2016-10-26

DISSOLVED OXYGEN MODELLING USING ARTIFICIAL NEURAL NETWORK: A CASE OF RIVER NZOIA, LAKE VICTORIA BASIN, KENYA

Kanda, Edwin

Journal of Water Security

eISSN 2345-0363
Journal of Water Security, 2016, Vol. 2
Article Number: jws2016004
DOI: http://dx.doi.org/10.15544/jws.2016.004

DISSOLVED OXYGEN MODELLING USING ARTIFICIAL NEURAL NETWORK: A CASE OF RIVER NZOIA, LAKE VICTORIA BASIN, KENYA

Edwin Kimutai Kanda[a,], Emmanuel Chessum Kipkorir[b], Job Rotich Kosgei[b]

[a,]*Department of Civil and Structural Engineering, Masinde Muliro University of Science and Technology, P.O Box 190 – 50100, Kakamega, Kenya. E-mail: kandaedwin@gmail.com*

[b]*Department of Civil and Structural Engineering, Moi University*

Submitted 30 August 2016; accepted 26 October 2016

Abstract. River Nzoia in Kenya, due to its role in transporting industrial and municipal wastes in addition to agricultural runoff to Lake Victoria, is vulnerable to pollution. Dissolved oxygen is one of the most important indicators of water pollution. Artificial neural network (ANN) has gained popularity in water quality forecasting. This study aimed at assessing the ability of ANN to predict dissolved oxygen using four input variables of temperature, turbidity, pH and electrical conductivity. Multilayer perceptron network architecture was used in this study. The data consisted of 113 monthly values for the input variables and output variable from 2009–2013 which were split into training and testing datasets. The results obtained during training and testing were satisfactory with R^2 varying from 0.79 to 0.94 and RMSE values ranging from 0.34 to 0.64 mg/l which imply that ANN can be used as a monitoring tool in the prediction of dissolved oxygen for River Nzoia considering the non-correlational relationship of the input and output variables. The dissolved oxygen values follow seasonal trend with low values during dry periods.

Keywords: a*rtificial intelligence technique, feed-forward propagation, pollution monitoring, water quality*.

Introduction

River Nzoia is the largest river in the Kenyan portion of Lake Victoria basin. The river receives both point and non-point sources of pollution from agricultural runoff and other diffuse sources, agro-based industries and municipal wastes. The major industries include Nzoia and Mumias sugar factories and Pan African paper mills. These industries contribute to the deterioration of water quality in River Nzoia (Abira, 2008; Kanda *et al*., 2015).

Dissolved oxygen (DO) is one of the most important parameters in surface waters as it indicates the level of pollution. River Nzoia had low DO levels due to effluent discharges from Mumias and Pan African paper mills especially during low flows when dilution effect was minimal (Kanda *et al*., 2015). The concentration of dissolved oxygen reflects the equilibrium, or its lack of, between oxygen producing processes and oxygen consuming processes and depends on many factors such as temperature, salinity, oxygen depletion, sources of oxygen and other water quality parameters (Ahmed, 2014).

Various models have been used in the simulation and prediction of dissolved oxygen in surface waters. Despite its relevance in decision making, few studies have been conducted on the use of models to predict water quality in River Nzoia. Most modelling studies in Nzoia catchment have concentrated on flood prediction such as Nyadawa *et al*. (2010) using Geospatial Stream Flow Model (Geo-SFM) and Ngaina *et al*. (2014), using Probability Distributed Moisture (PDM) model.

The limited water quality modelling studies in Nzoia catchment could be attributed to unavailability or limited data which are required for detailed water quality modelling studies (Kanda, 2014). A study by Kanda *et al*. (2015), found MIKE 11 model to be satisfactory in simulating DO and biochemical oxygen demand of River Nzoia. However, most water quality models have more calibration parameters which make their use difficult in situations like Kenya where water quality data is unavailable or limited for most of the rivers as according to Najah *et al*. (2011), limited water quality data and the high cost of water quality monitoring often pose serious problems for process-based modelling approaches. Moreover, according to Chau (2006), many water models require knowledge on model manipulation through real physical observations, mathematical description of water movement or water quality, discretization of governing equations, solution schemes for the equations and output analysis which many modellers do not possess thereby possibly producing inferior design and cause under-utilization, or even total failure, of these models.

Artificial intelligence is a technique with a flexible mathematical structure that is capable of identifying complex non-linear relationships between input and output data when compared with other classical modelling techniques (Najah *et al*., 2011). Chau (2006), reviewed the integration of artificial intelligence technologies (knowledge–based system, genetic algorithm, artificial neural network, and fuzzy inference system) into water quality modelling. In recent years, Artificial neural networks (ANN) have found a number of applications in the area of water quality modelling (Khalil *et al*., 2012) mostly in the fields of water quality prediction (Kisi, Murat, 2011). According to Chau (2006), the greatest advantage of ANN's over other modelling techniques is their capability to model complex, non-linear processes without having to assume the form of the relationship between input and output variables. ANN can be used to predict water quality variables in areas where there are missing values and thus help in pollution monitoring (Diamantopoulou *et al*., 2005). Pollution monitoring in developing countries is hindered by unavailability of funds and thus there is need for the development of simple models which use few and easily available inputs.

This study aimed at predicting DO using easily available data which included pH, turbidity, temperature and electrical conductivity (EC) with the help of ANN. The choice of these input variables was based on their availability and thus could be relevant in pollution monitoring and control, and also based on the relationship among the inputs and DO. Turbidity is an indicator of stream pollution and also has an influence on DO possibly due to increased light absorbency which lead to increased temperature levels (Emamgholizadeh *et al*., 2014). Csábrági *et al*. (2015) found pH as an effective parameter in the prediction of DO in the Hungarian section of River Danube while temperature had the highest impact in the Serbian section (Antanasijević *et al*., 2014). EC is an indicator of the dissolved solids in the stream which can be in the form of nitrates, ammonium, among others which are indicators of water pollution (Prathumratana *et al*., 2008). The land use of Nzoia catchment is mainly agricultural with intensive application of fertilizers for sugarcane production (Omwoma *et al*., 2012) and thus these nutrients can leach to the stream as nitrates or phosphorus which can be sources of dissolved ions as detected by EC. Kisi and Murat (2011), found that a combination of discharge, temperature, pH and EC gave good predictions of DO. These studies, therefore, suggest that the input parameters of temperature, pH, turbidity and EC could be used in the prediction of water quality parameters in rivers threatened with pollution.

Meterials and Methods

Study Area and Datasets

River Nzoia is in the western region of Kenya in the Lake Victoria basin. It has a catchment area of approximately 12,900 km^2 and a length of 334 km up to its outfall at Lake Victoria. It lies within the south-eastern part of Mt Elgon and the western slopes of the Cherangani Hills. The study area is the middle section of River Nzoia between Webuye and Mumias towns (a distance of about 55 km). It lies between latitudes 00º 35.157' N and 00º 22.165' N and longitudes 34º 48.411' E and 34º 28.962' E (Fig. 1).The elevation decreases gradually from 1459 m at Webuye to about 1297 m above sea level at Mumias town with an average slope of about 0.29%. The study area has a catchment of 1,942 km^2.

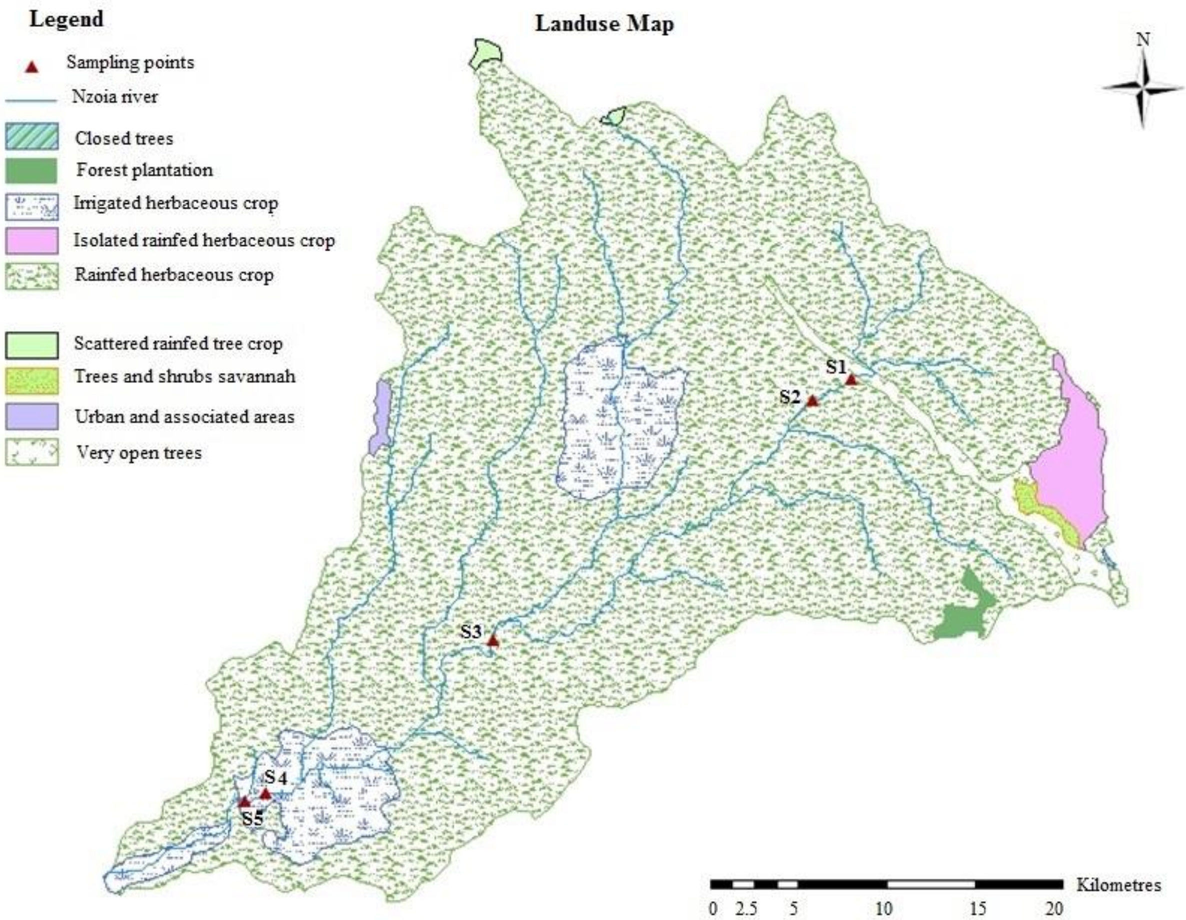

Fig. 1. Map of study area showing sampling points and land use

The predominant type of land use (Fig.1) is rain-fed herbaceous crops mainly sugarcane plantations owned by the out-growers of the two major sugar industries of Mumias and Nzoia Sugar. The irrigated herbaceous crops are found around the sugar factories forming the nucleus of the sugarcane farms owned by the sugar factories. The river section receives effluents from treated municipal wastewater from Mumias, Bungoma and Webuye towns. It also receives treated industrial wastewater from Pan African Paper mills in Webuye, Nzoia Sugar Company in Bungoma and Mumias Sugar Company in Mumias town.

The monthly data which was used in the study were from 2009 to 2013 obtained from Water Resources Management Authority (WRMA) regional office in Kakamega. This consisted of 113 data sets of DO, turbidity, temperature, pH and EC. WRMA measures DO, pH, EC and temperature in-situ using portable meters. The description of the sampling points is illustrated in Table 1. The data was split manually into training (80%) and testing (20%) by adjusting the default 70% to 30% since more data is necessary for training. This translates to 90 and 23 data sets for training and testing respectively.

Table 1. Description of the sampling points

Sampling point	Chainage, km	Description
Webuye Bridge (S1)	0	Upstream boundary. It is the upstream of effluent discharge point for Pan paper mills and Webuye municipal wastewater treatment plant.
S2	3	Downstream of discharge from Pan paper mills and Webuye municipal wastewater treatment plant.
S3	31	Nzoia Sugar factory discharge their industrial effluents to R. Nzoia via R. Kuywa.
S4	54	Upstream of Mumias Sugar factory and downstream of R. Khalaba which carries pollutants from Bungoma wastewater treatment plant.
Mumias Bridge (S5)	55	Downstream boundary. Downstream of Mumias Sugar factory.

The statistical parameters for the input and output data are shown in Table 2.

Table 2. Summary of input and output parameters

	Temperature, °C	pH	Turbidity, NTU	EC, µS/cm	DO, mg/l
Mean	23.34	7.38	229.24	179.27	4.58
Min	14	5.03	2.66	60.8	2.1
Max	29.3	9.48	960	1095	7.5
SD	2.44	0.63	242.32	164.77	1.14

ANN Architecture and Training

Determination of appropriate network architecture is one of the most important, but also one of the most difficult, tasks in the model building process (Sarda, Sadgir, 2015). Multilayer Perceptron (MLP) which is the most common form of feed-forward back-propagation (BP) model architecture (Maier *et al*., 2010) was chosen for this study. The basic and the most commonly used ANN architecture consists of an input layer, hidden layer and an output layer, where each of the layers consists of a number of interconnected neurons (Chau, 2006; Antanasijević *et al*., 2014). The number of input layers is normally determined by trial and error method but a one-hidden layered network is the most common (Palani *et al*., 2008). In the ANN structure, each node in the input and hidden layers receives input values, processes it, and passes it to the next layer using weights (W) and bias value (B) and uses its own transfer function (linear, sigmoid or the hyperbolic tangent function) to create an output value (Csábrági *et al*., 2015). In this case the MLP-BP architecture allows for information flows from the input layer to the output layer via the hidden layer (Heydari *et al*., 2013). The BP involves two steps where the effect of the input is passed forward through the network to reach the output layer and after the error is computed, a second step starts backward through the network and errors at the output layer are propagated back toward the input layer with the weights being modified (Emamgholizadeh *et al*., 2014). In the forward phase the weighted sum of the input components is calculated as in Equation 1 (Vicente *et al*., 2012).

$$u_j = \sum_{i=1}^{n} w_{ij} x_i + bias_j, \tag{1}$$

where: u_j is the input components, w_{ij} denotes the weight between the jth neuron and the ith neuron in the preceding layer, x_i denotes the output of the ith neuron in the preceding layer, and $bias_j$ denote the weight between the jth neuron and the bias neuron in the preceding layer.

The number of neurons in the input and output layer depends on the number of input and output variables respectively (Sengorur *et al*., 2015). The number of input neurons varied from 1 to 4 representing the input parameters that affect the DO while the output layer has one neuron representing DO. The number of neurons in the hidden layer was varied until an optimum performance is achieved for each input combination(s). The number of neurons ranged from 22 to 28. The training termination criterion was set based on the performance of a low mean square error of 0.001 or maximum epochs of 1000. The network structure adopted in this study is illustrated in Fig. 2.

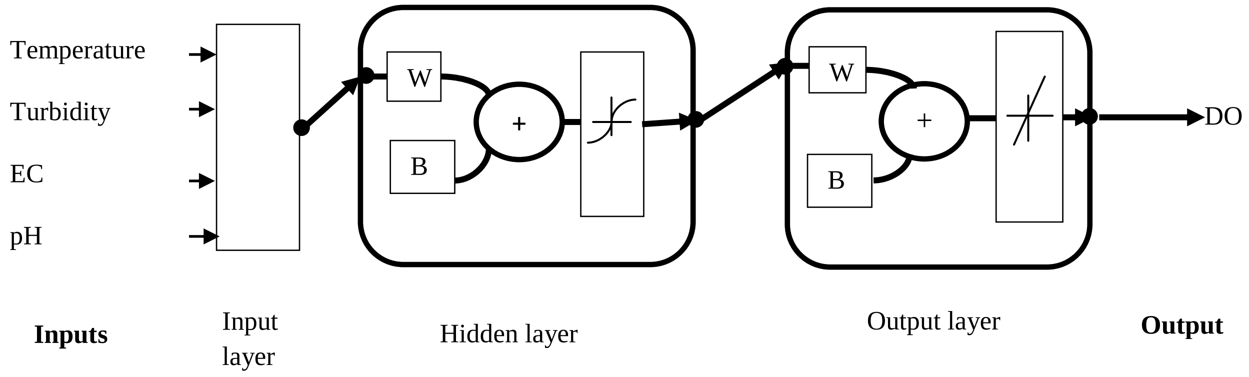

Fig. 2. ANN structure adopted for the study

Model Evaluation

The model performance was assessed using root mean square error (RMSE) and R^2.

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X-Y)^2}\ , \qquad 2)$$

Where: X is the observed value of DO, Y is the simulated value of DO, and n is the total number of observations. The lower the RMSE the better the model performance.
The goodness of fit of the model was assessed using the coefficient of determination, R^2.

$$R^2=\left[\frac{n(\sum XY)-(\sum X)(\sum Y)}{\sqrt{\left[n\sum X^2-(\sum X)^2\right]\left[n\sum Y^2-(\sum Y)^2\right]}}\right]^2 \qquad (3)$$

Results and Discussion

In order to select the most suitable input or input combination of the model, correlation analysis was carried out between the input and output and the results are as indicated in Table 3.

Table 3. Pearson correlation coefficients

	Temperature	pH	Turbidity	EC
Temperature	1			
pH	0.144	1		
Turbidity	-0.162	-0.204	1	
EC	0.274	0.390	-0.283	1
DO	-0.178	-0.079	0.284	-0.220

Turbidity with the highest correlation to DO (Table 3) was selected as the first input and trained in the network until a satisfactory performance was achieved. Then the input with second highest correlation (EC) was added to the turbidity model and network trained to satisfactory performance. A third model was developed which included turbidity, EC and temperature and the fourth model comprised all the four inputs. The results are shown in Table 4.

Table 4. Evaluation statistics

		Training		Testing		Overall	
SN	Model	RMSE	R^2	RMSE	R^2	R^2	RMSE
1	Turbidity	0.34	0.85	0.42	0.87	0.81	0.51
2	Turbidity +EC	0.53	0.87	0.56	0.79	0.83	0.48
3	Turbidity +EC + Temperature	0.46	0.86	0.68	0.89	0.85	0.46
4	Turbidity +EC + Temperature + pH	0.38	0.94	0.59	0.90	0.88	0.40

From the results in Table 4, the RMSE values ranged from 0.34 mg/l to 0.68 mg/l and R^2 ranged from 0.79 to 0.94 which indicate a satisfactory performance. According to the number of times the observations are greater than the mean error as discussed by Ritter and Muñoz-Carpena (2013), models 1, 2, 3 and 4 can be classified as good. The evaluation statistics indicated in Table 4 concurs with those found by Salami and Ehteshami (2015) where R^2 of 0.82, 0.85 and 0.92 for chloride, alkalinity and hardness respectively were obtained for prediction of dissolved oxygen. However, the RMSE values in Table 4 are slightly higher than those found by Zhang *et al*. (2010) since their inputs of temperature, BOD and ammonium had higher correlations with DO. The overall best fit plots i.e. using all the data are illustrated in Fig. 4 to 7.

The values of DO follow seasonal rainfall trend. Western region of Kenya experiences two rainy seasons from April/May–July/August (long rains) and October–November (short rains). DO is highest during the wet seasons and lowest during the dry season which normally occur from December/January to March/April (Fig. 3).

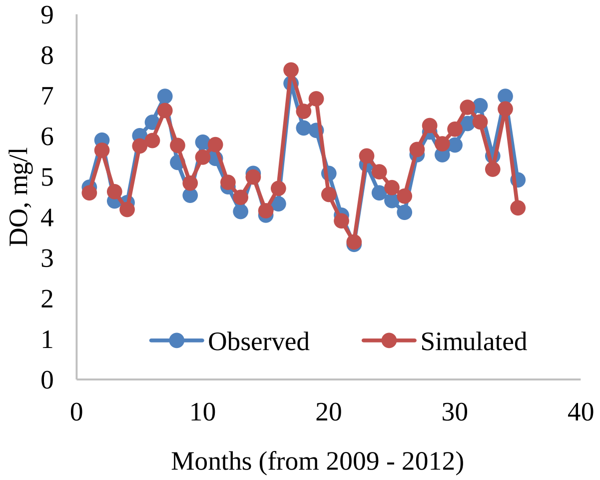

Fig. 3: Observed and simulated monthly DO concentrations at Webuye

Analysis of variance (ANOVA) between the DO values for dry months of December–March and wet months (April–August) showed significant variation ($p < 0.05$). During the dry months when the stream discharge is low, the value of DO is low which indicates that the river is susceptible to pollution during this time. During the wet periods, dilution mechanism enhance the self-purification mechanism of the river and thus the DO levels are high (Kanda, 2014). A study by Kanda *et al.* (2015) found that the DO decreased downstream from Webuye to Mumias which indicated the significant effect of the industrial effluent and municipal wastewater discharges from the towns situated in the river section between Webuye and Mumias. This was particularly important during low flows when the stream discharge is low.

From the results indicated in Fig. 4 to 7, it can be deduced that the ANN models developed from the readily available inputs of temperature, pH, electrical conductivity and turbidity can be used in prediction of DO in River Nzoia. The model can be reliably applied in pollution monitoring and control since the RSME values were less than 15% of the mean values of DO and also considering that the inputs have very weak relationship with the output as illustrated in Table 3. From pollution monitoring perspective, the ANN models developed for River Nzoia seems to be a good alternative tool in forecasting when DO measurements are unavailable.

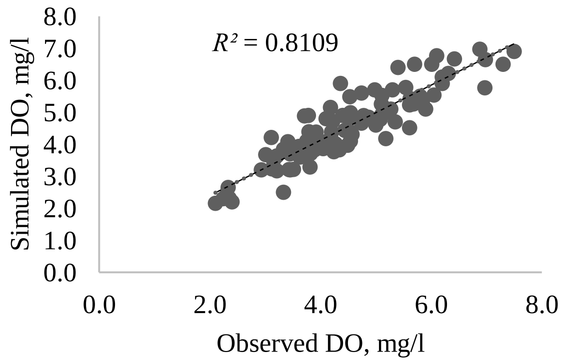

Fig. 4. Regression plot model 1

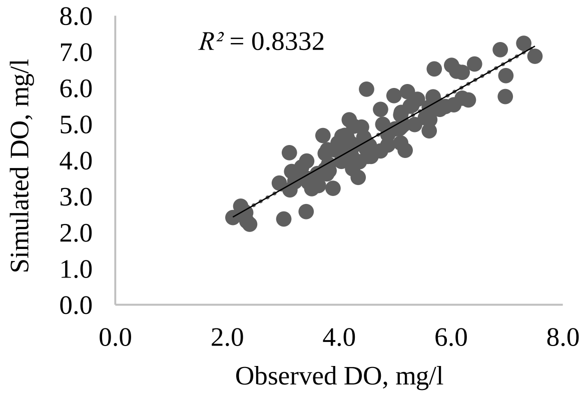

Fig. 5. Regression plot model 2

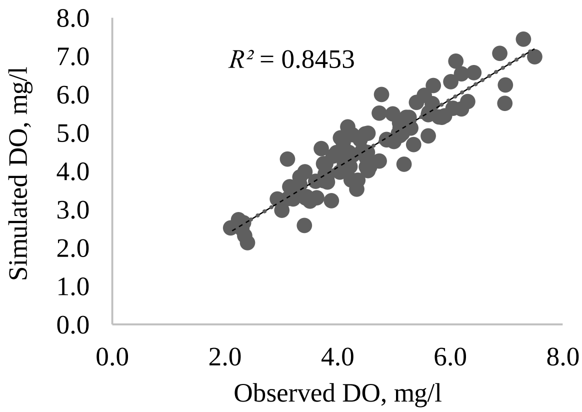

Fig. 6. Regression plot model 3

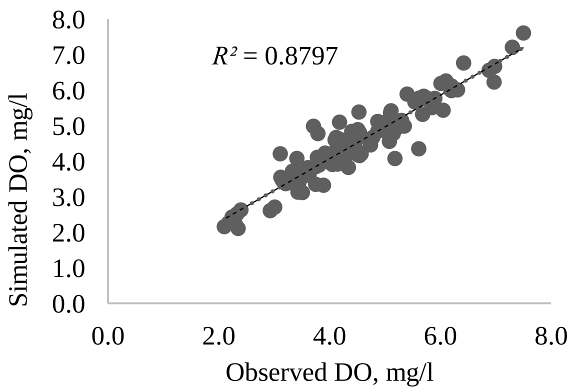

Fig. 7. Regression plot model 4

Conclusion

ANN is a popular forecasting tool in water quality studies. This study sought to determine the ability of feed forward back propagation artificial neural network in the prediction of DO in River Nzoia. This was accomplished using readily available input data of temperature, EC, turbidity and pH. The results indicated that the model combination of all the four inputs and the one which only excluded pH had good performance in predicting the DO for the river. For a country like Kenya where river water quality monitoring is hampered by insufficient funds, models such as ANN can be a good alternative to traditional process-based modelling which may require detailed data. DO is an important parameter in determining the pollution status of the river and therefore, the model developed in the study can be used to monitor the pollution levels in the river due to industrial effluent, municipal wastewater and agricultural runoff in Nzoia catchment.

References

Abira, M. A. 2008. *A pilot constructed treatment wetland for pulp and paper mill wastewater: performance, processes and implications for the Nzoia River, Kenya.* PhD Thesis, Wageningen University.

Ahmed, A. M. 2014. Prediction of dissolved oxygen in Surma River by biochemical oxygen demand and chemical oxygen demand using the artificial neural networks (ANNs). *Journal of King Saud University-Engineering Sciences.* (In Press) http://dx.doi.org/10.1016/j.jksues.2014.05.001

Antanasijević, D.; Pocajt, V.; Perić-Grujić, A.; Ristić, M. 2014. Modelling of dissolved oxygen in the Danube River using artificial neural networks and Monte Carlo Simulation uncertainty analysis. *Journal of Hydrology,* 519, 1895–1907. http://dx.doi.org/10.1016/j.jhydrol.2014.10.009

Chau, K.-W. 2006. A review on integration of artificial intelligence into water quality modelling. *Marine Pollution Bulletin,* 52(7), 726–733. http://dx.doi.org/10.1016/j.marpolbul.2006.04.003

Csábrági, A.; Molnár, S.; Tanos, P.; Kovács, J. 2015. Forecasting of dissolved oxygen in the River Danube using Neural Networks. *Hungarian Agricultural Engineering,* (27), 38–41. http://dx.doi.org/10.17676/HAE.2015.27.38

Diamantopoulou, M. J.; Papamichail, D. M.; Antonopoulos, V. Z. 2005. The use of a Neural Network technique for the prediction of water quality parameters. *Operational Research,* 5(1), 115–125. http://dx.doi.org/10.1007/bf02944165

Emamgholizadeh, S.; Kashi, H.; Marofpoor, I.; Zalaghi, E. 2014. Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *International Journal of Environmental Science and Technology,* 11(3), 645–656. http://dx.doi.org/10.1007/s13762-013-0378-x

Heydari, M.; Olyaie, E.; Mohebzadeh, H.; Kisi, Ö. 2013. Development of a neural network technique for prediction of water quality parameters in the Delaware River, Pennsylvania. *Middle-East Journal of Scientific Research,* 13(10), 1367–1376. DOI: 10.5829/idosi.mejsr.2013.13.10.1238.

Kanda, E. K. 2014. *Application of MIKE 11 simulation model for water quality management in River Nzoia.* MSc. Thesis, Moi University, Eldoret, Kenya.

Kanda, E. K.; Kosgei, J. R.; Kipkorir, E. C. 2015. Simulation of organic carbon loading using MIKE 11 model: a case of River Nzoia, Kenya. *Water Practice and Technology,* 10(2), 298–304. http://dx.doi.org/10.2166/wpt.2015.035

Khalil, B. M.; Awadallah, A. G.; Karaman, H.; El-Sayed, A. 2012. Application of artificial Neural Networks for the prediction of water quality variables in the Nile Delta. *Journal of Water Resource and Protection,* 4(6), 388–394. http://dx.doi.org/10.4236/jwarp.2012.46044

Kisi, O.; Murat, A. 2011. Modeling dissolved oxygen (DO) concentration using different Neural Network techniques. In: *International Balkans Conference on Challenges of Civil Engineering*, Tirana, Albania.

Maier, H. R.; Jain, A.; Dandy, G. C.; Sudheer, K. P. 2010. Methods used for the development of Neural Networks for the prediction of water resource variables in river systems: current status and future directions. *Environmental Modelling & Software,* 25(8), 891–909. http://dx.doi.org/10.1016/j.envsoft.2010.02.003

Najah, A.; El-Shafie, A.; Karim, O. A.; Jaafar, O.; El-Shafie, A. H. 2011. An application of different artificial intelligences techniques for water quality prediction. *International Journal of the Physical Sciences,* 6(22), 5298–5308. DOI: 10.5897/IJPS11.1180.

Ngaina, J. N.; Njoroge, J. M.; Mutua, F.; Mutai, B. K.; Opere, A. O. 2014. Flood forecasting over lower Nzoia Sub-Basin in Kenya. *Africa Journal of Physical Sciences* 1(1), 25–31.

Nyadawa, M.; Karanja, F.; Njoroge, T. 2010. Application of GIS-based spatially distributed hydrologic model in integrated watershed management: a case study of Nzoia basin, Kenya. *Journal of Civil Engineering Research and Practice,* 7(1), 61–76. http://dx.doi.org/10.4314/jcerp.v7i1.56748

Omwoma, S.; Omwoyo, W. N.; Alwala, J. O.; Ongeri, D. M.; Sylus, L. C.; Lalah, J. O. 2012. Nutrient reduction in runoff water from sugarcane farms by sedimentation method. *The Environmentalist,* 32(4), 494–502. http://dx.doi.org/10.1007/s10669-012-9416-3

Palani, S.; Liong, S.-Y.; Tkalich, P. 2008. An ANN application for water quality forecasting. *Marine Pollution Bulletin,* 56(9), 1586-1597. http://dx.doi.org/10.1016/j.marpolbul.2008.05.021

Prathumratana, L.; Sthiannopkao, S.; Kim, K. W. 2008. The relationship of climatic and hydrological parameters to surface water quality in the lower Mekong River. *Environment International,* 34(6), 860–866. http://dx.doi.org/10.1016/j.envint.2007.10.011

Ritter, A.; Muñoz-Carpena, R. 2013. Performance evaluation of hydrological models: statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology*, 480, 33–45. http://dx.doi.org/10.1016/j.jhydrol.2012.12.004

Salami, E.; Ehteshami, M. 2015. Simulation, evaluation and prediction modeling of river water quality properties (case study: Ireland Rivers). *International Journal of Environmental Science and Technology,* 12(10), 3235–3242. http://dx.doi.org/10.1007/s13762-015-0800-7

Sarda, P.; Sadgir, P. 2015. Computation of water quality parameters and prediction tool ANN for modeling of water quality of reservoir. *International Journal of Innovative Research in Science, Engineering and Technology,* 4(9), 8906–8911.

DOI: 10.15680/IJIRSET.2015.0409086

Sengorur, B.; Koklu, R.; Ates, A. 2015. Water quality assessment using artificial intelligence techniques: SOM and ANN – A case study of Melen River Turkey. *Water Quality, Exposure and Health,* 7(4), 469–490. http://dx.doi.org/10.1007/s12403-015-0163-9

Vicente, H.; Couto, C.; Machado, J.; Abelha, A.; Neves, J. 2012. Prediction of water quality parameters in a reservoir using artificial neural networks. *International Journal of Design & Nature and Ecodynamics*, 7(3), 310–319. http://dx.doi.org/10.2495/DNE-V7-N3-310-319

Zhang, Z.; Wang, X.; Ou, Y. 2010. Water simulation method based on BPNN response and analytic geometry. *Procedia Environmental Sciences,* 2, 446–453. http://dx.doi.org/10.1016/j.proenv.2010.10.049