Journal of Water and Soil
Vol. 29, No. 5, Nov.-Dec. 2015, p. 1262-1277

نشریه آب و خاک (علوم و صنایع کشاورزی)
جلد ۲۹، شماره ۵، آذر – دی ۱۳۹۴، ص. ۱۲۷۷-۱۲۶۲

Estimation of Rivers Dissolved Solids TDS by Soft Computing (Case Study: Upstream of Boukan Dam)

S. Zaman Zad Ghavidel[1] - K. Zeinalzadeh[2]*

Received: 11-23-2014
Accepted: 05-25-2015

Introduction: A total dissolved solid (TDS) is an important indicator for water quality assessment. Since the composition of mineral salts and discharge affects the TDS of water, it is important to understand the relationship of mineral salts composition with TDS.

Materials and Methods: In this study, methods of artificial neural networks with Levenberg-Marquardt training algorithm, adaptive neuro fuzzy inference system based on Subtractive Clustering and Gene expression programming were used to model water quality properties of Zarrineh River Basin at upstream of Boukan dam, to be developed in total dissolved solids prediction. ANN and ANFIS programs code were written using MATLAB programming language. Here, the ANN with one hidden layer was used and the hidden nodes' number was determined using trial and error. Different activation functions (logarithm sigmoid, tangent sigmoid and linear) were tried for the hidden and output nodes and the GeneXpro Tools 4.0 were used to obtain the equation of the best models. Therefore, water quality data from two hydrometer stations, namely Anyan and Safakhaneh hydrometer stations were used during the statistical period of 18 years (1389-1372). In this research, for selecting input variables to the data driven models the stepwise regression method was used. In the application, 75% of data set were used for training and the remaining, 25% of data set were used for testing, randomly. In this paper, three statistical evaluation criteria, correlation coefficient (R), the root mean square error (RMSE) and mean absolute error (MAE), were used to assess model's performances.

Results and Discussion: By applying stepwise method, the first significant (at 95% level) variable entered to the model was the HCO_3. The second variable that entered to the model was Ca. The third and fourth ones were Na and Q respectively. Mg was finally entered to the model. The optimal ANN architecture used in this study consists of an input layer with five inputs, one hidden and output layer with three and two neurons for Anyan and Safakhaneh hydrometer stations, respectively. Similar ANN, ANFIS-SC5 model had the best performance. It is clear that the ANFIS with 0/4 and 0/7 radii value has the highest R and the lowest RMSE for Anyan and Safakhaneh hydrometer stations, respectively. Various GEP models have been developed using the input combinations similar ANN and ANFIS models. Comparing the GEP5 estimations with the measured data for the test stage demonstrates a high generalization capacity of the model, with relatively low error and high correlation. From the scatter plots it is obviously seen that the GEP5 predictions are closer to the corresponding measured TDS than other models. As seen from the best straight line equations (assume the equation as $y=ax$) in the scatter plots that the a coefficient for GEP5 is closer to 1 than other models. In addition to previous operation, Gene expression programming offered mathematical relationships in the stations of Anyan and Safakhane with the correlation coefficients, respectively 0.962 , 0.971 and with Root-mean-square errors, respectively 12.82 , 29.08 in order to predict dissolved solids (TDS) in the rivers located at upstream of the dam. The obtained results showed the efficiency of the applied models in simulating the nonlinear behavior of TDS variations in terms of performance indices. Overall, the GEP model outperformed the other models. For all of applied models, the best result was obtained by application of input combination (5) including HCO_3, Ca, Na, Q and Mg. The results are also tested by using t test for verifying the robustness of the models at 95% significance level .Comparison results indicated that the poorest model in TDS simulation was ANN especially in test period. The observed relationship between residuals and model computed TDS values shows complete independence and random distribution. It is further supported by the respective correlations for GEP5 models (R^2 = 0.0011 for Anyan station and R^2 = 0.0123 for safakhaneh station) which are negligible small. Plots of the residuals versus model computed values can be more informative regarding model fitting to a data set. If the residuals appear to behave randomly it suggests that the model fits the data well. On the other hand, if non- random distribution is evident in the residuals, the model does not fit the data adequately. On the base of these results, we propose

1, 2- PhD Candidate and Assistant Professor, Department of Water Engineering, Urmia University
(* - Corresponding Author Email: k.zeinalzadeh@urmia.ac.ir)

GEP, ANFIS-SC and ANN methods as effective tools for the computation of total dissolved solids in river water, respectively.

Conclusion: It can be concluded that the ANN, ANFIS-GP, ANFIS-SC and GEP models can be considered as promising tools for forecasting TDS values, based on water quality parameters. It is notable from the results that the prediction accuracy of all applied models increases by increasing the number of input combinations. With attention to the aim of current research that is presenting the feasibility of artificial intelligence techniques for modeling TDS values, it is notable that the results presented in this paper are for research purpose and applying the abstained results for real-world needs some complicated steps and building artificial intelligences methods, based on complete data and parameters maybe affected the TDS values.

Keywords: Gene Expression, Dissolved Solids, Zarrineh River, Artificial Neural Networks