Comparison of Artificial Neural Network and State Space Model for Predicting River Water Quality

Jun K.W.[1], Kım S.D.[2,*] And Lee H.J.[2]

[1]Graduate School of Disaster Prevention, Kangwon National University, 346 Joongang-ro, Samcheok, Gangwon, 245-711, S.Korea

[2]School of Civil Engineering, Chungbuk National University, 52 Nassudong-ro, Cheongju, Chungbuk, 361-763, S.Korea

*corresponding author:

e-mail: soungdoug@hanmail.net

Abstract The purpose of this study is to investigate appropriate tools for river water quality forecasting as variations in water quality are difficult to predict due to the complicated nature within the range of various water quality factors. In this study, state space and neural network models are employed to mathematically analyze the intricate nonlinearity of processes that affect factors related to water quality. A monthly forecasting model is proposed that can predict water quality parameters, including dissolved oxygen (DO), biochemical oxygen demand (BOD), and suspended solid (SS) at the Miho river station in the Geum river basin (Korea). River water quality is predicted through the learning and the verification processes after applying the neural network theory to the proposed water quality forecasting model. Practical applications for predicting water quality prediction are examined by comparing the proposed model to the state space model (SSM). As a result, the artificial neural network (ANN) is estimated to have the ability to predict water quality more accurately than the state space model for each water quality item.

Keywords: Water quality, back propagation, neural networks, state space

1. Introduction

In recent years, researchers have been interested in improving water quality forecasting techniques. The surface water quality in a region largely depends on nature and on the extent of the industrial, agricultural, and other anthropogenic activities in or near the catchments. River systems are most adversely affected as a result of their dynamic nature and easy accessibility for waste disposal through drains and tributaries by direct or indirect means. Since, rivers and streams are among the most important sources of water used for irrigation, industry, and other purpose; they serve as lifelines for populations residing near the basins. Models for forecasting and examining river water quality are widely categorized into two types, which include a physical model and a system model. The physical model can be employed based on the use of empirical relationships between natural phenomena related to river water quality and mathematical models. The representation of a physical model can be achieved with the Enhanced Stream Water Quality Model (QUAL2E) and the Water Quality Analysis Simulation Program (WASP5). The system model has many advantages that can be employed. These advantages include the simplicity of composing the input and output data without having to rigorously understand the physical, chemical, and biological reactions in the water system. It also can be used for both short-term and long-term forecasting. Due to these advantages, the system model has been validated for applications in various fields. There are several representative system models, including the Autoregressive Integrated Moving Average (ARIMA) model, the State Space model (SSM), and the Artificial Neural Network model (ANN) (Faruk, 2009). The artificial intelligence algorithm generates numerical analysis by imitating the human thinking process. This algorithm is applicable not only for an interpretation of a simple relationship within data, but also for a non-linear analysis of its correlation. One of the representative system models is the artificial neural network (ANN) model (Yeon *et al*., 2009 and Zang, 2009). The ANN model is extensively used nowadays due to the simplicity of its application, its descriptive ability for non-linear characterizations, and its robust ability for prediction. Maier and Dandy (1996) applied the neural network model to predict the water quality of the Murray River in South Australia. Nouh (1996) applied the neural network model to estimate the optimal parameters of the SWMM model to model water quality of urban sewage. Kenichi *et al*. (1997) compared the back propagation (BP) algorithm to the field measurements for forecasting eutrophication in lakes. Palani *et al*. (2008) demonstrated the application of neural network theory to a model with values of selected seawater quality variables, having the dynamic and complex processes hidden in the monitored data itself. In this study, water quality forecasting is conducted to suggest water quality control and countermeasures river for abnormalities for the future. The objectives of the present study are to: (1) develop an ANN and a state space model that predict monthly water quality data, (2) assess the performance of each modeling approach using observed data versus predicted data, and (3) evaluate the predictive performance of ANN relative to the state space model by analyzing accuracy measures.

2. Theoretical Background

2.1. State Space Model

The state space model (SSM) is one of the tools that can be used in the field involving time series for water quality forecasting. The model was introduced by Kalman(1960) and is also widely known as the Kalman filter model. In order to predict the future state, the current state is defined using the smallest possible subset of current and past information that can be used to create a prediction for the future state. The SSM can be expressed by formulating the process equation (1) and the measurement equation (2).

$$x_{k+1} = F_{k+1,k} x_k + w_k, w_k \sim N(0, Q_k) \qquad (1)$$

$$y_k = H_k x_k + v_k, v_k \sim N(0, R_k) \qquad (2)$$

Where $F_{k+1,k}$ is the transition matrix taking the state x_k from time k to time $k+1$. y_k is the observable variable at time k and H_k is the measurement matrix. The two vectors w_k and v_k are the process and the measurement noise, respectively, where Q_k and R_k are the covariance matrices of the normal distributions.

2.2. Artificial Neural Network Model

The neural network model is usually comprised of single or multiple layers, thereby being referred to as a single-layer neural network or multi-layer neural network, respectively, to differentiate between models with different numbers of layers. In particular, the three layer neural network structure is the most extensively used, comprised of an input layer, output layer, and a hidden layer. The neural network model needs an algorithm that can obtain the value of its own connection weight by learning and repetition in order to produce reliable results that can expect an output for each corresponding input. In this study, the methods of moment and the adaptive learning rate were utilized as a form of a back propagation (BP) algorithm to improve the rate and the unstable learning outcome. The moment method can improve the converging rate, and it is useful to accelerate the learning speed. This method is extensively used because of its simplicity and efficiency. Moment coefficients are employed to avoid the vibration in errors generated by the learning process and to maximize the learning rate. The current adjusting volume of the connecting intensity can be computed by considering the variation of the connecting weight, which can be written expressed as follows:

$$\Delta W(t+1) = \eta \delta_y H + \alpha \Delta W(t) \qquad (3)$$

$$\Delta V(t+1) = \eta \delta_h X + \beta \Delta V(t) \qquad (4)$$

where, t is the number of iteration, η is the learning rate, δ_y is the error of output layer, δ_h is the error of the hidden layer, α and β are the moment coefficients of weights W and V, and H and X are vectors of the hidden layer and the input layer, respectively. The learning rate η can be adjusted and is controlled with the moment method of moments. This reate can be enhanced by equation (5)

$$\eta(t+1) = \begin{pmatrix} r_1\eta(t), & E(t+1)\langle E(t) \\ r_2\eta(t), & E(t+1)\langle r_3 E(t) \\ \eta(t), & otherwise \end{pmatrix} \qquad (5)$$

where, r_1, r_2, r_3 are parameters corresponding to the adjustment of the learning rate, and $E(t+1)$, $E(t)$ are current and previous step errors.

3. Application and Results

3.1. Study Area

In this study, Mihocheon station, which is part of the Guem river station and is located in the central part of the Korea Peninsula, is considered to be a tripped basin. The area of river basin is 1,850 km^2, and the river length is 87.3 km. Water quality data from the stripped basin is required to evaluate the proposed models. The data in the selected location is used to estimate the parameters needed, to verify and forecast the state space model, and to learn and forecast using the neural network model. The water quality data set used in this study from Mihocheon station is provided by the Ministry of Environment of Korea from their water quality measurement database. This data had been produced by measuring the representative water quality taken monthly from January 1998 to December 2008. The data set, taken from 1998 to 2007, is used to estimate the parameters necessary to learn the state space model and neural network model. The data taken in 2008 are used for the forecasting of the model.

3.2. State Space Model

Water quality indicators [including water temperature, dissolved oxygen (DO), biochemical oxygen demand (BOD), suspended solid (SS), total nitrogen (TN), total phosphorus (TP), and chemical oxygen demand(COD)] are used as input variables in the model to forecast water quality at Mihocheon station using the state space model. In this study, the estimation of DO, BOD, and SS concentrations are conducted at each selected location by using the SAS/ETS program. The canonical correlation analysis is used to select the time difference of the initial VAR(p) model and state space vectors, and water quality forecasting is performed after determining the state space model equation and the vector autoregressive moving average (VARMA) model equation. A simulation of the water quality concentration of each item is conducted at Mihocheon station though the proposed state space model and we subsequently evaluated the finalized state space model equation and the VARMA model equation. The result of the simulation is shown in Table 1 and Figure 1.

Table 1. Statistical analysis of state space model at Mihocheon.

Index	Model	Average	SD	Skewness	RMSE
DO	State space	9.276	2.380	0.882	1.871
BOD	State space	4.454	2.886	0.796	1.402
SS	State space	16.203	12.011	1.240	3.400

Table 2. Statistical analysis of state space model at Mihocheon.

Index	Input layer node(n)	Hidden layer node	Input data	Output data
	1	2n ~ 6n	DO_t	DO_{t+1}
DO	2	2n ~ 6n	DO_t, $TEMP_t$	DO_{t+1}
	1	2n ~ 6n	BOD_t	BOD_{t+1}
	2	2n ~ 6n	BOD_t , COD_t	BOD_{t+1}
BOD	2	2n ~ 6n	BOD_t , Q_t	BOD_{t+1}
	3	2n ~ 6n	BOD_t , COD_t, Q_t	BOD_{t+1}
SS	1	2n ~ 6n	SS_t	SS_{t+1}
	2	2n ~ 6n	SS_t, Q_t	SS_{t+1}

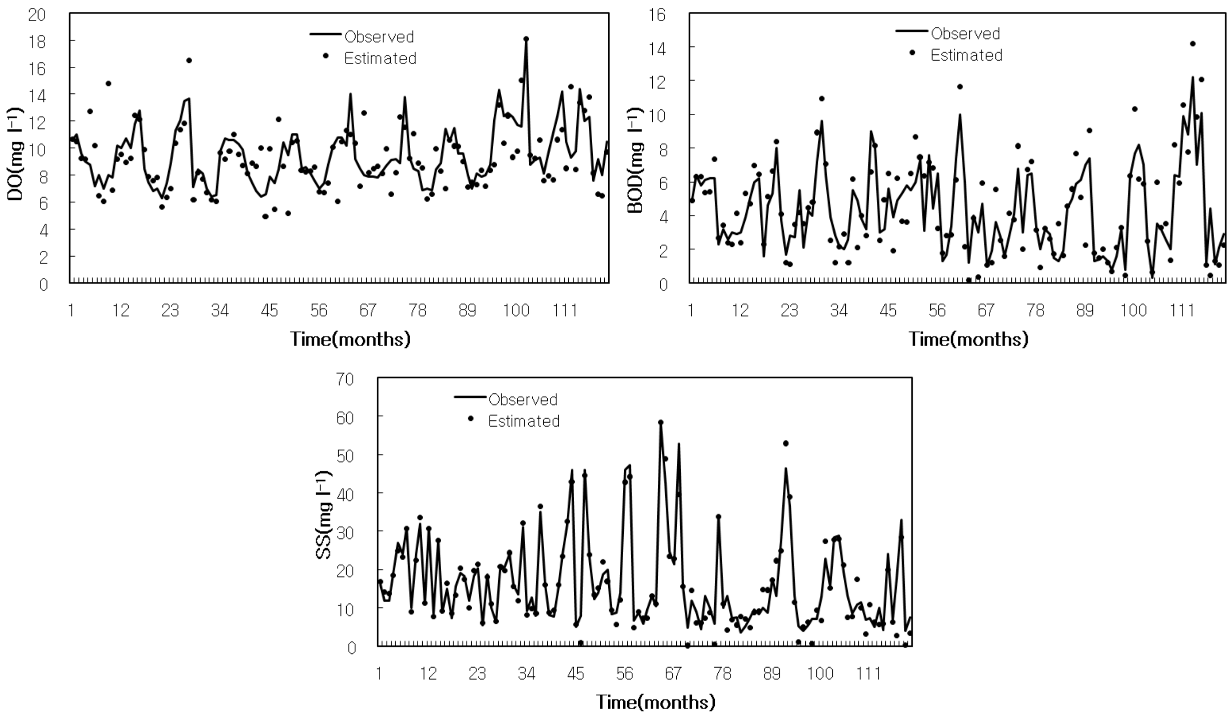

Figure 1. Comparison between observed and estimated values for DO, BOD, and SS (SS Model)

Table 3. Statistical analysis of neural network models at Mihocheon.

Index	Model	Average	SD	Skewness	RMSE
DO (mg l^{-1})	DO(1)4n5000	9.487	2.203	0.817	1.411
	DO(2)2n5000	9.467	2.262	0.814	1.394
	DO(2)4n3000	9.485	2.288	0.788	1.367
BOD (mg l^{-1})	BOD(1)2n3000	4.593	2.519	0.826	1.258
	BOD(2)2n5000	4.913	2.866	0.976	1.587
	BOD(3)4n5000	4.387	2.437	0.604	0.127
SS (mg l^{-1})	SS(1)4n5000	17.759	12.694	1.242	5.694
	SS(2)2n3000	16.983	12.502	1.552	4.652
	SS(2)4n1000	16.989	12.145	1.315	5.312

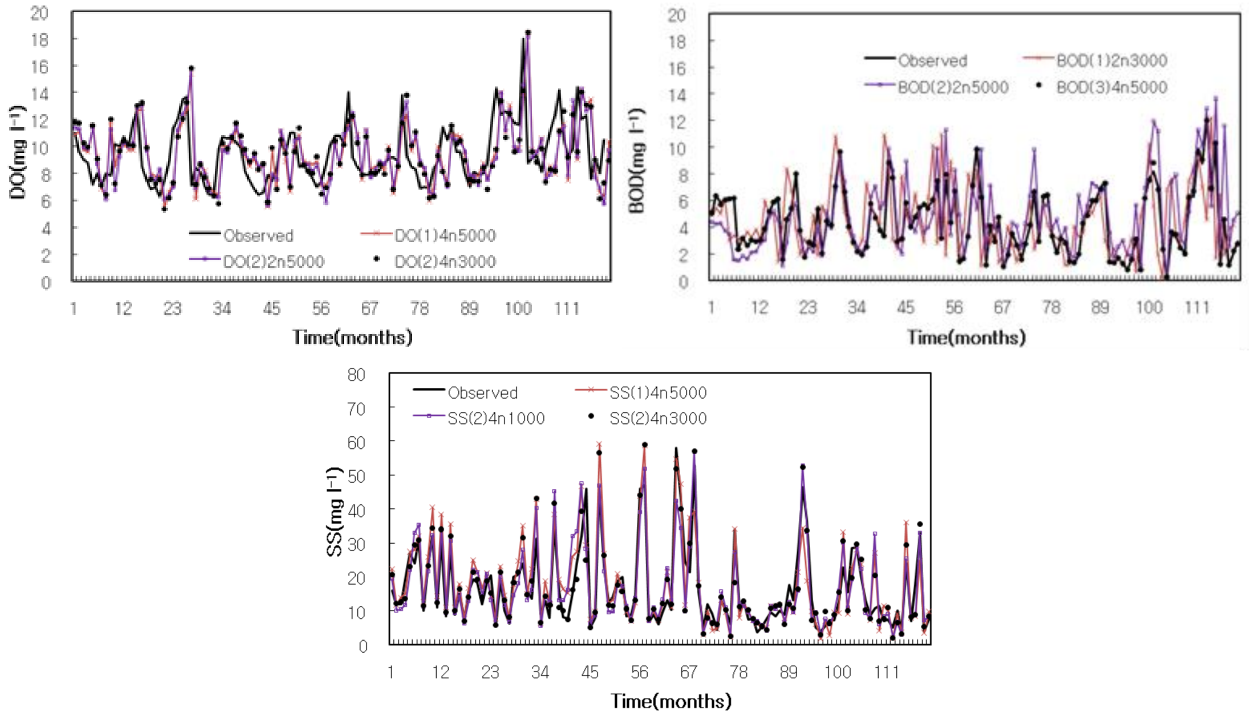

Figure 2. The learning results of neural network models for DO, BOD, and SS (ANN Model)

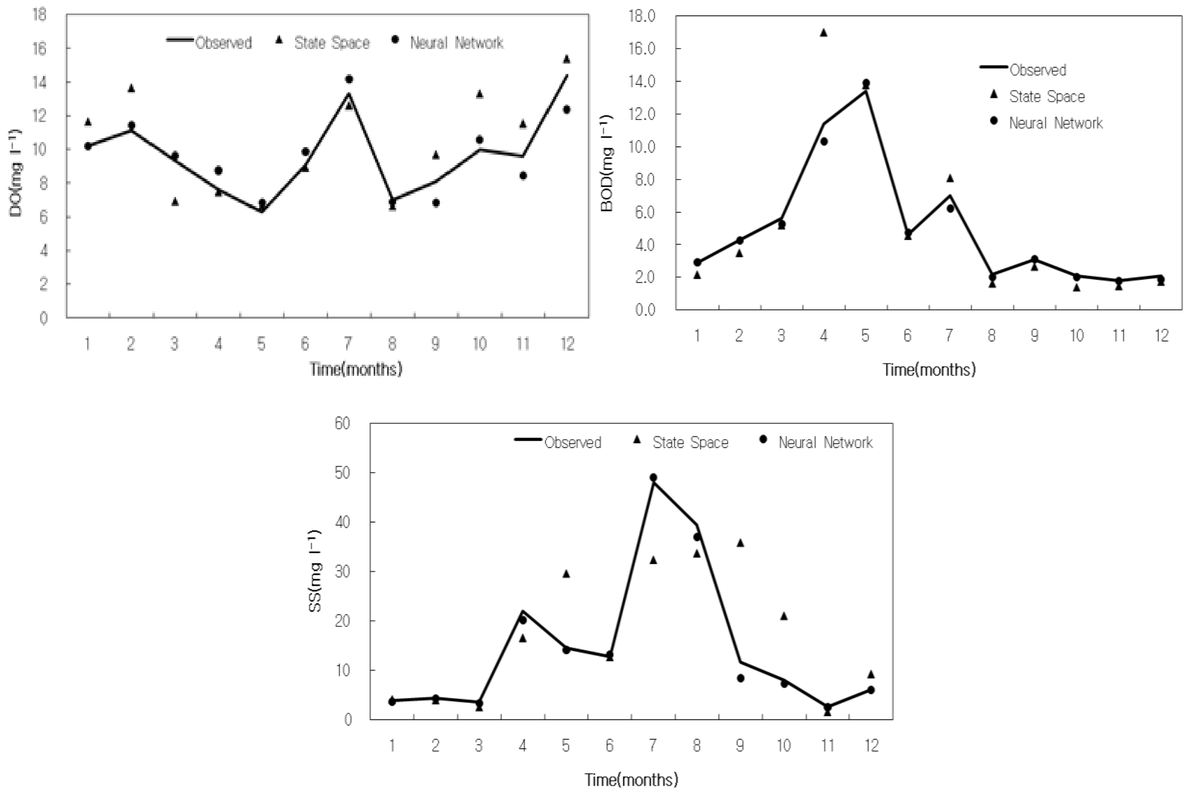

Figure 3. Comparison of observed and estimated water quality(SS and ANN Model) at Mihocheon for DO, BOD, and SS

Table 4. Characteristics of statistics on water quality at Mihocheon

Index	Model	Average	SD	Skewness	RMSE
	Observed	9.667	2.406	0.682	
DO (mg l^{-1})	Neural Network	9.663	2.301	0.429	0.948
	State Space	10.390	3.063	0.119	1.677
	Observed	5.042	3.807	1.430	
BOD (mg l^{-1})	Neural Network	4.863	3.764	1.611	0.434
	State Space	5.263	5.158	1.584	1.711
	Observed	14.725	14.794	1.544	
SS (mg l^{-1})	Neural Network	14.059	14.732	1.677	1.364
	State Space	16.900	13.205	0.278	10.412

In Table 2, the average, standard deviation and skewness values of both observed and estimated variables are in agreement, so the proposed state space model seems to be applicable.

3.3. Neural Network Model

The neural network model is proposed as time series form to forecast water quality using a dataset containing monthly values. Autocorrelation analysis and cross autocorrelation analysis are performed to determine the input shape for water quality forecasting at Mihocheon station. As a result, water temperature and DO show a clear periodicity over a 12-month period. However, BOD shows less correlation with lag time even though there are some constant characteristics. In the case of BOD, discharge is shown to be the most critical factor. In particular, Lag-1 affects Lag+2 for BOD, so the discharge is considered as input data. With respect to SS, there is no particular factor found in the cross autocorrelation analysis.

The construction of the model for water quality forecasting is shown in Table 3 based on the autocorrelation and cross autocorrelation analyses. Mihocheon station is in the first branch of the Geum river basin, and it has a characteristic higher average water quality concentrations compared to those of other learning stations in the main river. The learning results of neural network models, for each proposed one, are shown in Table 3 and Figure 2, The DO(2)4n3000 model is selected as a well-trained model for DO prediction through learning. In the DO(2)4n3000 model, 'DO' indicates the water quality parameter, '(2)' indicates it is the second model of itself in Table 3; '4n' indicates the number of nodes in the hidden layer with 'n' as the number of inputs, and '3000' is the number of iterations. For the BOD model, BOD(3)4n5000 seems to be in good agreement between estimated and observed values on average, SD, and skewness using overflow discharge as input data. For the SS model, SS(2)2n3000 is selected as the optimized model.

3.4. Validation of the presented model

The accuracy in prediction for water quality was examined using the proposed state space model and neural network model. This examination was compared to the data sets, collected from January in 2008 to December in 2008, for each water quality items. The most accurate model was selected as a water quality forecasting model at Mihocheon station. The result of the water quality forecasting was obtained by evaluating and applying an appropriate model from the proposed models. The result is shown in table 4 and figure 3. In table 4, average and standard deviation for BOD forecasting models are in a good agreement. The respective values are 4.863 and 3.764 for the neural network model and 5.263 and 5.158 for state space model. However, For DO and SS forecasting models, the state space model has a tendency to over or underestimate. Therefore, the neural network model was selected as a final water quality forecasting model at Mihocheon station.

Conclusions

In this study, the water quality forecasting model at Mihocheon station is implemented by performing the proposed state space and neural network models, and the suitability of the models is evaluated.

For the state space model at Mihocheon station, the order was determined to be VARMA (3,3) through the result of the forecast results by state vector modeling.

For the neural network model, three sub-categorized models were introduced, including DO, BOD, and SS forecasting models. The DO forecasting model was determined to be DO(2)4n3000, the BOD forecasting model was determined to be BOD3(4n)5000, and the SS forecasting model was determined to be SS(2)2n3000.

The statistical characteristics, RMSE, average, SD, and skewness were considered to verify the accuracy in predicting the water quality through the proposed models in this study. The neural network model was finally selected as a fundamental model for water quality forecasting at Mihocheon station. The neural network model developed for the Mihocheon station can be employed for the development of a water quality emergency management plan so as to ensure sustainable water resource management in the basin.

References

Faruk D.O. (2009), A hybrid neural network and ARIMA model for water quality time series prediction, *Engineering Applications of Artificial Intelligence* 1-9.

Kalman R.E. (1960), A new approach to linear filtering and prediction problems, *Journal of Basic Engineering Ser. D*, **82**, 35-45.

Kenichi, Y., Masaaki H. and Akihiho M. (1997), Novel application of a back propagation artificial neural network model formulated to predict algal bloom, *Water Science and Technology*, **36**, 89-97.

Maier H.R. and Dandy G.C. (1996), The use of artificial neural networks for the prediction of water quality parameters, *Water Resources Research*, **32**, 1012-1022.

Nouh, M. (1996), Simulation of water quality in sewer flow by neural networks, *Hydroinformatics Proceedings of the 2nd International Conference*, 885-891.

Palani, S., Liong S.Y. and Tkalich P. (2008), An ANN applications for water quality forecasting, *Marine Pollution Bulletin*, **56**, 1586-1597.

Yeon S.W., Jun K.W. and Lee H.J. (2009) The improvement of total organic carbon forecasting using neural networks discharge model, *Environment Technology*, **30**, 45-51.

Zhang P.G. (2003), Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing* **50**, 159-175.