A New ANN-Markov Chain Methodology for Water Quality Prediction

Xiu Li, Jingdong Song
Shenzhen Key Laboratory of Information Science and Technology
Graduate School at Shenzhen, Tsinghua University
Shenzhen, China
Email: qingjing0413@163.com

***Abstract*— In recent years, water quality prediction has attracted many attentions of governments and researchers. The safety of water quality seriously affects the human health, fishery economy and agricultural activities. If an early prediction to the water quality with an acceptable accuracy can be achieved, the negative impacts will be minimized or even be avoided. Many researchers have applied artificial neural networks (ANNs) to build the water quality models for there is a complicated non-linear relation between the prediction variables and measured inputs. However the ANN models are easy to be over-fitting for training them needs a large of samples. As the objective of this study, artificial neural network and Markov chain approach are used to develop a new hybrid methodology for predicting the biochemical oxygen demand which is the main indicator of water quality. ANN produces the primary values and then the results are modified by three regression methods using the Markov transitional probability matrices respectively. We use a 27-year water quality data set of Tolo Harbor which only has a total of 439 samples to test our method. The results are validated and a better prediction accuracy of the new ANN-Markov Chain Methodology is demonstrated through three criteria.**

Keywords*— *ANN, Markov chain, water quality prediction, BOD

I. INTRODUCTION

Water plays an important role in our daily life, and the quality of water in a region heavily affects the sustainable development of local normal industrial, agricultural and other anthropogenic activities. Conversely, the quality of water largely depends on the human activities and it is one of the main characteristics of a river or harbor. In recent years water quality in many coastal areas of the world has significantly dropped for the uncontrolled discharge of industrial and domestic wastewater. What's the worse, some solid waste is also poured into the rivers or harbors. Biochemical oxygen demand (BOD), which measures an approximate amount of bio-degradable organic matter present in water, is generally used as the criterion to measure and express the water quality [1]. The BOD is defined by the amount of oxygen required for the aerobic microorganisms present in the water samples to oxidize the organic matter to a stable organic form [2]. The value of BOD becomes higher which means that the value of dissolved oxygen (DO) is lower, and the quality of water is worse. However the test of BOD in traditional mechanistic methods is time taking and difficult. The results are influenced by the external environments greatly which usually need cost 5 days under a constant temperature like 20 ℃, leading to measurement errors.

Therefore, we need some more effective and accuracy approaches to predict the value of BOD, and more attentions have been concentrated on the machine learning methods which collect a large number of water samples and construct prediction models to find out the relationship between the monitoring variables. In [3], Reckhow expresses the complex determination in statistical mechanics, using Bayesian probability networks for water quality assessment and prediction of the Neuse River estuary in North Carolina. Chau et al. [4] used both data mining and multivariate statistical analysis to study the coastal water quality data from Tolo Harbour, Hong Kong, and they found out that the monitoring station, TM3, in the in the Harbour Subzone is the most prone to eutrophication and it is necessary to take measures to control pollutant loadings from internal and external sources. The main techniques of the two methods are box plots and factor analysis respectively. In [5] Hendrik et al. applied a single decision tree to predict the multiple physico-chemical properties of river water, and in [6] Hendrik et al. used regression trees to predict chemical parameters of river water quality from bio-indicator data. Both methods can also be utilized by BOD determination. As we have said before, the measurement of BOD is affected by a lot of factors, and there is a complicated non-linear relationships among those variables. As the ANN (artificial neural network) is a well-suited method with self-adaptability, self-organization, and error tolerance, which is quite suitable for nonlinear simulation, it may be the best choice to construct the prediction models. In recent years, ANNs [1, 6-9] have been successfully used in water quality problems, and it can approximate almost desired parameters of water. Sing et al. [1] adopt the Levenberg-Marquardt algorithm to improve the performance of traditional back-propagation (BP) neural network for it is too slow, and construct a model for the Gomti river water quality prediction. Besides, they also demonstrate its application to complex water quality data as how it can improve the

This research was partly supported by National Natural Science Foundation of China (NSFC, Project No.:71171121/61033005) and National "863"High Technology Research and Development Program of China. (863 Project No.: 2012AA09A408).

978-1-4799-1959-8/15/$31.00 ©2015 IEEE

interpretation of the results. [6] presents an efficient self-organizing RBF neural network for water quality prediction and the model can vary its structure dynamically in order to maintain the prediction accuracy. As well as the neural network proposed has fewer hidden neurons and fast convergence speed. Robert et al. [9] discuss the input variable selection (IVS) to ANN prediction model of water quality, and propose a newly non-linear IVS algorithm to reduce the need for arbitrary judgments and extensive trial-and-error during model development.

This paper is inspired by the previous work of combining ANN and Markov chain [10-12]. Pourmousavi et al. [10] use the ANN, Markov chain and linear regression to develop a new model for very short term wind speed prediction, and they also substitute a second ANN for the linear regression to improve the accuracy of final results [11]. Yoshua, et al. [12] propose an algorithm for global optimization of a neural network–hidden Markov model. Those models use the ANNs to train the primary models to give their basic prediction values, and then use the Markov chain to calculate transition probability matrices between the different states of predicted values. Finally, use the selected regression method to combine the ANN estimated values and Markov chain calculated probabilities to output the results. The prediction of BOD has a lot of uncertainties, and there is a big possibility of over-fitting if we applied ANN alone to train the prediction model, for it uses a large of training data set and need many iterators. Therefore, an ANN-Markov chain model is developed for BOD measurement.

The rest of the paper is organized as follows. In section 2, the study area of water quality and the training data set are described. And the detail of the proposed methodology is presented in section 3. Section 4 shows the measure criteria and the results of experiments compared with the ANN method alone. In the last section, conclusions are given.

II. Study Area and Water Quality Data Set

A. *Study Area*

Tolo Harbor is a semi-enclosed and shallow bay with a mean depth of 6–7 m in the northeastern part of Hong Kong (shown in Fig. 1) which receives Sha Tin Shing Mun River, Tai Po Lam Tsuen River and other tributaries. Various municipal and livestock waste discharges into the harbor during its course, and water pollution has been a major environmental concern since the 1970s. Due to its bottleneck topography that it connects to the open sea at Mirs Bay through a narrow tidal channel, the water exchange is poor and the accumulation of waste is more and more. The Hong Kong Government set up eight monitoring stations distributed spread over the harbor to prevent the declining of water quality of Tolo Harbour in 1986 called the Tolo Harbour Action Plan (THAP).The monitoring stations collect the biweekly water quality data of three different depths, and the stations are also shown in Fig.1.

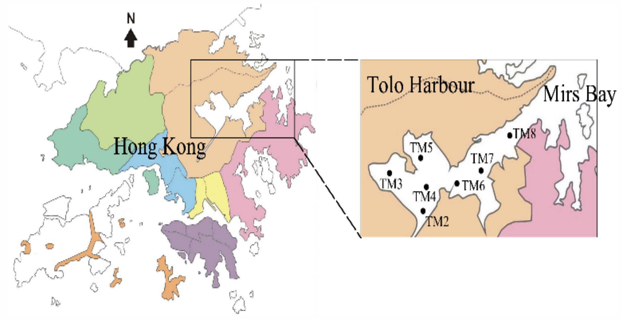

Fig. 1. Location of study area: Tolo Harbour

B. *Water Quality Data Set*

Reference the related work [13-14] in this area, we also choose the data of TM3 which is the most weakly flushed monitoring station that can isolate the hydrodynamic effects as much as possible. We choose the data from 1986.1 to 2012.12, and use the depth-averaged data for this work. Here for ANN modeling, we also choose the following 13parameters based on the previous studies: chlorophyll-a concentration, Chl-a (μg/l); water pH; total inorganic nitrogen, TIN (mg/l); orthophosphate, PO4 (mg/l); total phosphorus, TP (mg/l);total nitrogen, TN (mg/l);water temperature, TEMP (℃); secchi disc depth, SD (m); Kjeldahl nitrogen, TKN(mg/l); nitrate nitrogen, NO3-N (mg/l);silicon dioxide, SiO2(mg/l); dissolved oxygen, DO(mg /l), and 5-day biochemical oxygen demand, BOD(mg/l).All variables are normalized to the interval [-1, +1] by transformation to guarantee the analytical data quality.

III. The Proposed ANN-Markov Chain Methodology

The proposed ANN-Markov chain method is shown in Fig.2. It is mainly consist of three modules. First we use 351 records of water quality data of TM3 station in Tolo Harbor to train the ANN model to predict primary values and generally the results in test data set is not well fit with the measured values. Then the Markov chain module provides the state transition probabilities and here we use a second-order Markov chain state transition matrix. The final predictions of BOD are given by combing the primary determinations by ANN and the state transition probabilities of these predicted values by Markov chain approach through the specific regression methods.

A. *Back Propagation Neural Network and Levenberg-Marquardt Algorithm*

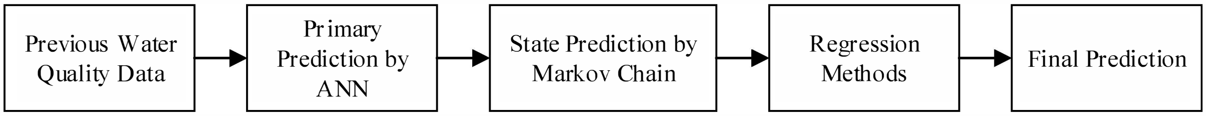
Previous Water Quality Data
Primary Prediction by ANN
State Prediction by Markov Chain
Regression Methods
Final Prediction

Fig. 2. Flowchart of the proposed ANN-Markov chain method

The artificial neural network as introduced above, which imitating the activities of human brain, is well suitable to deal with nonlinear problems. A typical neural network is comprised of an input layer which receives inputs of the model and computes the weighted sum of inputs, an output layer which provides the final results, and one or more intermediate hidden layers for processing. Each layer may has one or more processing nodes called neurons and each neuron has its own transfer function. As well as the weights of different inputs to a specific neuron are different. The signal passes the three modules and ends in the output layer. The most widely used neural network is the three-layer feed-forward neural networks with back propagation. The traditional back propagation neural network is based on a gradient descent algorithm which distributes the output error though the model after one iteration or 'epoch' in a back direction. And the weights will be adjusted during the algorithm to realize the minimum error. However the standard BP algorithm has a big deficiency that slow convergence speed for its complicated optimal objective function.

In this study, a Levenberg-Marquardt based BP network (LMBP) [15] is used for modeling water quality and it is much faster than other BP networks as the Levenberg-Marquardt algorithm is a modified Gauss-Newton method which has both the global characteristics of the gradient descent algorithm and the local characteristics of the Gauss-Newton algorithm. The core idea of this algorithm can be presented as

$$\Delta\omega = -\left[J^T J + \mu I\right]^{-1} J^T e \quad (1)$$

where the $\Delta\omega$ are the updates of weights of one iteration, the J is the Jacobian matrix obtained by calculating the first derivatives of the network output errors with respect to the weights. And μ is the step length or the learning rate of the optimal algorithm. From [16], we know when the μ is large the algorithm becomes gradient descent, while for small μ the algorithm becomes Gauss-Newton.

Our neural network is also a three-layer one for more hidden layers may cause unnecessary computation complexity. We have 12 neurons in the input layers and one neuron in the output layer which stands for the BOD value to be predicted. The appreciate number of neurons in the hidden layer and the transfer functions are chosen by trial and error processes as well as related experience [1]. Gradually varying the number of neurons in the hidden layer from 7 to 25, the optimal numbers for the model is found to be 12. Linear transfer function (purelin) is used as the activation function for the hidden layer and the tangent sigmoid transfer function (tansig) for the output layer. The two transfer functions are developed by matlab neural network toolbox. The convergence speed is sensitive to the value of learning rate. Too small can lead to slow and too large can produce the undue oscillations and fall into the local optimum. The initial weights are random generated between -1 and +1, and the mean square error (MSE) is given by (2) as one of the stopping criteria.

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(t_i - o_i\right)^2 \quad (2)$$

where t_i and o_i represent the model predicted and measured values of BOD respectively, and N is the total number of observations. The specific values of learning rate, maximum number of epochs, target error goal and minimum performance gradient are 0.05, 1000, 0.001 and 10-exp10.

B. Markov Chain Approach

A Markov chain is a stochastic process that the probability of the next state can be deduced from the preceding states, and it just depends on the current state not the state sequence before it. Let X_n, (n= {1,2,…,k}) is stochastic process which meets the characteristic of Markov chains, then the conditional probabilities should be :

$$\begin{aligned} \Pr &= \left\{X_{n+1} = x \mid X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1\right\} \\ &= \Pr\left\{X_{n+1} = x \mid X_n = x_n\right\} \end{aligned} \quad (3)$$

Transitional probability matrices of various time steps are the basic tools to analyze the Markov chains. In general if the number of states we divide is n, the transitional matrix size will be n×n between two successive time instances. Each element of the matrix, p_{ij} represents the probability that a state moves to another over time. A first order transition matrix P is shown as:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{21} & \cdots & p_{n1} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \quad (4)$$

In (4), the p_{ij} represents the probability a state at time t to another state at time t+1, and be approximately calculated by (5).

$$p_{ij} = \frac{n_{ij}}{\sum_j n_{ij}} \quad (5)$$

where the n_{ij} is the number of transitions from state i to state j in the time sequence of precede data. And the matrices have some valid properties by definition: any state probability varies between zero and one and the row summation in the transition matrices is equal to 1 [17].

Many natural processes can be considered as Markov chain processes. In this study, we use the (6) to weigh the difference between the modeled values by ANN and measured values. The results calculated by (6) have a large stochastic volatility. And according to distribution of the results, we divide the results into four states as shown in Table 1. What's more, we also use a second order transition probability matrix for four states which can be obtained by a multiplication of P matrix by itself.

$$er = \frac{BOD_{measured} - BOD_{modeled}}{BOD_{modeled}} \qquad (6)$$

TABLE I. THE DIVIDED STATES OF ER

States	*ER*
S1	(-0.8, -0.4]
S2	(-0.4, 0]
S3	(0, 0.4]
S4	(0.4, ∞)

C. Regression Methods

After the two modules above, the next step is combining the primary values by ANN and the transition probability matrices to produce the final prediction results. Herein we apply three regression approaches, and they are linear method, ANN method and SVR method respectively.

- Linear Regression. The linear regression function is made by ourselves, as shown in (7).

$$BOD_{final} = BOD_{modeled}\left[1 + \sum_{j=1}^{4} \frac{\left(p_{ij}^{(1)} + p_{kj}^{(2)}\right)}{2} \Delta_j\right] \quad (7)$$

Before we use the equation, we must know the specific states precede the value we want to predict one and two steps. The $p_{ij}^{(m)}$ represents the probability of state i transferring to state j in m steps. And the Δ_j represents the mean er value of state j, while the value of the fourth state is 0.6.

- ANN method. A second ANN model is designed to fit the final results. It is also a typical three-layer feed-forward neural network and uses the traditional gradient descent algorithm. There are five inputs and one output that is the final BOD. And the five inputs are the modeled value by the first ANN model and average values of first and second order transition probabilities from the previous states to the four states respectively. The number of hidden layer neurons is five.
- SVR method. Similar to the second ANN model, we also use the six values to train the regression model. And we choose the ε-SVR of MATLAB neural network toolbox which use a RBF kernel function. In SVM approach, we use a suitable kernel function to map the input into a high dimensional feature space and find out a maximal separating plane, however in the case of SVR we should find the minimal distance plane from the data points[18].

IV. RESULTS AND DISCUSSION

A. Model performances

The performance of models is evaluated using the following measures: the root mean square error (RMSE), the Bais and the coefficient of determination (R^2) [1]. The goodness-of-fit measures are defined as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2} \qquad (8)$$

$$Bias = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right) \qquad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{N}\left(Y_i - \overline{Y}_i\right)^2} \qquad (10)$$

where N is the total number of data; Y_i and $\hat{Y}_i$ are the actual and simulated data respectively; and the average value of the associated variable is represented by 'bar' above the variable. The RMSE describes the performance in average error for predicting the dependent variables which is well suited for the measure of the goodness-of-fit. While the Bias calculates the mean errors of all individual prediction results, it can used to indicate the degree of estimation of variables. And the R^2 represents the percentage of variability that can be explained by the model[1].

B. Results and Analysis

In this section, we compare the proposed model with the first ANN model, and the proposed method also can be divided into three kinds of forms with respect to the regression methods. 351 records of water quality data which is about 80% of the total samples is used to train the whole model and data employed in each module is different. The rest of records are applied to produce the final predictions and test the model performances.

The RMSE, Bias and R^2 are presented in the Table 2. We can conclude that the ANN-Markov chain models show better performance than the ANN model alone in all measure indicators, and the ANN-Markov chain model using SVR is the most fitting method with RMSE of 0.6174, Bias of -0.0426 and R^2 of 0.8782. Fig. 3 shows the comparison of the predicted BOD values of different models with measured values, and for a better comparison, just 20 predicted values are illustrated. In Fig. 4, the errors produced by all methods are depicted. From these pictures, we can see the ANN-Markov chain models better meet the measured value plot and the error distribution is more concentrated near the zero value. Also the ANN- Markov chain model using SVR reduces the absolute errors more.

TABLE II. PERFORMANCE PARAMETERS OF THE PREDICTION MODELS

MODELS	**RMSE**	**Bias**	**R^2**
ANN alone	0.6597	-0.13032	0.8610
ANN-Markov chain-linear	0.6504	-0.0954	0.8649
ANN-Markov chain-ANN	0.6311	-0.0974	0.8727
ANN-Markov chain-SVR	0.6174	-0.0426	0.8782

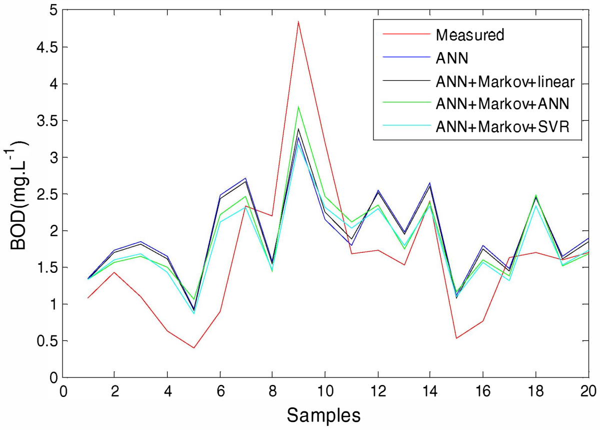

Fig. 3. BOD prediction by ANN-Markov chain methods and ANN alone.

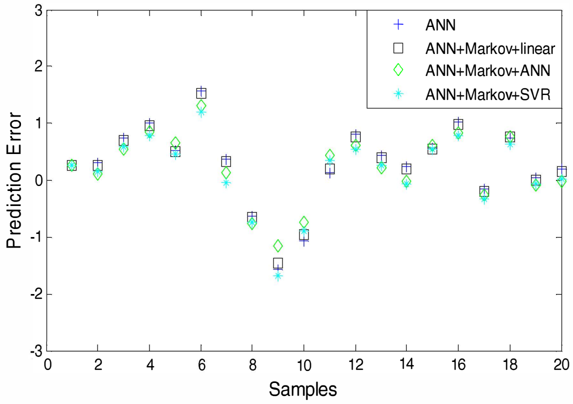

Fig. 4. Error generated by ANN-Markov chain methods and ANN alone.

V. Conclusion

This paper has studied the problem of water quality prediction, and proposes a new ANN-Markov chain methodology to predict the value of BOD which is the main indicator of water quality. A data set with 27 years of Tolo Harbor is used to train and evaluate the method. The results show that the proposed methodology can help make more accurate prediction than ANN approach considering three criteria. In our proposed methodology, the ANN module plays a primary role to obtain the initial results. Then the results are modified by the Markov transition probability matrices and a specific regression method. Among the three regression methods we provided, the SVR with ANN and Markov chain achieves the best overall performance.

The Markov chain approach is applicable for time series prediction which has a big stochastic fluctuation, and suitable for the small sample sets. We can see that although we use a 27 years data set, the available samples are not very large. In this study, we just use the first and second order transition probability matrices, and in the future we can apply higher order matrices to evaluate the accuracy.

References

[1] K. P. Singh, A. Basant, A. Malik, and G. Jain. "Artificial neural network modeling of the river water quality—a case study." Ecological Modelling, vol 220, no.6, pp: 888-895,2009.

[2] D. V. Chapman., ed. "Water quality assessments: a guide to the use of biota, sediments and water in environmental monitoring." 1996.

[3] K. H. Reckhow. "Water quality prediction and probability network models." Canadian Journal of Fisheries and Aquatic Sciences vol.56, no.7, pp.1150-1158,1999.

[4] K. Chau and N. Muttil. "Data mining and multivariate statistical analysis for ecological system in coastal waters." Journal of Hydroinformatics vol.9, no.4, pp. 305-317,2007.

[5] H. Blockeel, S. Džeroski, and J. Grbović. "Simultaneous prediction of multiple chemical parameters of river water quality with TILDE. " Springer Berlin Heidelberg, 1999.

[6] S. Džeroski, D. Damjan, and J. Grbović. "Predicting chemical parameters of river water quality from bioindicator data." Applied Intelligence vol.13, no.1, pp. 7-17,2000.

[7] E. Dogan, B. Sengorur, and R. Koklu. "Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique." Journal of Environmental Management vol.90, no.2, pp. 1229-1235, 2009.

[8] K. Chau. "A split-step PSO algorithm in prediction of water quality pollution." Advances in Neural Networks–ISNN 2005. Springer Berlin Heidelberg, 2005, pp.1034-1039.

[9] R.J. May, G. C. Dandy, H. R.Maier, and J. B. Nixon. "Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems."Environmental Modelling & Software vol. 23, no. 10, pp. 1289-1299,2008.

[10] S. P. Kani, S. M. Mousavi, A. K. Kaviani, and G. H. Riahy. "A new integrated approach for very short-term wind speed prediction using linear regression among ANN and Markov Chain." Proceeding on International Conference on Power System Analysis, Control and Optimization, 2008.

[11] S. P. Kani and G. H. Riahy. "A new ANN-based methodology for very short-term wind speed prediction using Markov chain approach." Electric Power Conference, 2008. EPEC 2008. IEEE Canada. IEEE, 2008, pp. 1-6.

[12] Y. Bengio, R. D. Mori, and G. Flammia. "Global optimization of a neural network-hidden Markov model hybrid." Neural Networks, vol. 3, no. 2, pp. 252-259, 1992.

[13] H. Han, Q. Chen, and J. Qiao. "An efficient self-organizing RBF neural network for water quality prediction." Neural Networks, vol. 24, No. 7, pp.717-725, 2011.

[14] C. Sivapragasam, N. Muttil, S. Muthukumar, and V. M. Arun, "Prediction of algal blooms using genetic programming," Marine pollution bulletin, vol. 60, pp. 1849-1855, 2010.

[15] J. Hu, P. Ji and C. Zhang. "Prediction Model for Red Tide at Yantai Sishili Bay Based on LMBP Algorithm," Journal of System Simulation, vol. 19, p. 068, 2009.

[16] M. T. Hagan, and M. B. Menhaj. "Training feedforward networks with the Marquardt algorithm." Neural Networks, vol. 5, no. 6, pp. 989-993, 1994.

[17] A. D. Sahin, and Z. Sen. "First-order Markov chain approach to wind speed modelling." Journal of Wind Engineering and Industrial Aerodynamics, vol.89, no. 3, pp. 263-269, 2001.

[18] J. Qu and M. J. Zuo. "Support vector machine based data processing algorithm for wear degree classification of slurry pump systems." Measurement, vol. 43, no.6, pp. 781-791, 2010.