An Approach to Determine Land Use Factors to Study Their Influence on Stream Water Quality

Satish Nagalapalli, Jagadeesh Anmala, Murari R R Varma and K Rajitha

December 19, 2021

AN APPROACH TO DETERMINE LAND USE FACTORS TO STUDY THEIR INFLUENCE ON STREAM WATER QUALITY

NAGALAPALLI SATISH[1], JAGADEESH ANMALA[2], MURARI R R VARMA[3], K. RAJITHA[4]

[1]*Graduate Student (Ph.D. Research Scholar), Department of Civil Engineering, Birla Institute of Technology and Science Pilani, Hyderabad Campus, Jawahar Nagar, Kapra (Mandal), Medchal District, Hyderabad, Telangana, India – 500 078,* Email**:** *p20190416@hyderabad.bits-pilani.ac.in*

[2] *Associate Professor, Department of Civil Engineering, Birla Institute of Technology and Science Pilani, Hyderabad Campus, Jawahar Nagar, Kapra (Mandal), Medchal District, Hyderabad, Telangana, India – 500 078,* Email: *jagadeesh@hyderabad.bits-pilani.ac.in*

[3]*Assistant Professor, Department of Civil Engineering, Birla Institute of Technology and Science Pilani, Hyderabad Campus, Jawahar Nagar, Kapra (Mandal), Medchal District, Hyderabad, Telangana, India – 500 078,* Email: *murari@hyderabad.bits-pilani.ac.in*

[4]*Assistant Professor, Department of Civil Engineering, Birla Institute of Technology and Science, Pilani Hyderabad Campus, Jawahar Nagar, Kapra (Mandal), Medchal District, Hyderabad, Telangana, India – 500 078,* Email: *rajitha@hyderabad.bits-pilani.ac.in*

Abstract

Stream water quality is gaining significant attention due to increasing population and anthropogenic sources of pollution. These accelerate water quality stress conditions in developing countries especially, India. Streams are one of the significant water resources, and their quality impacts the watershed. The surrounding land use and land cover (LULC) within the watershed also influence them. In this study, we attempted to find out the primary land use factors of the Godavari river basin. The information of water quality monitoring stations was obtained from the Telangana State Pollution Control Board (TSPCB) website and was delineated using QGIS software. The GlobeLand30 open source classified dataset developed by China compiling the Landsat, and Chinese HJ-1 satellite data was utilized in the current study. The urban, forest, agriculture, grassland, and shrubland were the primary land cover classes considered in this study. First, QGIS software was used to delineate the upper portion or catchment area of sample stations individually. The land use factors for each class were calculated using the arcsine square root of respective LULC areas to their corresponding basin areas. Furthermore, these land use factors are combined with meteorological parameters as inputs to the water quality prediction and analysis model.

Keywords: Land use/Land cover, Godavari river basin, Water Quality Parameters, Watershed delineation

1. Introduction

Nonpoint source pollution is the primary threat to the aquatic environment, the major contribution of these pollutants is from the urban, agricultural, and forest areas. By using advancement in Satellite remote sensing technology, continuous monitoring and identifying the critical riparian areas along the streams will help to reduce the none point source pollution load on streams (Goetz, 2006). Large scale agriculture, bare land, and forest area in the basin protect the nutrient loss by obstructing the runoff from the basin, whereas riparian areas influence the streams with organic matter, woody debris, and nutrient accumulation, majorly nitrate and phosphate (Narumalani et al., 1997).

There have been many studies reported that land use land cover influence the water quality parameters by adopting the Geographic Information System (GIS) and remote sensing techniques with land use land cover quantification and landscape of the watershed (Roth et al., 1996; Allan et al., 1997; Bolstad & Swank, 1997; L. Johnson et al., 1997; L. B. Johnson & Gage, 1997; Huang et al., 2013). Different land occupational or land use effects on the water quality parameters of the Cuiaba river located at the southern part of the Brazilian state of Mato Grosso were studied (Zeilhofer et al., 2006). The study reveals that extensive cattle farming increases the total coliform count and intensive agriculture is the reason for the increase of Chemical Oxygen Demand (COD) and nutrient concentrations. Pisciculture activates the nitrate concentration; moreover, urban agglomeration areas were subjected to the severe chemical, and biological pollutant loads. Basnyat et al. 1999 investigated the impact of land use and land cover on nutrient pollution by quantifying the nitrate pollution contributing zones. Based on satellite imagery, land use/cover, a nutrient linkage model was built and it was observed that urban regions supply more nutrients than agriculture sectors. In contrast, forest operates as a sink for nutrients. Researchers have used linear, nonlinear regression, and machine learning models to predict river water quality.

Maillard and Pinheiro Santos, 2008 developed a cartography model based on GIS and statistics to study the influence of LULC on the water quality parameters. Parameters like Turbidity, Dissolved Oxygen (DO), phosphorus, and fecal coliform revealed a distinct fluctuation with the distance in the stream from sample station. Amiri and Nakane, 2009 applied multilinear regression models along with remote sensing, GIS techniques to study LULC data and observed spatial variability in the distribution of water quality parameters. By coupling the GIS-derived landscape metrics, it was a proportional relationship exists between the DO, pH, and total phosphorus (TP). These are the main effects of stream bank erosion, which further leads to rise in phosphorus load (Ishee et al., 2015).

The representation of stream water quality dynamics is very complex in a model due to multiple variables that could be of natural or anthropogenic influence. Manty studies attempted to develop the linkage models by incorporating the watershed parameters, LULC classifications, meteorological data using artificial neural networks (Girija et al., 2007; Song et al., 2010; Anmala et al., 2015; Kumar et al., 2016). Recent techniques have made it possible to develop more effective techniques for dealing with uncertainty in water quality prediction, such as GIS-based artificial neural network models for precise water quality analysis (Maier and Dandy, 1996; Turuganti et al., 2019; Turuganti et al., 2020; Pujar et al., 2020; Giri, 2021). The Artificial Neural Network (ANN) has been reported to be the most effective and predictive tool used to model nonlinear environmental relationships (Zhang and Stanley, 1997, Jain and Indurthy, 2003, Keskin et al., 2015).

Singh et al. (2004) developed a three layer feedforward -ANN model with a backpropagation learning algorithm to predict the DO and Biochemical Oxygen Demand (BOD) of the Gomti river. After analyzing ten years' monthly data, it was suggested that ANN is capable of predicting long-term trends. Juahir et al. 2004 developed an ANN model to predict the water quality index (WQI) by splitting the data into three sets and performing training, testing, and validation with various independent water quality variables as input parameters. Multilinear regression analysis was used to determine the association between water quality parameters and WQI. The study reveals that ANN models were capable of estimating WQI with satisfactory accuracy when trained without the independent variables. Extending this -technique, Gazzaz et al. 2015 developed an ANN model based on the LULC area as predictors to forecast the WQI. It was discovered that the majority of the mining area influences the water quality. Fu et al. (2013) devised a coupled model based on GIS and improved backpropagation algorithm in ANN to forecast dissolved organic carbon (DOC) in a river network. The model is linked to geographic factors, resulting in a more accurate DOC forecast.

2. Methodology

2.1 Study Area and Data Source

2.1.1 Godavari River basin

Godavari river basin is one of the largest river basins in peninsular India and is called Dakshina Ganga. After the Ganges Basin, the Godavari Basin is the country's second-largest basin, covering roughly 9.50 % of the country's total land area. Over all the basin covers the drainage area of 3,12,812 Km^2 bounded between the Longitudes from 70°24' to 83°4' E and Latitudes from 16° 19' to 22° 34' N. The river originates from Sahyadris, at an altitude of 1,067m above MSL near Trimbakeshwar in the Nashik district of Maharashtra and flows across the Deccan Plateau from the Western to the Eastern Ghats. The river flows through many states and forms interstate boundaries between Telangana and Maharashtra, Telangana and Chattisgarh etc. The river covers the major geographic areas in Maharashtra (48.7%), Telangana (19.87%), Andhra Pradesh (3.53%), and Chhattisgarh (10.69%), Madhya Pradesh (10.17%), Odisha (5.67%), Karnataka (1.41%) and less area in Pondicherry respectively (source: Ministry of Jal Shanti –Godavari river management. grmb.gov.in accessed on 15th September 2021). The basin receives most of its rainfall during the southwest monsoon season (July-September), with minimum and maximum annual rainfall ranging from 881 mm to 1395 mm, with an average of 1110 mm. The annual maximum temperature ranges from 31^0C to 33.5 ^{0}C (India Meteorological Department website accessed on 15th September 2021). The western portion of the basin is much hotter than the central, northern, and eastern regions of the basin.

2.1.2 Data collection

For qualitative and quantitative assessment of stream waters, three types of data sets were required as input to the model. They are water quality data, meteorological data, and geospatial data.

2.1.2.1 Water Quality Data

The monthly water quality data from January 2019 to April 2021 was collected from the Telangana State Pollution Control Board website (tspcb.cgg.gov.in). This data was spatially referenced into the QGIS platform based on the latitude and longitude of the sampling stations. In this monthly observed data, continuous monitoring stations are sorted and considered for the study. The LULC factors were computed for these consistent stations where the data is available on a continuous basis.

2.1.2.2 Meteorological Data

The rainfall and temperature data were collected from the IMD website (cdsp.imdpune.gov.in). The rainfall data was available from 1901 to 2020 in the NetCDF format at 0.25° x 0.25° grid resolution and temperature data was available from 1951 to 2020 at 1° x 1° grid resolution. To process NetCDF file, python console was used in the QGIS software and all files were run in loop and extracted in the excel format. For 2021-year, the rainfall and temperature data were collected from the NASA web site (power.larc.nasa.gov/data-access-viewer). The data were downloaded by specifying the latitude and longitude of the sampling stations.

2.1.2.3 Geospatial Data

Land use factors are generated by dividing the watershed area into five basic classes: urban, forest, agricultural, grass, and shrub land uses, in order to characterize the non-point source influences of urban, agricultural, and industrial discharges on stream water quality. To arrive at these factors, the data for GlobeLand30 land use land cover (global land cover mapping at 30 m resolution) was taken from the website www.globallandcover.com. The LULC data was processed into the requisite basin form, and sampling sites were divided into sub-basins. All of the sampling stations' upstream drainage zones have been identified, and rasters have been converted to vector files. To obtain the LULC into the designated sample station region, the basin vectors were projected into processed LULC and a clip extent operation was done. The size of each field is then determined using the field calculator, and then added to all of the related field areas as another type of classed areas (urban, forest, agricultural, grass, and shrub). Similarly, all the consistent data stations were delimited, merged, and clipped for corresponding sample stations to extract the land use components for the study using QGIS software. To put it another way, the upper drainage part of each sampling station is delimited, land use/land cover is added, and the shape file is extracted. The areas of each classification are then calculated when similar fields are recognized. The land use land cover factors such as urban land use factor (ULUF), forest land use factor (FLUF), agricultural land use factor (ALUF), grass land use factor (GLUF), and shrub land use factor (SLUF) are calculated as below:

$$ULUF = \arcsin\sqrt{\frac{urban\,area}{catchment\,area}} \qquad (1)$$

$$FLUF = \arcsin\sqrt{\frac{forest\,area}{catchment\,area}} \qquad (2)$$

$$ALUF = \arcsin\sqrt{\frac{agricultural\,area}{catchment\,area}} \qquad (3)$$

$$GLUF = \arcsin\sqrt{\frac{grassland\,area}{catchment\,area}} \qquad (4)$$

$$SLUF = \arcsin\sqrt{\frac{shrubland\,area}{catchment\,area}} \qquad (5)$$

The above land use factors are combined with climate parameters precipitation and temperature to predict for the three water quality parameters namely: conductivity, dissolved oxygen (DO), and total dissolved solids (TDS). The land use map of the Godavari watershed is shown in Figure 1.

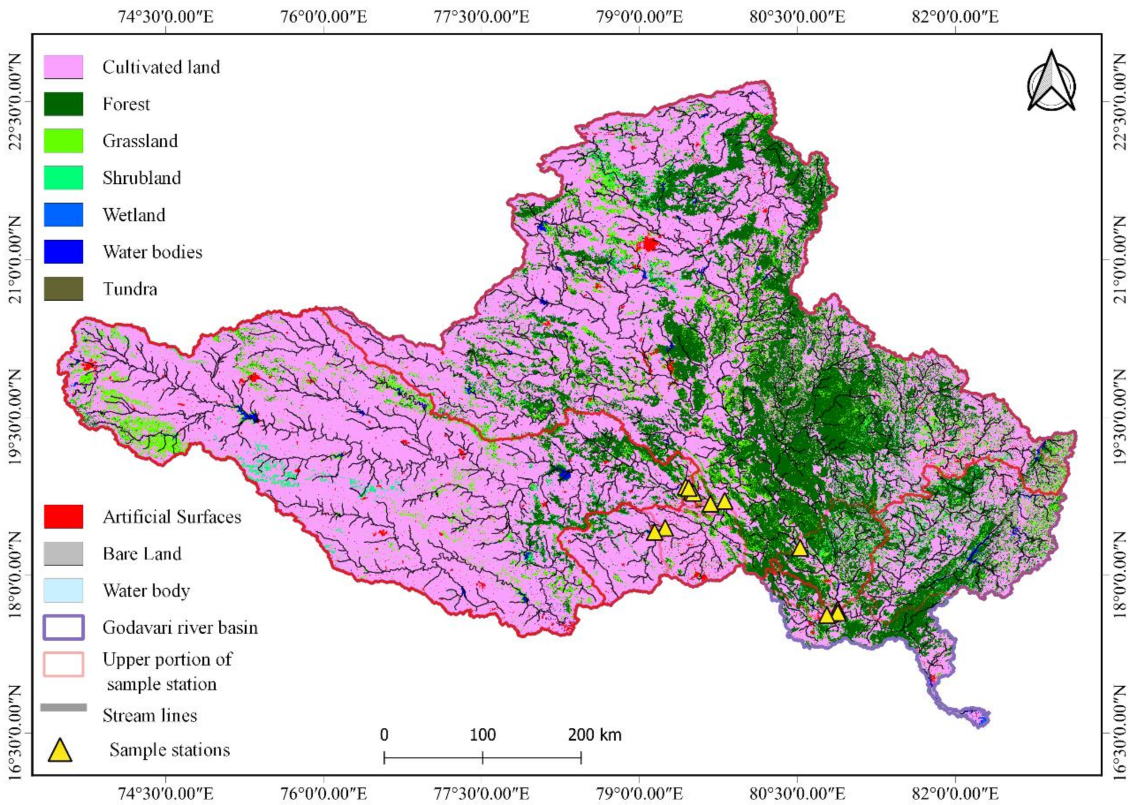

Figure 1 Land use map of study area

2.2 Artificial Neural Network Modeling

Artificial neural network modeling has been carried out using MATLAB R2020b and SPSS frameworks. Three-layer feedforward ANNs consisting of one input layer, one hidden layer and an output layer were considered. The neurons /nodes were connected from input to hidden layers, and hidden to output layers with biases at each layer. The interconnections between the layers are associated with weights which are determined for example using TRAINLM function within the MATLAB environment. The mean daily precipitation (P) in a month, maximum temperature (T_{max}), minimum temperature (T_{min}), land use factors of agriculture, forest, grass, shrub, urban represented as ALUF, FLUF, GLUF, SLUF, ULUF, a total of eight input parameters were considered for each data sample, respectively. Three water quality parameters namely, conductivity, DO, and TDS, were taken as individual outputs for the neural network modeling. The inverse distance weighted interpolation method (IDW) has been used to estimate rainfall and temperatures at sample stations. The data was available for a length of 255 sample observations from the year 2019 to 2021 at 13 sample stations. The data was normalized and separated into training, testing and validation groups with -proportions of 70%, 15%, and 15%, respectively. The three training functions Levenberg-Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient were tried out and the best results are obtained using Levenberg-Marquardt training algorithm in the present ANN model.

3. Results and Discussions

The SPSS multilayer perceptron neural (MPN) networks were used first to model water quality parameters. The networks were trained with 70% dataset and tested with 30% dataset to predict for the water quality parameters of conductivity, DO, and TDS. This was done using trial and error method for network architecture. Finally, with a single hidden layer consisting of 5 hidden nodes was considered for conductivity and DO, and a hidden layer consisting of 8 hidden nodes was considered for TDS. Using these network architectures of 8-5-1, 8-8-1, the neural network training has been performed for the three parameters. The performance of SPSS MPN was analyzed with Root Mean Squared Error (RMSE), Correlation Coefficient (CC), Index of Agreement (D), Mean Absolute Error (MAE), Mean Bias Error (MBE), and Nash-Sutcliffe Coefficient of Efficiency (NSCE). The results of above parameters are given in Table 1 for the three parameters.

Table 1 RMSE, R^2, D, MAE, MBE, NSCE, network architecture using SPSS with Scaled conjugate training algorithm

S/NO	Parameter/ Test	RMSE	R^2	D	MAE	MBE	NSCE	Network architecture
1	Conductivity (mS/cm)	133.07	0.79	0.92	91.43	-2.76	0.78	8-5-1
2	DO (mg/L)	0.50	0.52	0.71	0.40	0.04	0.51	8-5-1
3	TDS (mg/L)	84.88	0.76	0.93	59.35	5.76	0.76	8-8-1

The 'ideal' or 'perfect' values of these parameters are given in Table 2.

Table 2 List of statistical indices used (Willmott et al., 1985).

Statistical Index	**Unit**	**Equation**	**Perfect value**
Root Mean Squared Error (RMSE)	mm	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n}}$	0
Correlation Coefficient (CC)	NA	$r = \frac{\sum_{i=1}^{n}(O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O})^2}\sqrt{\sum_{i=1}^{n}(P_i - \overline{P})^2}}$	1
Index of agreement (d)	NA	$d = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(\lvert P_i - \overline{O}\rvert + \lvert O_i - \overline{O}\rvert)^2}$	1
Mean Absolute Error (MAE)	mm	$MAE = \frac{1}{n} \times \sum_{i=1}^{n} \lvert O_i - P_i \rvert$	0

Mean Bias Error (MBE)	mm	$MBE = \frac{1}{n} \times \sum_{i=1}^{n} (P\iota - O\iota)$	0
Nash-Sutcliffe Coefficient of Efficiency (NSCE)	NA	$NSE = 1 - \frac{\sum_{i=1}^{n} (O\iota - P\iota)^2}{\sum_{i=1}^{n} (O\iota - \overline{O})^2}$	1

where $O\iota$ refers to observation data and $P\iota$ shows the prediction of model data, n stands for number of samples, $\overline{O}$ and $\overline{P}$ refers to the observed and predicted average values respectively (Willmott et al., 1985).

The regression results between predicted and observed values showed reasonable value of correlation. The conductivity, DO, and TDS have shown the RMSE values of 133.07 mS/cm, 0.50 mg/l, 84.88 mg/l respectively. They have also shown a strong index of agreement of above 0.71 and R^2 values of 0.79, 0.52, and 0.76 respectively. The RMSE of conductivity and TDS are higher than DO even after having the strong R^2 values and higher NSCE and D values. The RMSE values also depend on the parameter range values. The MAE and MBE of the conductivity parameter (underestimation if MBE < 0, and overestimation if MBE > 0) are 91.43 and -2.76 respectively. The DO had a low average error of 0.4, a good MBE, NSCE values of 0.04, 0.51. Furthermore, it has a low RMSE of 0.50 and a reasonably high index of agreement of 0.71, indicating that MPN prediction is effective for DO.

The results of ANNs using MATLAB environment are shown the Table 3. In MATLAB environment, the ANNs are trained and tested using Levenberg-Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient Algorithm.

Table 3 R values of ANN model with Levenberg-Marquardt training algorithm

S.NO	Parameter	Training	Validation	Testing	All	Network architecture
1	Conductivity (ms/cm)	0.94	0.95	0.93	0.94	8-17-1
2	DO (mg/L)	0.76	0.72	0.78	0.75	8-17-1
3	TDS (mg/L)	0.92	0.95	0.96	0.92	8-17-1

The Levenberg-Marquardt training function gives the best results when the correlation coefficients of all the three algorithms are compared. The testing of conductivity, DO, and TDS gives R values: 0.93, 0.78, and 0.96. We obtain an overall R values of 0.94, 0.75, and 0.92 for conductivity, DO, and TDS. All of the training, testing and validation values are more than 0.72. The regression model outputs are shown in Figure 2, 3, and 4 for all the three water quality parameters considered. Only one hidden layer consisting of 17 hidden nodes was considered after several trial and error attempts of fixing network architecture. This is keeping in view of the result of Kolmogorov mapping theorem (Hecht-Nielsen, 1987). Both the attempts of MATLAB and SPSS predicted almost same R value for DO with different network model architectures. The predictions using Levenberg-Marquardt are better than Scaled Conjugate Gradient algorithm which can be seen from MATLAB and SPSS results.

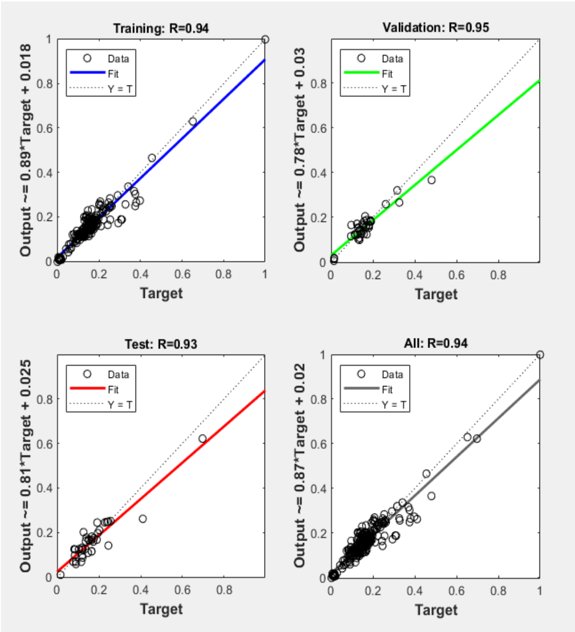

Figure 2. Training, Validation, Testing, Overall for Conductivity

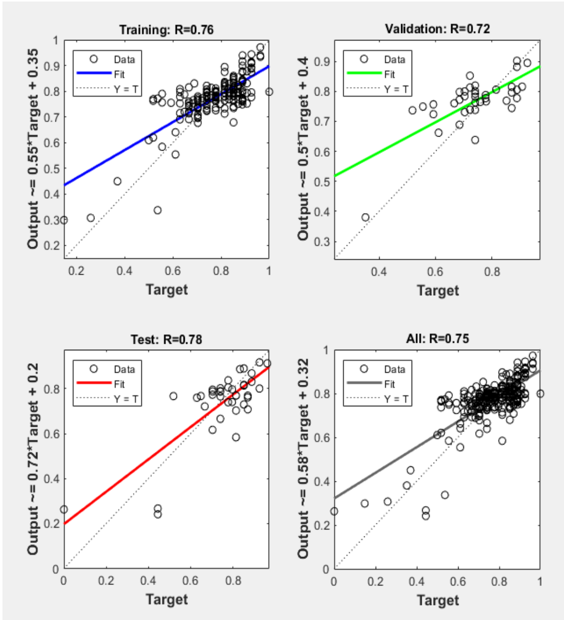

Figure 3. Training, Validation, Testing, Overall for DO

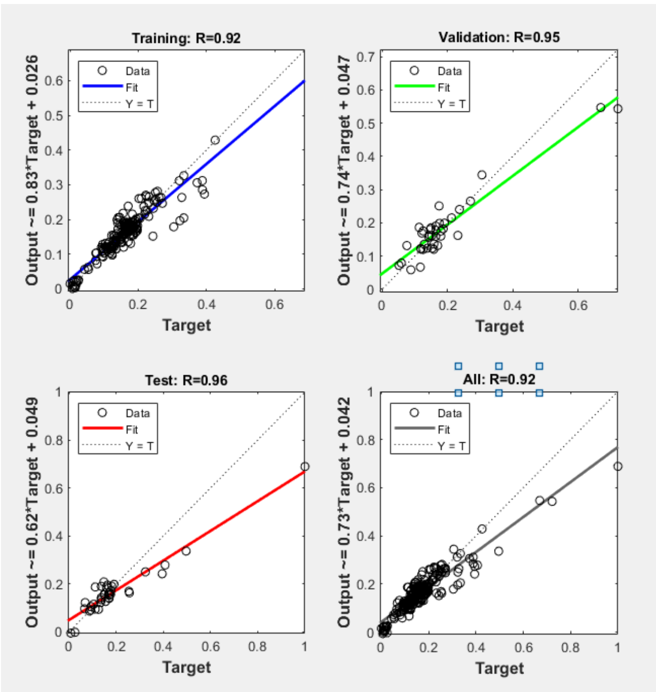

Figure 4. Training, Validation, Testing, Overall for TDS

4. Conclusions

The following conclusions are derived from the above results and discussion.

(i) The spatial dataset of LULC map and DEM data were used to correlate the field derived water quality parameters to comprehensively address LULC impacts on watershed basis.

(ii) The Levenberg-Marquardt training algorithm gives the best results in terms of R^2 compared to Bayesian Regularization and Scaled Conjugate Gradient algorithm in MATLAB environment for prediction of the three water quality parameters.

(iii) The simulations of MPN of SPSS with a single hidden layer consisting of 5 to 8 hidden nodes gives reasonable results compared to MATLAB results for the three parameters. Though the data variability is less for DO, the scaled conjugate gradient learning algorithm of MPN of SPSS predicts with reasonable accuracy and a low MBE for DO.

Acknowledgements

The authors would like to thank Council of Scientific and Industrial Research (CSIR), India grant (No. 24(0356)/19/EMR-II) of the project titled ‘Experimental and Computational studies of Surface Water Quality parameters from Morphometry and Spectral Characteristics’. The authors would also like to thank TSPCB for providing us with the required data.

References

Allan, D., Erickson, D., & Fay, J. (1997). The influence of catchment land use on stream integrity across multiple spatial scales. *Freshwater Biology*, *37*(1), 149–161. https://doi.org/10.1046/j.1365-2427.1997.d01-546.x

Amiri, B. J., & Nakane, K. (2009). Modeling the Linkage Between River Water Quality and Landscape Metrics in the Chugoku District of Japan. *Water Resources Management*, *23*(5), 931–956. https://doi.org/10.1007/s11269-008-9307-z

Anmala, J., Meier, O. W., Meier, A. J., & Grubbs, S. (2015). GIS and Artificial Neural Network–Based Water Quality Model for a Stream Network in the Upper Green River Basin, Kentucky, USA. *Journal of Environmental Engineering*, *141*(5), 04014082. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000801

Basnyat, P., Teeter, L. D., Flynn, K. M., & Lockaby, B. G. (1999). Relationships Between Landscape Characteristics and Nonpoint Source Pollution Inputs to Coastal Estuaries. *Environmental Management*, *23*(4), 539–549. https://doi.org/10.1007/s002679900208

Bolstad, P. V., & Swank, W. T. (1997). Cumulative Impacts Of Landuse On Water Quality In A Southern Appalachian Watershed. *Journal of the American Water Resources Association*, *33*(3), 519–533. https://doi.org/10.1111/j.1752-1688.1997.tb03529.x

Fu, Y., Zhao, Y., Zhang, Y., Guo, T., He, Z., & Chen, J. (2013). GIS and ANN-based spatial prediction of DOC in river networks: A case study in Dongjiang, Southern China. *Environmental Earth Sciences*, *68*(5), 1495–1505. https://doi.org/10.1007/s12665-012-2177-y

Gazzaz, N. M., Yusoff, M. K., Ramli, M. F., Juahir, H., & Aris, A. Z. (2015). Artificial Neural Network Modeling of the Water Quality Index Using Land Use Areas as Predictors. *Water Environment Research*, *87*(2), 99–112. https://doi.org/10.2175/106143014X14062131179276

Giri, S. (2021). Water quality prospective in Twenty First Century: Status of water quality in major river basins, contemporary strategies and impediments: A review. *Environmental Pollution*, *271*, 116332. https://doi.org/10.1016/j.envpol.2020.116332

Girija, T. R., Mahanta, C., & Chandramouli, V. (2007). Water Quality Assessment of an Untreated Effluent Impacted Urban Stream: The Bharalu Tributary of the Brahmaputra River, India. *Environ Monit Assess*, 16.

Goetz, S. J. (2006). Remote Sensing Of Riparian Buffers: Past Progress And Future Prospects. *Journal of the American Water Resources Association*, *42*(1), 133–143. https://doi.org/10.1111/j.1752-

1688.2006.tb03829.x

Hecht-Nielsen, R., Drive, O., & Diego, S. (1987). *Kolmogorov's Mapping Neural Network Existence Theorem*. 4.

Huang, J., Zhan, J., Yan, H., Wu, F., & Deng, X. (2013). Evaluation of the Impacts of Land Use on Water Quality: A Case Study in The Chaohu Lake Basin. *The Scientific World Journal*, *2013*, 1–7. https://doi.org/10.1155/2013/329187

Ishee, E. R., Ross, D. S., Garvey, K. M., Bourgault, R. R., & Ford, C. R. (2015). Phosphorus Characterization and Contribution from Eroding Streambank Soils of Vermont's Lake Champlain Basin. *Journal of Environmental Quality*, *44*(6), 1745–1753. https://doi.org/10.2134/jeq2015.02.0108

Jain, A., & Indurthy, S. K. V. P. (2003). Comparative Analysis of Event-based Rainfall-runoff Modeling Techniques—Deterministic, Statistical, and Artificial Neural Networks. *Journal of Hydrologic Engineering*, *8*(2), 93–98. https://doi.org/10.1061/(ASCE)1084-0699(2003)8:2(93)

Johnson, L. B., & Gage, S. H. (1997). Landscape approaches to the analysis of aquatic ecosystems. *Freshwater Biology*, 20.

Johnson, L., Richards, C., Host, G., & Arthur, J. (1997). Landscape influences on water chemistry in Midwestern stream ecosystems. *Freshwater Biology*, *37*(1), 193–208. https://doi.org/10.1046/j.1365-2427.1997.d01-539.x

Juahir, H., Zain, S. M., Toriman, M. E., Mokhtar, M., & Man, H. C. (2004). Application Of Artificial Neural Network Models For Predicting Water Quality Index. 14.

Keskin, T. E., Düğenci, M., & Kaçaroğlu, F. (2015). Prediction of water pollution sources using artificial neural networks in the study areas of Sivas, Karabük and Bartın (Turkey). *Environmental Earth Sciences*, *73*(9), 5333–5347. https://doi.org/10.1007/s12665-014-3784-6

Kumar, V., Sharma, A., Chawla, A., Bhardwaj, R., & Thukral, A. K. (2016). Water quality assessment of river Beas, India, using multivariate and remote sensing techniques. *Environmental Monitoring and Assessment*, *188*(3), 137. https://doi.org/10.1007/s10661-016-5141-6

Maier, H. R., & Dandy, G. C. (1996). The Use of Artificial Neural Networks for the Prediction of Water Quality Parameters. *Water Resources Research*, *32*(4), 1013–1022. https://doi.org/10.1029/96WR03529

Maillard, P., & Pinheiro Santos, N. A. (2008). A spatial-statistical approach for modeling the effect of non-point source pollution on different water quality parameters in the Velhas river watershed – Brazil. *Journal of Environmental Management*, *86*(1), 158–170. https://doi.org/10.1016/j.jenvman.2006.12.009

Narumalani, S., Zhou, Y., & Jensen, J. R. (1997). Application of remote sensing and geographic information systems to the delineation and analysis of riparian buffer zones. *Aquatic Botany*, *58*(3–4), 393–409. https://doi.org/10.1016/S0304-3770(97)00048-X

Pujar, P. M., Kenchannavar, H. H., Kulkarni, R. M., & Kulkarni, U. P. (2020). Real-time water quality monitoring through Internet of Things and ANOVA-based analysis: A case study on river Krishna. *Applied Water Science*, *10*(1), 22. https://doi.org/10.1007/s13201-019-1111-9

Roth, N. E., Allan, J. D., & Erickson, D. L. (1996). Landscape influences on stream biotic integrity assessed at multiple spatial scales. *Landscape Ecology*, *11*(3), 141–156. https://doi.org/10.1007/BF02447513

Singh, K. P., Malik, A., Mohan, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—A case study. *Water Research*, *38*(18), 3980–3992. https://doi.org/10.1016/j.watres.2004.06.011

Song, S., Zheng, X., & Li, F. (2010). Surface water quality forecasting based on ANN and GIS for the Chanzhi Reservoir, China. *The 2nd International Conference on Information Science and Engineering*, 4094–4097. https://doi.org/10.1109/ICISE.2010.5689328

Turuganti, V., Agrawal, R., Purna, B., & Anmala, J. (2019). Prediction of river water quality of upper green river watershed, kentucky, usa using cluster-based neural network models. 10.

Venkateswarlu, T., Anmala, J., & Dharwa, M. (2020). PCA, CCA, and ANN Modeling of Climate and Land-Use Effects on Stream Water Quality of Karst Watershed in Upper Green River, Kentucky. *Journal of Hydrologic Engineering*, *25*(6), 05020008. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001921

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'Donnell, J., & Rowe, C. M. (1985). Statistics for the evaluation and comparison of models. *Journal of Geophysical Research: Oceans*, *90*(C5), 8995–9005. https://doi.org/10.1029/JC090iC05p08995

Zeilhofer, P., Lima, E. B. N. R., & Lima, G. A. R. (2006). Spatial Patterns of Water Quality in the Cuiabá River Basin, Central Brazil. *Environmental Monitoring and Assessment*, *123*(1–3), 41–62. https://doi.org/10.1007/s10661-005-9114-4

Zhang, Q., & Stanley, S. J. (1997). Forecasting raw-water quality parameters for the North Saskatchewan River by neural network modeling. *Water Research*, *31*(9), 2340–2350. https://doi.org/10.1016/S0043-1354(97)00072-9