practice permanent cultivation in low lying and flat terrains using modern techniques of cultivation. But on the other hand the tribal people still depend on "Jhum Cultivation" or shifting cultivation along the hill slopes without adopting any kind of soil conservation measures. The supplementary employment is very limited due to the limitation in industrial growth.
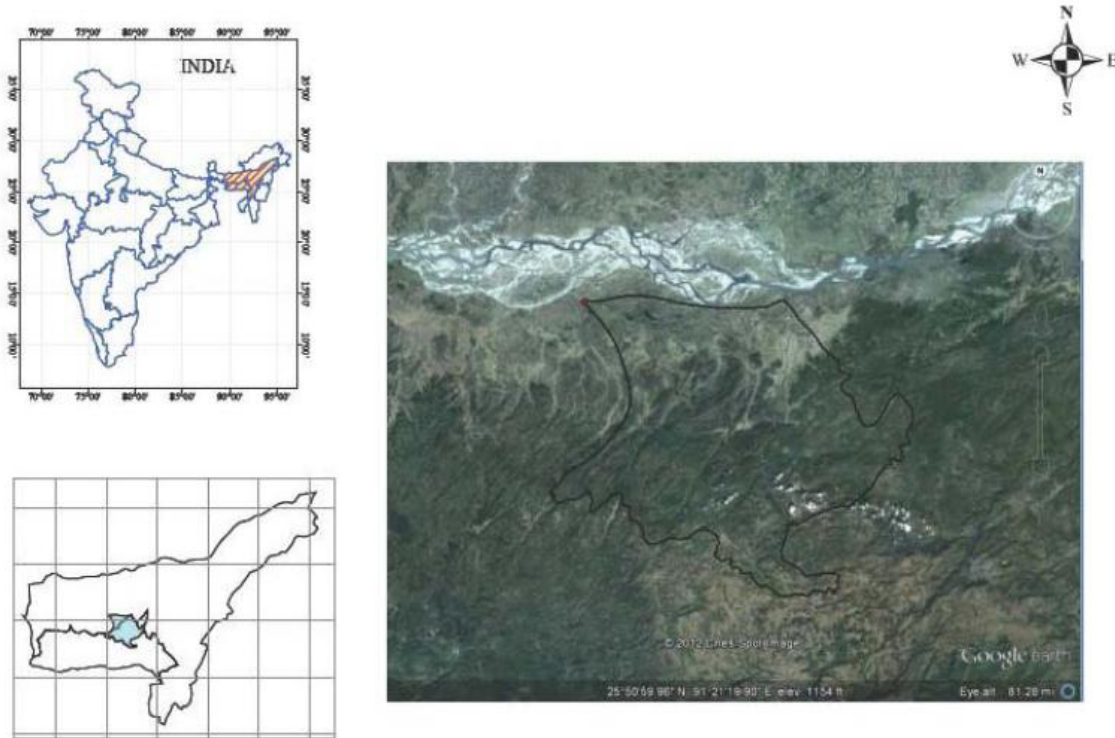
Fig. 1. The location of Kulsi River Basin

III. METHODOLOGY

Fifty ground water samples were collected each during pre- and post-monsoon seasons during 2012 from various abstraction sources in clean polyethylene bottles and preserved by adding an appropriate reagent (Jain and Bhatia, 1988; APHA, 1992). All the samples were stored in sampling kits maintained at 4oC and brought to the laboratory for detailed physico-chemical analysis. The source and depth wise distribution of sampling locations are given in Table 1. The sample collection was limited to Assam boundary of the Basin (Fig. 2).

All general chemicals and reagents used in the study were of analytical reagent grade (Merck/BDH). De-ionized water was used throughout the study. All glassware and other containers used for analysis were thoroughly cleaned by soaking in detergent followed by soaking in 10% nitric acid for 48 h and finally rinsed with de-ionized water several times prior to use. The physico-chemical analysis was performed following standard methods (Jain and Bhatia, 1988; APHA, 1992).

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ETWQQM -2014 Conference Proceedings

Table 1. Distribution of Sampling Sites

Source	Depth range			Total number
	0-20 m	20-40 m	> 40 m	
Hand Pumps	5	1,7,8,9,10, 20,21,22, 23,24,25	6	13
Dug Wells	2-4,11-19, 27-32, 34-41,43,47, 49,50	-	26	30
Tube Wells	-	33,42,44, 45,46,48	-	7
Total	31	17	2	50
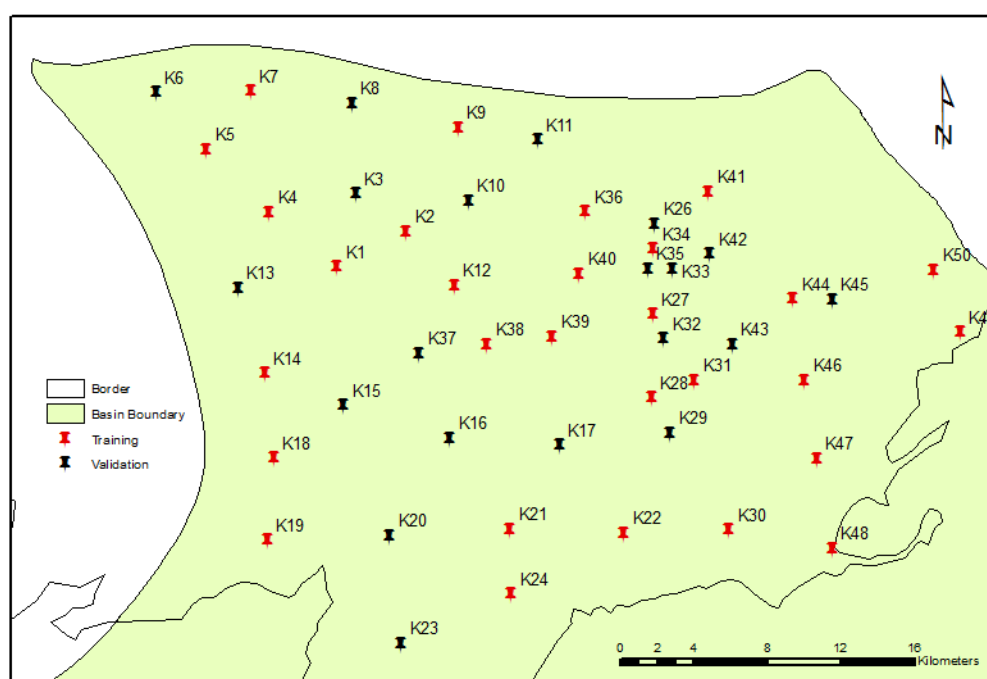
Fig. 2. Location of Sampling Points

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ETWQQM -2014 Conference Proceedings

Data Processing

The input mean monthly rainfall data for the ANN model were extracted from historical monthly concurrent rainfall data for 21 years (1990-2010) procured from Indian Meteorological Department (IMD), corresponding to five weather stations i.e. North Lakhimpur (S_1), Mohanbari (S_2), Guwahati (S_3), Tezpur (S_4) and Silchar (S_5) from Assam. The coordinates of the weather stations were imported into the ArcGIS^{TM} software in Geographic Projection. The processing of rainfall data (S_1 - S_5) in Assam was carried out using Spatial Analyst Tool Box of ArcGIS software. Monsoon season rainfall corresponding to monthly means of June to September were computed and interpolated within Assam using Inverse Distance Weighted (IDW) Method of interpolation. Monthly rainfall raster grids for June, July,

August and September were created. The topography of the basin was characterized with SRTM DEM of 90 m by 90 m resolution (http://www.srtm.csi.cgiar.org). The slope (in degrees) is a measure of the maximum rate of change of elevation between each cell and its neighbors. The slope grid for Assam was created.

The sampling locations point vector shapefile was overlaid on top of raster grids corresponding to June, July, August, September and Slope. The individual pixel values corresponding to the fifty sampling locations were identified and tabulated. These values constituted input variables of the ANN model. The basic statistics (min, max and mean) of the input and output variables of ANN model is given in Table. 2

Table 2. Basic Statistics of Input-Output Variables of ANN Model

Variables	Minimum	Maximum	Mean
Rainfall (Jun), mm	303	327	315
Rainfall (Jul), mm	316	332	323
Rainfall (Aug), mm	292	397	341
Rainfall (Sep), mm	179	200	190
Slope (in Degrees)	0.11	13.86	2.44
pH	6.5	7.6	6.9
TDS, mg/L	40	393	141
Sodium, mg/L	2.8	47	13
Potassium, mg/L	0.8	15	3.3
Calcium, mg/L	2.4	40	14
Magnesium, mg/L	1.5	32	6.4
Chloride, mg/L	1.5	52	8.5
Sulphate, mg/L	0.3	31	4.4
Nitrate, mg/L	0.1	39	5.2
Fluoride, mg/L	0.04	1.40	0.50

Neural Network Modelling

In this study, neural network modelling was performed using Neuropath software Minasny and McBratney (2002). Mathematically, neural networks can be represented by a set of simple functions linked together by weights. A network with an input vector of elements x_l ($l = 1...., N_l$) is transmitted through a connection that is multiplied by weight W_{jl} to give the hidden units z_j ($j = 1,...,N_h$) :

$$z_j = \sum_{l=1}^{N_l} w_{jl} x_l + w_0$$

(1)

where N_h is the number of hidden units and N_i is the number of input units. The hidden units consist of the weighted input (w_{jl}) and the bias (w_0). The outputs from hidden layer pass another layer of filters with weights (u_{kj})

and bias (u_0) and are fed into another activation function F to produce output y ($k = 1,....N_0$):

$$y_k = F\left[\sum_{j=1}^{N_h} u_{kj} f\left(\sum_{l=1}^{N_l} w_{jl} + w_0\right) + u_0\right]$$

(2)

The weights are adjustable parameters of the network and are determined from a set of data though the process of training. The NL2SOL adaptive nonlinear least squares algorithm Dennis et al (1981) implemented in the Neuropath software was used for training the networks. The objective of training was to minimize the sum of squares of the residuals between the measured and predicted outputs. Fig. 2 shows the architecture of neural network used in the study.
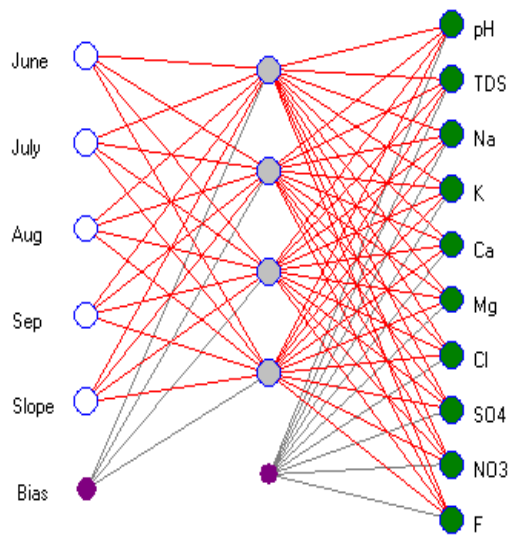
Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ETWQQM -2014 Conference Proceedings

Fig. 2. Architecture of Neural Network Model

IV.   RESULTS AND DISCUSSION

The fifty input-output datasets were randomly divided in the ratio of 3:2 to generate training and validation datasets respectively as shown in Fig. 2. The maximum number of epochs was set to 100. Performance of the neural networks was evaluated by spearman's correlation coefficient (r), Root Mean Square Error (RMSE) and Mean Error (ME) for the validation datasets. The performance of ANN model for estimation of different water quality parameters is shown in Table 3. The scatter plots between estimated and measured water quality attributes are shown in Fig. 3 along with the 1:1 line.

Table 3. Performance of ANN Model for Prediction of Water Quality Attributes

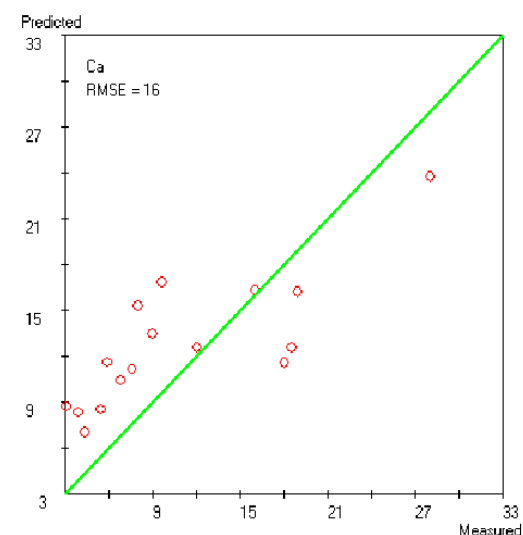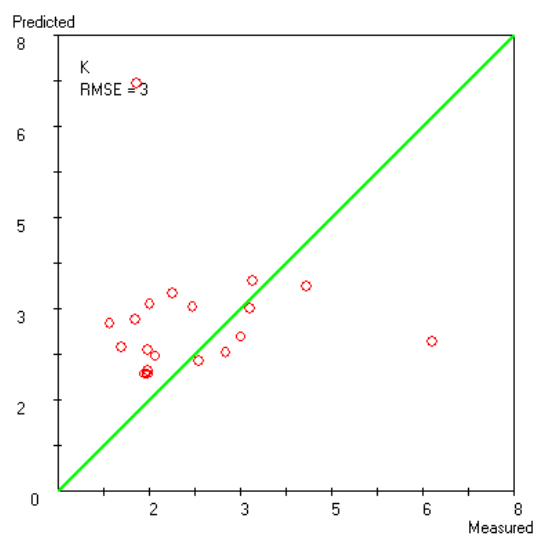| Water Quality attributes	r	RMSE	ME
pH	0.779	0.253	0.066
TDS, mg/L	0.652	225.143	85.718
Sodium, mg/L	0.671	20.954	7.538
Potassium, mg/L	0.719	2.876	0.512
Calcium, mg/L	0.751	15.784	6.515
Magnesium, mg/L	0.568	16.027	6.094
Chloride, mg/L	0.484	13.282	2.809
Sulphate, mg/L	0.632	3.832	0.998
Nitrate, mg/L	0.544	5.955	0.729
Fluoride, mg/L	0.568	0.471	0.161

Based on the correlation coefficient the ANN model performed best for estimation of pH and worst for estimation of chloride. Highest RMSE and ME were observed for TDS due to the range of the values and lowest for pH. The estimation of pH, K and Ca were better (r > 7.0) compared to other parameters, while estimations of Mg, Cl and and NO3 were least (r <6.0) compared to other water quality parameters.

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
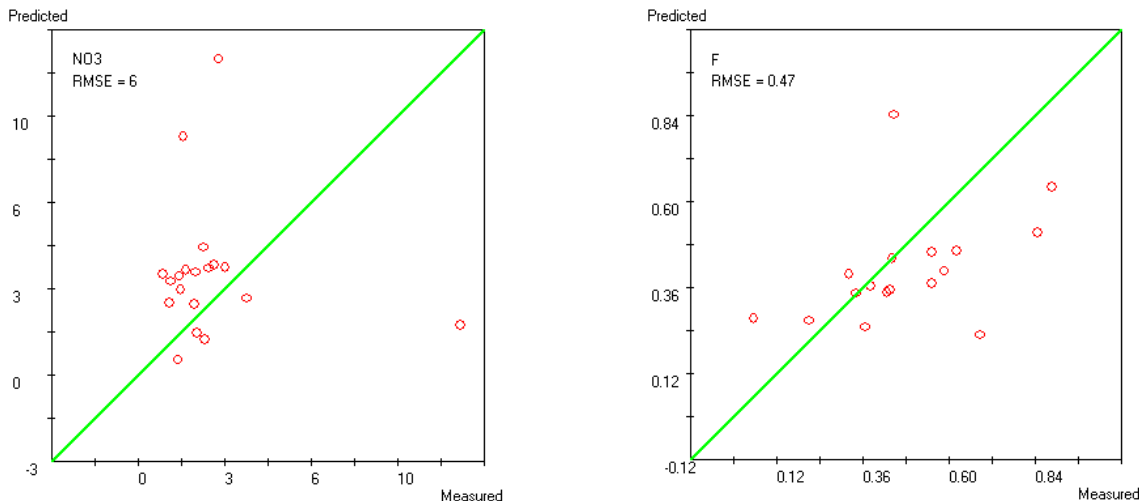**ETWQQM -2014 Conference Proceedings**
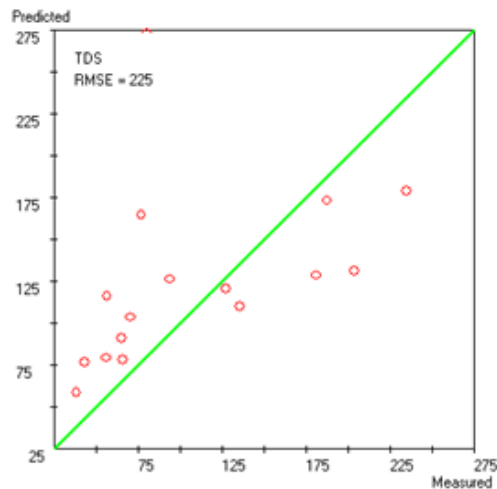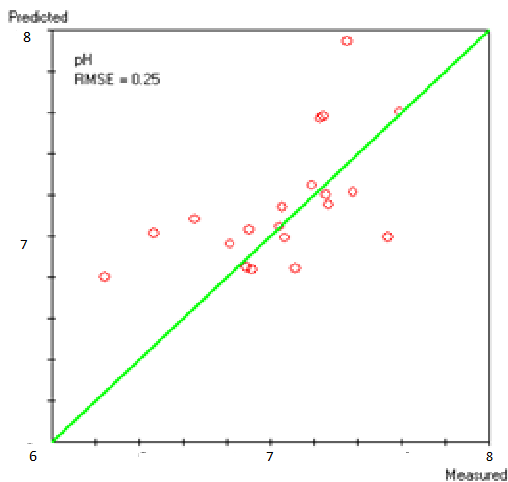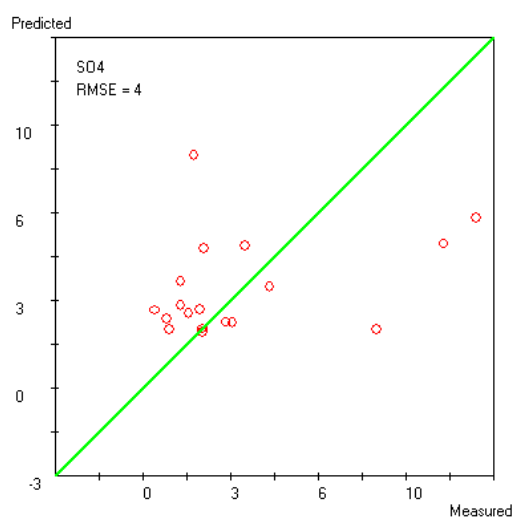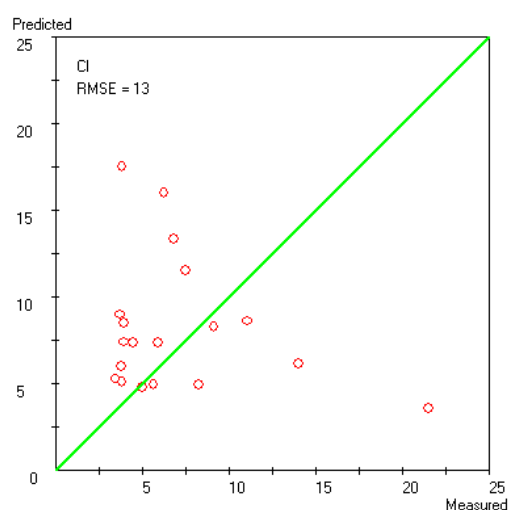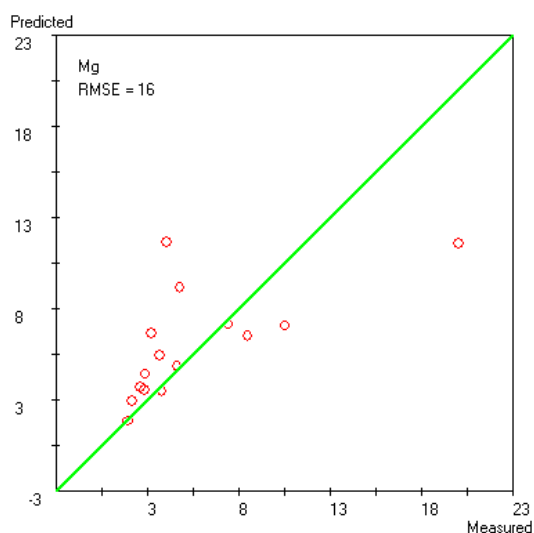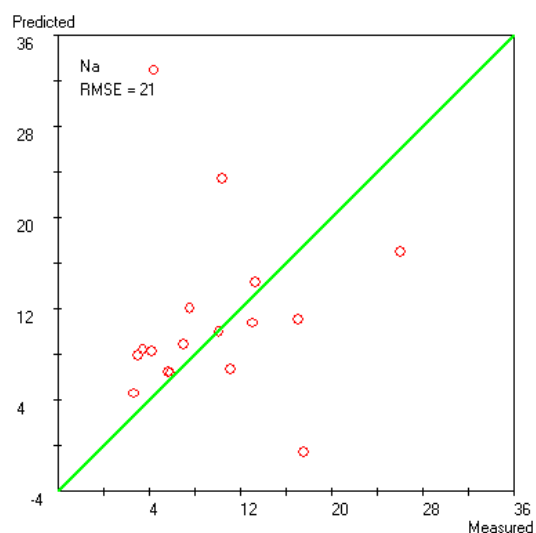
Fig. 3. Scatter Plots of Validation Datasets for different Water Quality Attributes

The developed ANN model was used to predict the water quality parameters at Shillong, Meghalaya. The input values for weather station at Shillong were obtained based on the methodology given in Section III. The monthly mean rainfall of Shillong weather station for the months of June to September was determined as 428 mm, 425 mm, 304 mm and 288 mm respectively. The slope value was determined to be 12.24 degrees. These input variables were fed to ANN model and the resulting output is shown in Table 4.

Table 4. Prediction of Water Quality Attributes using ANN model

Station	pH	TDS, mg/L	Na, mg/L	K, mg/L	Ca, mg/L	Mg, mg/L	Cl, mg/L	SO_4, mg/L	NO_3, mg/L	F, mg/L
Shillong	6.58	66.99	7.07	6.43	17.43	3.56	2.76	4.61	9.44	4.22

V.   CONCLUSIONS

The integration of GIS with ANN model show promising potential and can be used as a handy tool by policy planners to obtain meaningful and usable estimates of water quality parameters from inaccessible and remote areas. The ANN model can be further refined by inclusion of soil, geomorphology and vegetation attributes along with sources of pollution as possible inputs.

REFERENCES

[1]   APHA, Standard Methods for the Examination of Water and Waste Waters, American Public Health Association, 18th Edition, Washington, DC, 1992.
[2]   ASCE. 2000a. Artificial neural networks in hydrology. I: preliminary concepts. J. Hydrol. Eng., 5(2):115-123.
[3]   ASCE. 2000b. Artificial neural networks in hydrology II: hydrologic applications. J. Hydrol. Eng., 5(2):124-137.
[4]   Jain, C. K. and Bhatia, K. K. S. 1988. Physico-chemical Analysis of Water and Wastewater, User's Manual, UM-26, National Institute of Hydrology, Roorkee.
[5]   Dawson, C. W., Wilby, R. L., 2001. Hydrological modelling using artificial neural networks. Progress in Physical Geography 25 (1) , 80–108.
[6]   Dennis, J.E., Gay, D.M., Welsch, R.E., 1981. NL2SOL - An adaptive nonlinear least squares algorithm. ACM Transactions on Mathematical software 7, 348-368.
[7]   Hsu, K.-L., Gupta, H.V., Sorooshian, S., 1995. Artificial neural network modeling in rainfall-runoff process. Water Resources Research 31 (10), 2517-2530.
[8]   Maier, H. and Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. Environmental Modeling Software 15, pp. 101–124
[9]   Minasny, B., McBratney, A. B., 2002. The neuro-m method for fitting neural network parametric pedotransfer functions. Soil Science Society of America Journal 66, 352-361.
[10]   GWP (2000) IWRM at a Glance: Technical Advisory Committee (Stockholm: Global Water Partnership Secretariat).
[11]   Sanjay K. Sharma, Binayak P. Mohanty and Jianting Zhu., 2006 Including Topography and Vegetation Attributes for Developing Pedotransfer Functions, Soil Sci Soc Am J 70:1430-1440.
[12]   Sharma, S. K. and Tiwari, K. N. 2009. Bootstrap based artificial neural network (BANN) analysis for hierarchical prediction of monthly runoff in Upper Damodar Valley Catchment. J. Hydrol., 374:209-222.