World Applied Sciences Journal
14 (Exploring Pathways to Sustainable Living in Malaysia: Solving the Current Environmental Issues): 60-65, 2011
ISSN 1818-4952; © IDOSI Publications, 2011

Sensitivity Analysis for Water Quality Index (WQI) Prediction for Kinta River, Malaysia

[1,3]Munirah Abdul Zali, [2]Ananthy Retnam, [3]Hafizan Juahir, [4]Sharifuddin M. Zain,
[5]Mohd Fadhil Kasim, [6]Baharuddin Abdullah and [5]Syaiful Bahren Saadudin

[1]Department of Chemistry of Malaysia, Jalan Sultan, 46661, Petaling Jaya, Selangor, Malaysia
[2]Centre of Excellence for Environmental Forensic,
[3]Department of Environmental Sciences,
Faculty of Environmental Studies, Universiti Putra Malaysia, 43400 Serdang, Malaysia
[4]Department of Chemistry, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
[5]Water Quality and Environmental Research Division, National Hydraulic Research
Institute of Malaysia (NAHRIM), Ministry of Natural Resources and Environment,
5377, Jalan Putra Permai, 43300 Seri Kembangan, Selangor, Malaysia
[6]Department of Irrigation and Drainage,
District Kinta and Batang Padang, Jalan Bendahara, 31650 Ipoh, Perak, Malaysia

Abstract: Water quality index (WQI) serves as the basis for environment assessment of watercourse in relation to pollution load categorization and designation of classes and beneficial uses as provided by Interim National Water Quality Standards (INWQS) in Malaysia. This index is calculated based on six parameters DO, BOD, COD, pH, NH_3-NL and SS. This research was need as it will give the preliminary judgement on the importance of each water quality parameter for WQI calculation at the Kinta River, Malaysia. This study revealed the used of sensitivity analysis based on ANN to evaluate the significant of each parameter for WQI determination. Sensitivity analysis was carried out for seven models (ANN-WQI-AP, ANN-WQI-LDO, ANN-WQI-LBOD, ANN-WQI-LCOD, ANN-WQI-LpH and ANN-WQI-LNH_3-NL) and a model performance criterion (R^2, RMSE and SSE) was used for model performance evaluation. DO, SS and NH_3-NL were selected as the best input models for WQI prediction. The ANN-WQI-LDO, ANN-WQI-LSS and ANN-WQI-LNH_3-NL model have R^2 values of 0.8301, 0.9265 and 0.9369 respectively; RMSE values of 4.888, 3.214 and 2.978 respectively; SSE values of 3106.534, 1343.286 and 1152.902 respectively. The low R^2 values and higher RMSE and SSE value compared to the ANN-WQI-AP model suggest the importance of these three parameters significantly affect the fitness and residual measurement of the ANN models in WQI prediction. The result also suggests that water quality of Kinta River was affected by agricultural activities and vicinity animal farm. Moreover the use of less parameter for WQI is much more applicable for our water resource management since its time and cost consuming.

Key words: Artificial Neural Network · Water Quality Index · Sensitivity Analysis · Water Quality · River Pollution · Kinta River

INTRODUCTION

Water quality become continuous concerning problems even its purposes other than human water supply. Water quality is the surface part started to deteriorated as waste water or uncontrolled inlets being discharged to the surface water and receiving ground. Even water quality is responsible to controlling health and the state of disease for our flora and fauna [1]. Human activities are a major factor in determining the quality of our water bodies through municipal and industrial wastewater discharge, eroded soils and land use, atmospheric pollution.

In Malaysia, major pollution is coming from domestic waste, industrial effluents, land clearance with suspended solid (SS) as the major source contributing up to 42% to

Corresponding Author: Hafizan Juahir, Department of Environmental Science, Faculty of Environmental Studies, University Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
Tel: +603-89467460, Fax: +603-8943 8109.

poorly planned land development, 30% from biological oxygen demand (BOD) due to industrial waste and 28% from ammoniacal nitrogen (NH_3-NL) attributed from domestic sewage disposal and animal farming activities [2]. For this reason, a mathematical instrument used to transform physico-chemical water characterization data into a single number, which represents the water quality level called water quality index (WQI). The WQI serves as the basis for environment assessment of a watercourse in relation to pollution load categorization and designation of classes of beneficial uses as provided for under the National Water Quality Standards for Malaysia (INWQS). The Department of Environment (DOE) of Malaysia used WQI to evaluate the status of the river water quality. Pollution status was estimated using WQI range and water quality classes were evaluated using values of six water quality parameter for WQI. The six parameters are dissolved oxygen (DO), BOD, chemical oxygen demand (COD), pH, NH_3-NL and SS. The values of 81-100, 60-80 and 0-59 were classified as clean, slightly polluted and polluted respectively.

However, the influence of these parameters on the surface water quality and subsequently in calculating WQI is questionable and very much dependent on the land use of the region [3]. Study has showed that certain parameters such as pH and COD have less contribution on WQI for Langat River basin [4] thus, the prediction of WQI eliminating these parameters exhibit best predictive performance. As such there is a need to evaluate the influence of each of the WQI parameters on the WQI itself for different region or different river basin. Instead of this there is no reported study on influence of WQI parameters on WQI for Kinta River basin.

The purpose of this study is to evaluate the contribution of each parameter used in calculation of WQI on the prediction of WQI. For this purpose Artificial Neural Network (ANN) sensitivity test with leave-one-out technique will be employed. Sensitivity analysis is tool for ranking the importance of the model-input variables by assessing their contribution to the variability of the model output [5]. The use of ANN was a widely applied in many hydrological integrated researches [6-8] as it capability in noisy data when the underlying physical and biological relationship was not fully understood. The extensive application owing to reliability on non-linear data set [9] and able to determine the best model input and network structure in optimization of the desired model. ANN also has advantages as its function is to stimulate the functionality and decision-making process of the human brain.

MATERIALS AND METHODS

The Study Area: The Kinta River flow for approximately 100 km length and it is located in the central-eastern section of Perak State. The topography of the catchment consists of steep forest-covered mountains and hills in the north and east, which pass through to the Kinta Valley to the south of Ipoh. Land use of the Kinta valley consists of agriculture (e.g. rubber, oil palm and fruit trees), urban development and unproductive examining land, including tailings and ponds [10]. The major tributary of Kinta River from the northwest is the Pari River and others tributaries from the steeper eastern catchment include the Raia River and Kampar River, which join the Kinta River at Tg Tualang.

Data Set and Sensitivity Analysis Technique: The water quality data set from Kinta River was chosen in this study as its locality changes in the past years may describe the effectiveness of this index. Data set of Kinta River was obtained from regular monitoring programme by DOE of Malaysia from year 2002 to 2006. Six parameters (from WQI calculation) were selected as input selection and WQI as the output for all model developed. The first model was run using all parameters as input variables and named as artificial neural network-water quality index- all parameter (ANN-WQI-AP) which serves as a reference model. In order to evaluate the importance of the input parameters of ANN-WQI-AP sensitivity analysis was carried out by excluding one parameter from six all parameters and ANN performance model was evaluated using correlation of coefficient (R^2), root mean square error (RMSE) and sum squares error (SSE).

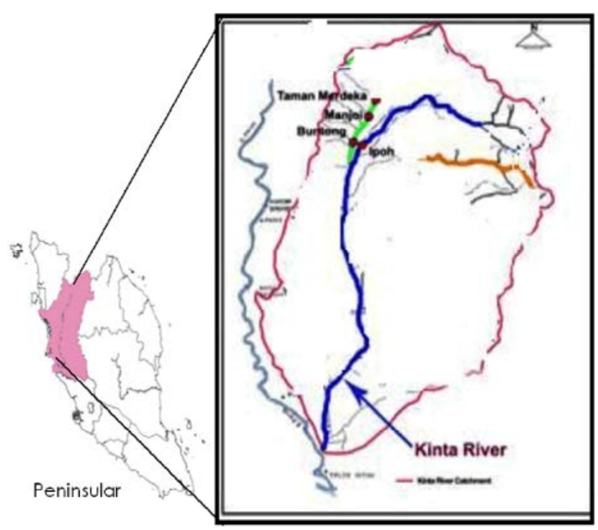

Fig. 1: Kinta River, Malaysia

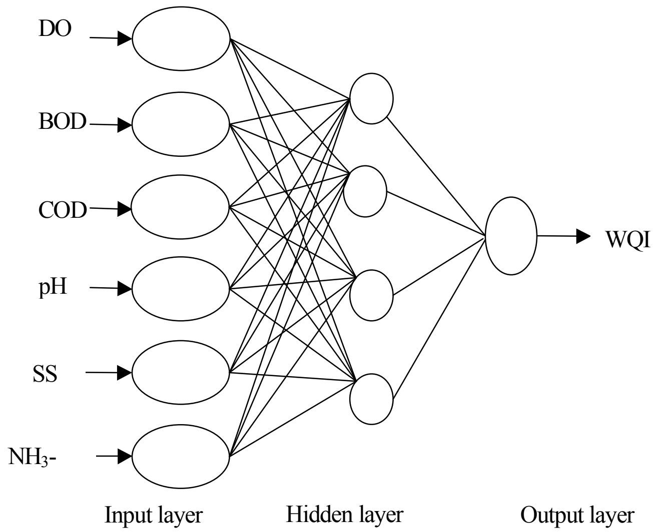

Fig. 2: ANN architecture for ANN-WQI-AP model

Moreover sensitivity analysis was very useful and reliable when sufficient data available and to assess the relative importance of the parameter [11]. The second model was developed named as artificial neural network-water quality-leave DO (ANN-WQI-LDO) which means DO is excluded in forecasting the WQI values. The third model is artificial neural network-water quality index-leave BOD (ANN-WQI-LBOD); the forth model is artificial neural network-water quality index-leave COD (ANN-WQI-LCOD); the fifth model is artificial neural network-water quality index-leave pH (ANN-WQI-LpH); the sixth model is artificial neural network-water quality index-leave SS (ANN-WQI-LSS) and the seventh is artificial neural network-water quality index-leave NH_3-NL (ANN-WQI-LNH_3-NL). Figure 2 show the ANN-WQI-AP model architecture with six inputs variable, 4 hidden nodes and WQI as output variables. A total of 135 observations from year 2002 and 2006 were selected as data set and all models were run using JMP8 software.

Determination of Best Input Selection for WQI Prediction: A multi criteria approach was used for estimating the goodness of the seven models developed subsequently determining the best input selection for WQI prediction. The model performance was evaluated using goodness-of-fit measures and statistical error including R^2, RMSE and SSE. The R^2 value provides an indication of the similarity on the actual WQI values to the predicted WQI values. R^2 provides the variability measure of the data reproduced in the model and fitness of the model but not on how well it perform when applying the unknown data set [12]. Thus, model evaluation also was perform using values of RMSE as it measures the residual error which give the information of the difference between observed and modelled values [13] and minimum SSE values indicating the good forecasting accuracy of the models [14].

RESULT AND DISCUSSION

The selection of input parameter is very vital in obtaining the effective neural network modelling. Table 1 show the overall result of seven ANN-WQI models developed for sensitivity analysis. The model ANN-WQI-AP was used as reference to others models developed. Figure 3 show the plotted graph of actual value of WQI versus predicted value of WQI. The lowest value of RMSE and SSE signify that the high model robustness with R^2 value of 0.9818 for both training and testing sets with 98.18% of WQI variability explained by the six parameters used for WQI prediction.

ANN-WQI-AP model show goodness of accuracy and exhibit minimum residual errors compared to other model as this model has the lowest RMSE value. A slight reduction of R^2 value was noticed when excluding DO parameter in WQI predication. This revealed that DO display the most significant parameter for WQI forecasting. The highest values of RMSE and SSE of ANN-WQI-LDO also suggest that the model fitness was decreased and high residual error occurred. ANN-WQI-LSS and ANN-WQI-LNH_3-NL models also indicating the significant of SS and NH_3-NL parameter in WQI numerical modelling as both of the model have high RMSE and SSE

Table 1: Result of sensitivity analysis for WQI prediction

Model	R^2	RMSE	SSE	Equation
ANN-WQI-AP	0.9818	1.598962	332.3681	WQI= -1.45*DO -0.52*BOD + 0.30*COD -0.11*SS -0.12*pH -0.09*NH3-NL
ANN-WQI-LDO	0.8301	4.888396	3106.5340	WQI= -1.55B*BOD + 2.58*COD -0.42*SS -1.18*pH +2.16*NH3-NL
ANN-WQI-LBOD	0.9678	2.129072	589.2833	WQI= 0.35*DO -1.23*COD +2.43*SS +0.28*pH +0.77*NH3-NL
ANN-WQI-LCOD	0.9674	2.140067	595.3855	WQI= 0.61*DO -0.78*BOD -0.25*SS -0.07*pH -1.24*NH3-NL
ANN-WQI-LSS	0.9265	3.214494	1343.2860	WQI= -0.19*DO -0.32*BOD + 0.62*COD +0.37*pH +2.40*NH3-NL
ANN-WQI-LpH	0.9831	1.542089	309.1450	WQI= 0.03*DO -0.08*BOD + 0.38*COD -3.16*SS -0.25*NH3-NL
ANN-WQI-LNH3-NL	0.9369	2.977999	1152.9020	WQI= 0.37*DO -0.95*BOD -0.42*COD -2.68*SS -0.30*pH

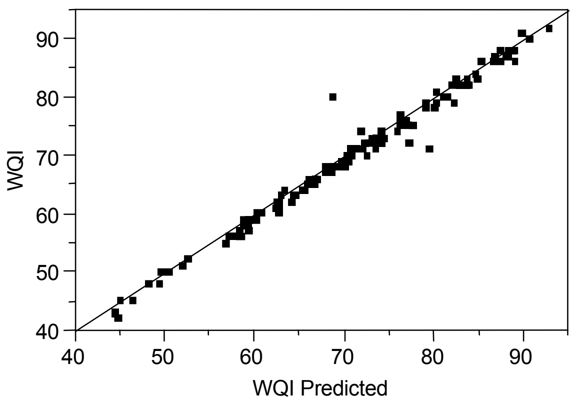

Fig. 3: Fitness of ANN-WQI-AP model

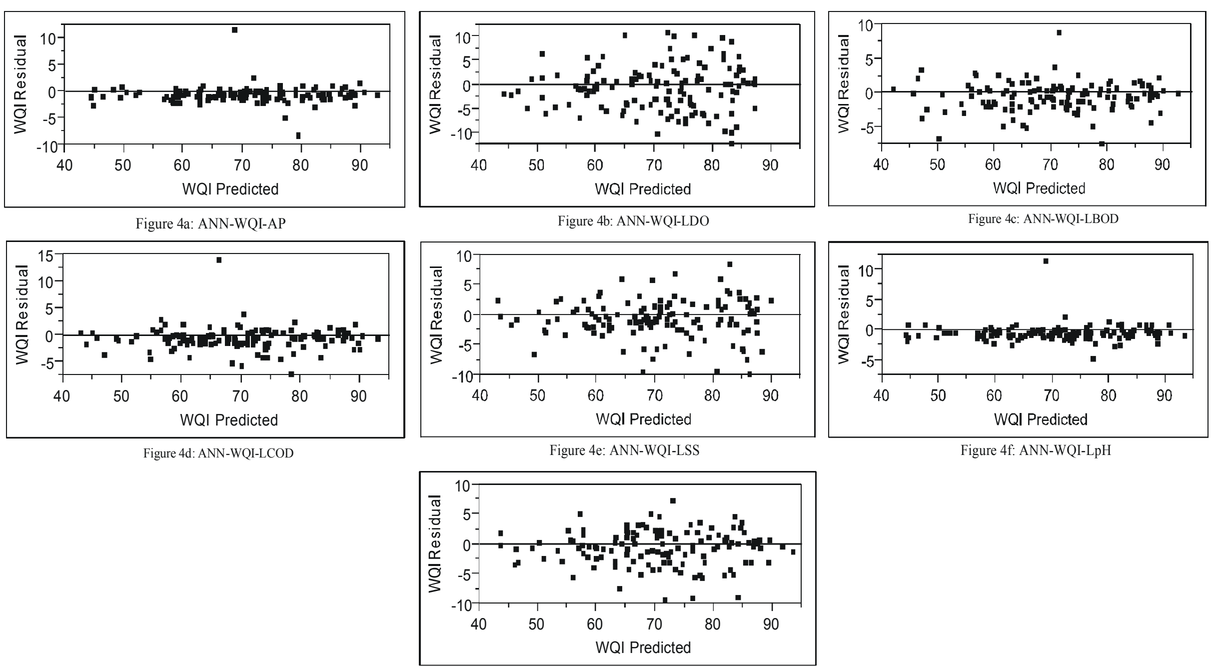

Figure 4a: ANN-WQI-AP

Figure 4b: ANN-WQI-LDO

Figure 4c: ANN-WQI-LBOD

Figure 4d: ANN-WQI-LCOD

Figure 4e: ANN-WQI-LSS

Figure 4f: ANN-WQI-LpH

Figure 4g: ANN-WQI-LNH$_3$-NL

Fig. 4a-4g: Residual error of the seven models developed for WQI estimation based on sensitivity analysis

values compared to the others model. This again suggests the importance of SS and NH_3-NL as significant input for WQI prediction.

Figure 4a- 4g show the image of distribution of residual error of the seven models developed. The highly distributed residuals difference variation was observed for ANN-WQI-LDO, ANN-WQI-LSS and ANN-WQI-LNH$_3$-NL models with few data points for each model are out of the 50% error band[11]. Model of ANN-WQI-LBOD, ANN-WQI-LCOD and ANN-WQI-LpH demonstrate the less residual error as the residual values distributions were near to 0 lines. The significant of DO, SS and NH_3-NL parameters may attribute to the agricultural non-point source and vicinity animal farm [15] along the river as well as the land development that may affect the surface water quality. Therefore good model reproducibility may result by using DO, SS and NH_3-NL as input parameters for WQI prediction as these parameters correlated to the locality of Kinta River. The less significant input may eliminated as this parameter contribute less variance for WQI prediction.

CONCLUSION

The potential of sensitivity analysis coupled with ANN application provide WQI estimation with good judgement by using various model performance evaluations. R^2, RMSE and SSE values was successfully describe all the models developed; and DO, SS and NH_3-NL were identified as the three significant variables for WQI estimation at Kinta River. This study was performed to reduce the less significant parameter for WQI calculation using ANN model. BOD and COD parameter can be excluded as these parameters were time and cost consuming and have minimum correlation for WQI forecasting. Moreover the less reliable parameter which is pH may need to justify for WQI calculation as this parameter give minimum variability for WQI estimation. Based on the comparison result, the sensitivity analysis based on ANN was performing well suited. However the model require occasionally monitoring programmes as the land used of Kinta River may affect the effectiveness of the model. This model also was potentially applied at others rivers that having similar water quality trends and land-used for water quality related agencies in Malaysia as it will help in water quality monitoring programme towards better environmental management in Malaysia.

ACKNOELDGEMENT

The author would like to thank to Department of Environment (DOE) for providing the data from the river monitoring programme in Malaysia, University Putra Malaysia, National Hydraulic Research Institute of Malaysia (NAHRIM) and Department of Irrigation and Drainage (DID), District Kinta and Batang Padang, Perak for their contributions in this project.

REFERENCES

1. Kazi, T.G., M.B. Arain, M.K. Jamali, N. Jalbani, H.I. Afridi, R.A. Sarfraz, J.A. Baig and G. Abdul Shah, 2009. Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. Ecotoxicology and Environmental Safety, 72: 301-309.
2. Juahir, H., S.M. Zain, M.K. Yusoff, T.I.T. Hanidza, A.S.M. Armi, M.E. Toriman and M. Mokhtar, 2010. Spatial water quality assessment of Langat River Basin (Malaysia) using envirometric techniques. Environmental Monitoring Assessment.
3. Juahir, H., S.M. Zain, A.A. Aris, M.K. Yusoff and M. Mokhtar, 2010. Spatial assessment of Langat River water quality using chemometrics. J. Environmental Monitoring, 12: 287-295.
4. Juahir, H., S.M. Zain, M.E. Toriman and M. Mokhtar, 2004. Application of Artificial Neural Network Models for Predicting Water Quality Index. Jurnal Kejuruteraan Awam, 16(2): 42-55.
5. Manache, G. and C.S. Melching, 2008. Identification of reliable regression- and correlation-based sensitivity measures for importance ranking of water quality model parameters. Environmental Modelling and Software, 23: 549-562.
6. Garcia, L.A.A.S.A., 2006. Using neural networks for parameter estimation in ground water. J. Hydrol., 318: 215-231.
7. Govindaraju, R.S. and A. Ramachandra Rao, 2000. Arti?cial Neural Networks in Hydrology. Kluwer Academic Publishing, The Netherlands.
8. Palani, S., S.Y. Liong and P. Tkalich, 2008. An ANN application for water quality forecasting. Marine Pollution Bulletin, 56: 1586-1597.
9. Bandyopadhyay, G. and S. Chattopadhyay, 2007. Single layer artificial neural network models versus multiplelinear regressio model in forecasting the time series of total ozone. International Journal Environmental Sci. Technol., 4(1): 141-149.
10. Azamathulla, H.M., C.K. Chang, A. Ab Ghani, J. Ariffin, N.A. Zakaria and Z. Abu Hassan, 2009. An ANFIS-based approach for predicting the bed load for moderately sized rivers. J. Hydro-environment Res., 3: 25-44.
11. Dogan, E., B. Sengorur and R. Koklu, 2009. Modelling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. J. Environmental Management, 90: 1229-1235.

12. Aertsen, W., V. Kint, J.V. Orshoven, K. Ozkan and B. Muys, 2010. Comparison and ranking of different modelling techniques for prediction of site index in mediterranean mountain forests. Ecological Modelling, 221(): 1119-1130.
13. Sousa, S.I.V., F.G. Martins, M.C.M. Alvim-Ferraz and M.C. Pereira, 2007. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentration. Environmental Modelling and Software, 22: 97-103.
14. Kisi, O., 2005. Daily river flow forecasting using artificial neural networks and auto-regressive models. Turkish J. Eng. Env. Sci., 29: 9-20.
15. Crowther, J., D. Kay and M.D. Wyer, 2001. Relationships between microbial water quality and environmental conditions in coastal recreational waters: The Flyde Coast, UK. Water Res., 35(17): 4029-4038.