Thesis for the Degree of Master of Engineering

# Prediction of Wastewater Treatment Plant Performance using Artificial Neural Network Model



by

Yong-Na Zhao

Department of Environmental Engineering

The Graduate School

Pukyong National University

August 2008

# Prediction of Wastewater Treatment Plant Performance using Artificial Neural Network Model

# 인공신경망 모델을 이용한 하수처리장 공정효율 예측

Advisor: Prof. Byung-Hun Lee

by

Yong-Na Zhao

A thesis submitted in partial fulfillment of the requirements
for the degree of

Master of Engineering

in the Department of Environmental Engineering
The Graduate School
Pukyong National University

August 2008

# Prediction of Wastewater Treatment Plant Performance using Artificial Neural Network Model

A thesis

by

Yong-Na Zhao

Approved as to style and content by:

__________________________
(Chairman) Seok-Jun Yoa

__________________________ __________________________
(Member) Il-Kyu Kim (Member) Byung-Hun Lee

August 2008

# Contents


List of Figures ························ iii

List of Tables ························ v

Abstract ························ vi

I. Introduction ························ 1

II. Method Study ························ 7

2.1 Plant Layout ························ 7

2.2 Artificial Neural Network ························ 10

2.3 Principal Component Analysis ························ 17

2.4 Model Development Process Steps ························ 19

2.4.1 Data collection ························ 20

2.4.2 Data preprocessing ························ 22

2.4.3 Data division ························ 25

III. Results and Discussions ························ 27

3.1 Prediction of Primary Settlement Tank Effluent ····································27

3.1.1 Prediction by an artificial neural network model··························27

3.1.2 Prediction by a hybrid artificial neural network model ·················34

3.2 Prediction of Secondary Settlement Tank Effluent ·································41

3.2.1 Prediction by a hybrid artificial neural network model ·················41

IV. Conclusions ······························································· 51

References······································································· 54

Acknowledgments ························································· 61

# List of Figures

Fig.2.1 A schematic diagram of the South Wastewater Treatment Plant··············· 9

Fig.2.2 Schematic of a multi-layer ANN structure ···········································11

Fig.2.3 Steps of the model development process ··············································20

Fig.3.1.1 General scheme of the prediction ANN stages ··································29

Fig.3.1.2 Prediction of BOD with ANN in the primary settlement tank···············31

Fig.3.1.3 Prediction of COD with ANN in the primary settlement tank···············32

Fig.3.1.4 Prediction of SS with ANN in the primary settlement tank··················32

Fig.3.1.5 Prediction of TN with ANN in the primary settlement tank ·················33

Fig.3.1.6 Prediction of TP with ANN in the primary settlement tank··················33

Fig.3.1.7 General scheme of the prediction ANN&PCA stages ··························36

Fig.3.1.8 Prediction of BOD with ANN&PCA in the primary settlement tank·····38

Fig.3.1.9 Prediction of COD with ANN&PCA in the primary settlement tank·····38

Fig.3.1.10 Prediction of SS with ANN&PCA in the primary settlement tank·······39

Fig.3.1.11 Prediction of TN with ANN&PCA in the primary settlement tank ······39

Fig.3.1.12 Prediction of TP with ANN&PCA in the primary settlement tank·······40

Fig.3.2.1 Scheme of the proposed hybrid ANN model procedure························44

Fig.3.2.2 Prediction of BOD with hybrid ANN in the secondary settlement tank 47

Fig.3.2.3 Prediction of COD with hybrid ANN in the secondary settlement tank 47

Fig.3.2.4 Prediction of SS with hybrid ANN in the secondary settlement tank ····48

Fig.3.2.5 Prediction of TN with hybrid ANN in the secondary settlement tank····48

Fig.3.2.6 Prediction of TP with hybrid ANN in the secondary settlement tank ····49

# List of Tables

Table 2.1 Statistical properties of influent composition ······································23
Table 2.2 Statistical properties of PST composition ···········································23
Table 2.3 Statistical properties of AT composition ·············································24
Table 2.4 Statistical properties of SST composition ···········································24
Table 3.1.1 Model architectures and prediction results by ANN ··························29
Table 3.1.2 Eigenvalues of the correlation matrix I for BOD, COD, SS ··············37
Table 3.1.3 Eigenvalues of the correlation matrix II for TN, TP ···························37
Table 3.1.4 Model architectures and prediction results by ANN&PCA················40
Table 3.2.1 Inputs and output targets for the ANN ··············································44
Table 3.2.2 ANN architecture and prediction result ·············································46

# Prediction of Wastewater Treatment Plant Performance using Artificial Neural Network Model

Yong-Na Zhao

Department of Environmental Engineering, the Graduate School,
Pukyong National University

## Abstract

Proper operation and control of municipal wastewater treatment plants is important in producing an effluent which meets quality requirements of regulatory agencies and in minimizing detrimental effects on the environment. Predicting the plant water quality parameters using conventional experimental techniques is also a time consuming step and is an obstacle in the way of efficient control of such processes. For control and automation of the plant treatment processes, lack of reliable on-line sensors to measure water quality parameters is one of the most important problems to overcome. And the accuracy of existing hardware sensors is also not sufficient. This paper deals with the development of software sensor techniques that estimate the target water quality parameter from other water quality parameters.

Here an artificial neural network (ANN) and a hybrid ANN model, combining with principal component analysis (PCA), both of them were applied

to predict the wastewater effluent quality parameters, biological oxygen demand (BOD), chemical oxygen demand (COD), suspended solids (SS), total nitrogen (TN), total phosphorous (TP) of the primary settlement tank based on past information. The PCA was used to synthesize the input water quality parameters in order to reduce the dimension of the inputs. And the back-propagation feed-forward neural network (FBNN) was chosen to model the wastewater treatment plant through this study, which is the South Wastewater Treatment Plant (WWTP) at Busan City, Korea. The tan-sigmoid function was used as activation function to transfer signal at the neural network. And the Levenberg-Marquart algorithm was used as learning algorithm to train neural network. All the 364 data sets, which were collected from the plant during 2005, 200 data sets and other 164 data sets, were used for training and validation, respectively. The hybrid ANN&PCA models for prediction of water quality parameters was also used in the primary settlement tank (PST) effluent, comparing with the prediction results by ANN. Following the prediction of first physical and chemical process in the wastewater treatment, it is the biological wastewater treatment process, which is commonly used to treat municipal and industrial wastewaters and so important process in the treatment. Special attention has been paid to biological processes modeling, both for wastewater treatment and sludge stabilization processes. In the prediction of secondary settlement tank (SST) effluent presented here is another hybrid ANN model, which is using some data from the activated sludge model (ASM) simulator in order to strength and advance ANN model. The hybrid ANN techniques show an enhancement of prediction capability and reduce the over-fitting problem of neural networks. The results showed that the hybrid ANN technique can be used to extract information from noise data and can provide more accurate predictions of the primary and secondary settlement tank effluent stream and then further to describe the nonlinearity of complex wastewater treatment.

**Key words:** wastewater treatment plant, artificial neural network, principal component analysis, activated sludge model

# I. Introduction

With increasing stringent regulation of the effluent quality, process monitoring and concentration control have become more important. Serious environmental and public healthy problems may result from improper operation of a WWTP, as discharging contaminated effluent to a receiving water body can cause or spread various diseases to human beings. Accordingly, environmental regulations set restrictions on the quality of effluent that must be met by any WWTP. However, due to the complex biological characteristic of the activated sludge process, it is difficult to measure water quality parameters using on-line sensors. Many parameters required for control can't be measured on-line as yet, and the reliability of existing on-line sensors is not sufficient for process automation. Although some parameters can be measured by laboratory analyses, a significant time delay in a range of tens of minute to few hours is usually unavoidable. It is normally too late to achieve well-timed adaptive process control accommodating influent fluctuation and other disturbances, especially for advanced wastewater treatment requiring more precise and timely controls. Two ways of dealing with these problems which have been discussed among researchers are to develop new methods for monitoring the desired parameters and to develop software sensors based on information from the existing on-line sensors. Safer operation and control of a WWTP can be achieved by developing a modeling tool for predicting the plant performance based on past observations of a certain key product quality parameters. Wastewater treatment plants involve several complex physical, chemical and biological processes. Often these processes exhibit non-linear behaviors which are difficult to describe by linear mathematical models. In addition, the variability of the influent characteristics, in terms of composition, strength and flow rates, might influence model parameters, and consequently operational control, significantly. Therefore, modeling a WWTP is a difficult task and most of the available models are just approximations based on, probably severe, assumptions.

In the recent decades, both domestic and overseas scholars have studied on the comprehensive evaluation methods of WWTP processes. A lot of evaluation methods and evaluation parameters were put forward, such as single-factor method, parametric methodology, graphic method and others. After 1980s, with the rapid progress of computers, modern mathematics theories were applied into the water environmental evaluation, and more and more complex statistical methods were used. In recent years, the techniques and algorithms of mathematics methods applied to WWTP performance evaluation integrated with computer technology are continuously appearing. These approaches and techniques are gathered together with a common purpose: to find a solution (usually in the form of models) to a wide variety of problems (such as pattern recognition, systems control, prediction, optimization and others) which share some characteristics: the nature of the problem is usually non-linear, and data is disturbed by noise, imprecision or uncertainty, and is often missing. Moreover, the source of these data can be very heterogeneous, ranging from discrete to continuous variables, which can also be scalar, vectorial combination, etc, and include a spatial or temporal component. The most common theories and methods employed make use of fuzzy mathematics, stochastic modeling, genetic algorithms, probabilistic reasoning, grey system, rough set theory, artificial intelligence and other methods. The artificial neural network are also applied this field as a part of artificial intelligence. Owing to their high accuracy, adequacy and quite promising applications in engineering, artificial neural network can be used for modeling some WWTP processes. Artificial neural network (ANN) has been proved to be able to model nonlinear systems. ANN modeling approaches have been embraced enthusiastically by practitioners in water resources, as they are perceived to overcome some of the difficulties associated with traditional statistical approaches. Artificial neural network is a statistical tool for data analysis. In the words of Sarle (1994), users of ANN want their networks to be black boxes requiring no human intervention-data in, predictions out. More recently, researchers have examined ANN models from a statistical perspective (e.g. Hill et al., 1994; Sarle, 1994; Wisra and Warner, 1996). Such studies indicate that certain models obtain when ANN geometry, connectivity and parameters changed are either equivalent, or very close to, existing statistical models. Consequently, some neural network models are not really new inasmuch as they represent variations on common statistical themes. Although ANN models are not significantly different from a number of standard statistical models, they are extremely valuable as they provide a flexible way of implementing them. Model complexity can be varied simply by altering the transfer function or network architecture. In addition, unlike some statistical models, ANN models can be extended easily from univariate to multivariate cases. However, as the number of different models that can be generated using ANNs is so vast, many have not yet been examined from a statistical perspective. As pointed by White, "the field of statistics has much to gain from the connectionist literature. Analyzing neural learning procedures poses a host of interesting theoretical and practical challenges for statistical method; all is not cut and dried." Until recently, there has been little interaction between the neural network and statistical communities and ANN and statistical models have developed virtually independently.

As preceding mentioned, artificial neural network (ANN) is a black-box approach that depends only on the observed values, which attracts many researchers' attention. Cote et al. (1995) used a two-step procedure to improve the accuracy of the mechanistic model of the activated sludge process. The first step is the parameter optimization of the mechanistic model using a least squares regression analysis based on a large set of experimental data for five key process variables. The second step is to use feed-forward neural network models simulating the prediction errors of the mechanistic model. Zhu et al. (1998) proposed a time-delay neural network modeling method for predicting the effectiveness of a biological treatment process. A procedure has been developed by Cote et al. (1995) using a neural network to improve the accuracy of an existing mechanistic model of the activated sludge process. Oliveira-Esquerre et al. (2002) obtained satisfactory prediction of the BOD in the output stream of a local biological wastewater treatment plant for the pulp and paper industry in Brazil. Hong et al. (2003) used the Kohonen Self-Organizing Feature Maps (KSOFM) neural network to analyze the multidimensional process data and to diagnose the inter-relationship of the process variables in a real activated sludge process. Hamed et al. (2004) developed ANN models to predict the performance of a WWTP based on past information. A hybrid neural network approach, which combines mechanistic and neural network models, has also been used to model wastewater treatment process. Krovvidy and Wee (1992) developed an intelligent hybrid system, combining inductive learning, artificial neural network approached and case-based reasoning for a wastewater treatment plant. Capodaglio et al. (1991) analyzed the input and output of the activated sludge system by applying stochastic models and artificial neural system models to treatment plant data. Wilcox et al. (1995) utilized a neural network simulation to classify potentially damaging events during the anaerobic digestion process with an on-line bicarbonate alkalinity sensor.

In this study, the effluent of the primary settlement tank and the secondary settlement tank was predicted here by artificial neural network. In the first section, it is the primary settlement tank effluent prediction, which is the physical and chemical process in the wastewater treatment. This section focuses on the development of a hybrid ANN as a software sensor in the wastewater treatment processes, which has been successfully applied in modeling wide range of non-linear systems, especially chemical/physical engineering processes. Here it is the back-propagation feed-forward neural network (FBNN) that was developed to predict the performance of the South Wastewater Treatment Plant, based on the available historical data. As well known, ANN predicts values of unmeasured target parameters using the correlation between measured and target parameters. It is possible to reduce the measured parameters' dimension for prediction of unmeasured parameters using PCA. And also PCA can minimize the influence of noise and outlier that is much harmful to ANN training. Therefore, a hybrid ANN model, back-propagation feed-forward neural network (FBNN) and principal component analysis (PCA) was also applied to predict the effluent wastewater quality parameters, comparing the prediction by ANN. Theoretically, the hybrid ANN model, combining with principal component analysis, would get better prediction results than the ones by ANN. The results that got from the ANN model and the hybrid ANN&PCA model proposed here also exemplify the hybrid ANN technique is a strong tool to predict the wastewater quality parameters. In the second section, it is the secondary settlement tank effluent prediction, which is the biological process in the wastewater treatment. This process includes aeration tank and secondary clarifier, which are the most important step and more complex in the treatment. Special attention has been paid to biological processes modeling, both for wastewater treatment and sludge stabilization processes. In this section, some data, getting from the activated sludge model (ASM) that is a mathematical model of the activated sludge process and done through a task group pointed by the International Water Association, was used here in order to strength and advance the artificial neural network model.

The main object of this research is to develop an estimation model that provides accurate predictions of BOD, COD, SS, TN and TP of primary and secondary settlement tank at a WWTP. This thesis is organized in four sections. In section 2 some concepts of the ANN and ANN hybrid model and data preprocessing are briefly described. The results obtained are given in section 3. Finally, section 4 presents the conclusion.

# II. Method Study

## 2.1 Plant Layout

Wastewater treatment plants typically have two principal stages: first, the primary stage, which includes the bar racks, grit chamber and primary settlement tank whose objective is the removal of the organic load and solids in the wastewater to a degree of 30%-50%; and second, the secondary stage, whose objective is the biological treatment of the organic load and which is essential when a higher degree of treatment is required. In our study, some sensor models were applied to the South Wastewater Treatment Plant, at Korea, Busan City, the developed ANN and the ANN&PCA hybrid mode for physical and chemical process and the ANN with data materials from activated sludge model simulator for biological process, respectively. A schematic diagram of the plant, the South Wastewater Treatment Plant, is shown in Fig. 2.1. The crude sewage from different pumping stations is collected and screened for floating debris and removal of grit is carried out by the grit collector and grit chambers. Primary settlement tanks (PST) are utilized to settle 50%–70% of the solids. Settled solids are scrapped down in the hoppers of the PST with the help of mechanical drive scrappers. These settled solids are removed by the Hydro Valves which open in the Sludge Thicker Tank. And then this sludge is removed to the Sludge Consolidation Tank. Aerobic bacteria are activated by aeration and mixing with activated sludge to reduce the volume of mixed liquor. Primary treated effluent is mixed with the returned activated sludge from the secondary settlement tank and uniformly distributed in channels for aeration with the help of mechanically driven aerators. Mixed liquor out of the aeration tank is made to settle in the secondary settlement tanks. The resulting stream, designated as final effluent, flows down into the wet well. The return-sludge from secondary settlement tank is returned to the aeration tank. And the waste-sludge from secondary settlement tank is removed to the centrifugal sludge tank and then mixed into the sludge consolidation tank with the waste-sludge from the primary settlement tank. The last step is the sludge treatment that is sludge digestion first and then sludge dewatering.

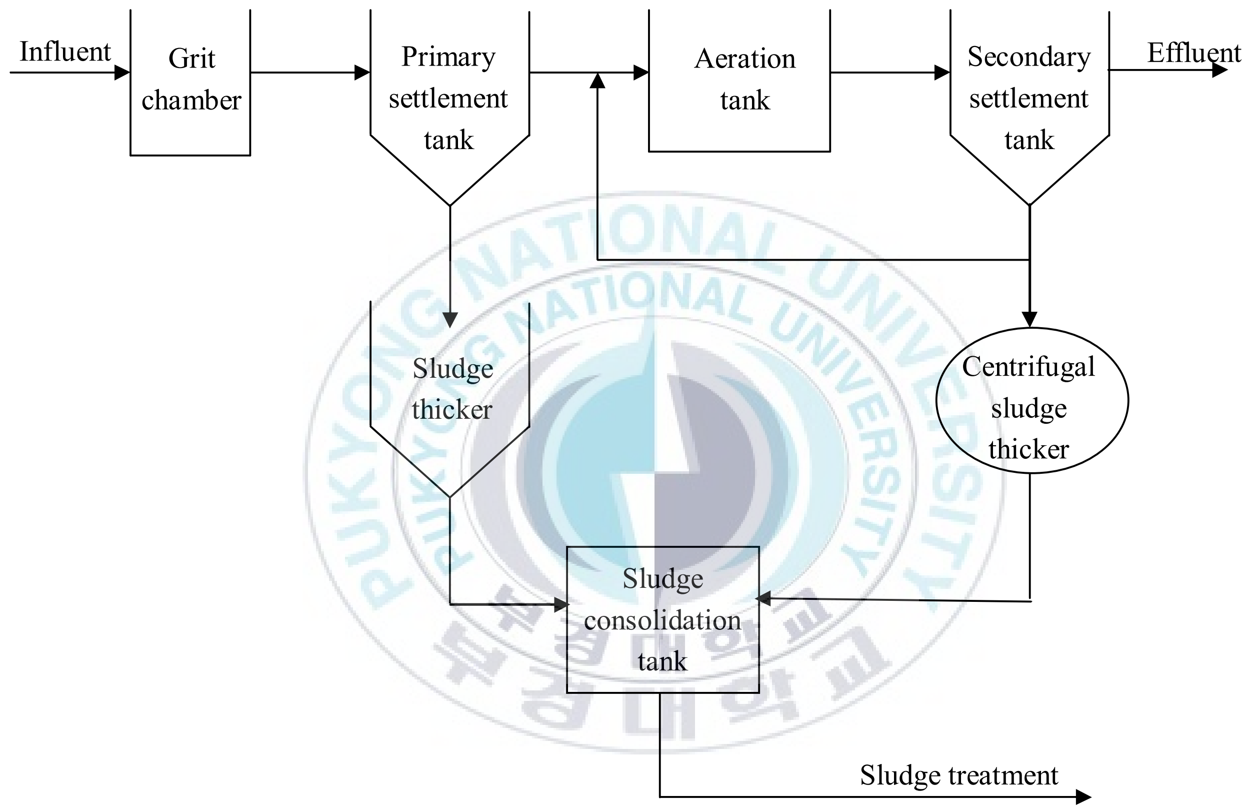


**Fig. 2. 1 A schematic diagram of the South Wastewater Treatment Plant**

## 2.2 Artificial Neural Network

Artificial neural network (ANN) is a black-box approach that depends only on the observed values. It can model complex nonlinear system. The ANN modeling approach simulated the operation features of human nervous system. Many simple computational elements called artificial neurons that are connected by variable weights are used. Artificial neural network is a statistical tool for data analysis, taking into account factors such as data pre-processing, the determination of adequate model inputs and a suitable network architecture, parameter estimation and model validation. In addition, careful selection of a number of internal parameters is also required. A typical neural network model consists of three independent layers: input, hidden and output layers. Each layer is comprised of several operation neurons. Input neurons receive the values of input variables that are fed to the network and store the scaled input values, while the calculated results in output layer are assigned by the output neurons. The hidden layer performs an interface to fully interconnect input and output layers. Each neuron is connected to every neuron in adjacent layers before being introduced as inputs to the neuron in the next layer by a connection weight, which determines the strength of the relationship between two connected neurons. Each neuron sums all of the inputs that it receives and the sum is converted to an output value based on a predefined activation or transfer function.

The neural network used here, as shown in Fig. 2.2, is a multi-layered, supervised feed-forward neural network with back-propagation algorithm (FBNN). Back propagation artificial neural network is one of the most widely used neural networks. Typical FBNN structure consists of one input layer, one (or multiple) hidden layers and one output layer. Each layer could have a number of nodes (processing elements), which are connected linearly by weights to the nodes in the neighboring layers. The training process adjusts weights to minimize the error between the measured output and the output produced by the network. Through this adjustment, the neural network learns the input-output behaviors of the system. Actually, the BP artificial neural network is a kind of highly non-linear mapping from input to output. In this work, the supervised FBNN neural network model with Levenberg-Marquardt (LM) algorithm was implemented in MATLAB, which is a common scientific programming language.

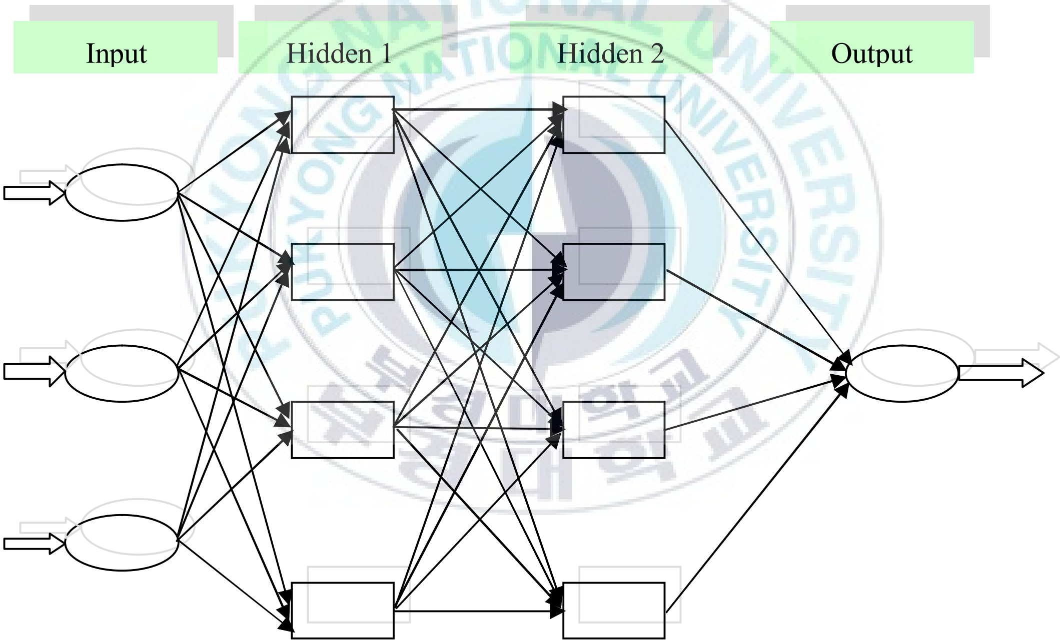


**Fig. 2. 2 Schematic of a multi-layer ANN structure**

And we give some detail information of artificial neural network as following:

(1) Learning rule:

A learning rule is defined as a procedure for modifying the weights and biases of a network. The learning rule is applied to train the network to perform some particular task. Learning rules in the MATLAB toolbox fall into two broad categories: supervised learning and unsupervised learning. In un-supervised learning, the artificial neural network is not provided with the correct results during training, where the weights and biases are modified in response to network inputs only. Un-supervised artificial neural network usually performs some kind of data compression, such as dimensionality reduction or clustering. In supervised learning, the correct results (target values, desired outputs) are known and are given to the artificial neural network during training so that the artificial neural network can adjust its weights to try to match its outputs to the target values. The learning rule is used to adjust the weights and biases of the network in order to move the network outputs closer to the targets with the goal of minimizing the error between the model-predicted value and the actual value of the output variable by modifying the weights between neurons according to a learning rule. Therefore, here the supervised learning is a perfect choice.

(2) Feed-forward neural network:

Feed forward back propagation neural network is a single-direction network. In a feed-forward neural network, feed back loops are absent. Information is processed in a forward manner only from input to output, and thus it always gives the same output result for the same input. The output of a feed-forward neural network with one hidden layer and one output neural network is given by

$$\bar{Y}_t = f_o[\sum_{j=1}^{HN} WO_j \times f_h\left(\sum_{i=1}^{m} WH_{ij} \bullet X_{it} + b_j\right) + b_o]$$

where,

$WH_{ij}$: the weight of the link between the $i$th input and the $j$th hidden neuron;

$m$: the number of input neurons;

$WO_j$: the weight of the link between the *j*th hidden neuron and the output neuron;

$f_h$: the hidden neuron activation function;

$f_o$: the output neuron activation function;

$b_j$: the bias of the *j*th hidden neuron;

$b_o$: the bias of the output neuron,;

HN: the number of hidden neurons;

(3) Back-propagation:

Back-propagation is the generalization of the Widrow-Hoff learning rule to multiple-layer network and nonlinear differentiable transfer functions. The back-propagation algorithm is a basic training optimization procedure for multi-layer networks, in which the weights are moved in the direction of the negative gradient and a mean square error performance index is minimized. In a back-propagation algorithm, the prediction error is generated at the output layer of neurons and then propagates backwards through the network. Each network has a learning rule that defines how the weights are modified to minimize prediction and the back-propagation algorithm is the most common rule used in process modeling.

(4) Levenberg-Marquardt:

The default FBNN training algorithm is the Levenberg-Marquardt (LM) algorithm. This is the fastest and the most robust method in the toolbox, but it can use large amounts of memory. The Levenberg-Marquardt used here may be considered to be a hybrid between the classical Newton and steepest descent algorithm. It is a modification of the classical Newton algorithm and its behavior is similar to that of gradient descent methods.

(5) Layer and neuron:

ANN consists of an input layer, output layer and hidden layer between the input and output layers. The hidden layer is one or multiple. The number of neurons in the input layer is usually equal to the number of input variables. The number of output layer neurons is usually the same as the target variable number. The number of neurons in the hidden layer is determined by trial and error. Larger numbers of neurons in the hidden layer give the network more flexibility because the network has more parameters it can optimize. But if the hidden layer is set too large, it might cause the problem to be under-characterized and optimize too much parameter to constrain them. So the perfect number of neurons in the hidden layer is according to the optimization performance.

(6) Learning function:

Here it is the learngdm that used in this paper, which is gradient descent with momentum weight/bias learning function. Gradient descent with momentum allows a network to respond not only to the local gradient, but also to recent trends in the error surface. With momentum a network can slide through a shallow local minimum. This learning function depends on two training parameters: learning rate (lr) and momentum constant (mc).

(7) Activation function:

The activation function defines how the net input received by a neuron is combined with its current state of activation to produce a new state of activation. These transfer functions that are most commonly used are linear transfer function (purelin) and sigmoid type functions such as the logistic (log) and hyperbolic tangent functions (tan). The output layer with the three transfer functions lets the network produce values in the range $-\infty$ to $+\infty$, 0 to 1, and -1 to +1, respectively. Generally, the same transfer function is used in all layers.

In this work, the tan-sigmoid activation function was used with the following formula:

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

(8) Error function:

The error function is the function that is minimized during training. The representative error functions are root mean squared error (RMSE), mean squared error (MSE) and correlation coefficient (R). Here the mean squared error (MSE) function was used to display the performance of the ANN model through the Matlab toolbox. We could know the difference between normalized target values and model predicted values through the MSE. The correlation coefficient R, which is also got directly from the Matlab program, was used to show the correlation between normalized target values and model predicted values. The last root mean squared error (RMSE) function shows the difference between the measured values and the model predicted values. Here the analysis of the ANN and ANN&PCA hybrid model performance will be hampered by the large standard deviations for R and MSE.

$$RMSE = \sqrt{\sum_{n=1}^{N} (Y_M - Y_P)^2 / N}$$

$$MSE = \sum_{n=1}^{N} (Y_M - Y_P)^2 / N$$

where,

$Y_M$: model predicted value;

$Y_P$: measured value;

$N$: data points;

(9) Training:

The function train carries out a loop of calculation, proceeding through the specified sequence of inputs, calculating the network output, error, and network adjustment for each input vector in the sequence as the inputs are present. Training proceeds through repeated presentation of data patterns to the ANN and subsequent weight modification until the prediction error is sufficiently small, as defined by the user, or until a maximum number of iterations have been reached. The training consisting of the following steps:

1) initialize of the connection weights of the network;

2) select a sample from the training data set as the input of the network;

3) calculate the output value or output vector of the network;

4) calculate the errors of the network;

5) adjust the connection weights from the output layer back to the input layer;

6) repeat 3), 4), 5) until the errors are acceptable;

(10) Validation:

Once the training phase has been completed, the performance of the trained network has to be validated on an independent data set. Because the function train doesn't guarantee that the resulting network does its job and it is not sufficient to evaluate a model by testing the prediction capability for the data used for model construction. Verification of prediction capability for the data that is not used for setting up the model has to be followed. In addition, the decision of initial weights is an important step of using back-propagation neural network program during the training and validation phases. In the same back-propagation neural network, the connection weights will be the same when the training is accomplished with the same initial values. And different initial weights will lead to different connection weights. It is best to training several times with different initial weights until the errors of the network are acceptable.

## 2.3 Principal Component Analysis

As in any prediction/forecasting model, the selection of appropriate model inputs is extremely important. However, in most ANN applications, little attention is given to this task. The main reason for this is that ANNs belong to the class of data driven approaches, whereas conventional statistical methods are model driven. Though ANN has the ability to determine which model inputs are critical, presenting a larger number of inputs to ANN model and relying on the network to determine the critical model inputs usually increases network size and bad results.

Principal component analysis (PCA) is a conventional statistical method for data preprocessing. It's also a factorial method because the number of the variables reduction is not effected by a simple selection, but constructing new synthetic variables obtained by combining initial variables. The PCA proceeds by reduction of the variables space dimension while eliminating correlations between initial variables.

When $P$ quantitative variables $X_1, X_2 \ldots X_p$ are correlated, the expressed information is characterized by some redundancy. PCA extracts a non-redundant list of $K$ new variables or factors $Y_1, Y_2 \ldots Y_k$ (k≤p) from a redundant list of $p$ variables $X_1, X_2 \ldots X_p$. The non-redundancy condition of the factor list $Y_1, Y_2 \ldots Y_k$ (k≤p), each factor explains a part of the variability observed on the original variables. It makes it possible to summarize the information contained in the $p$ initial variables thanks to a number of factors lower than $p$. This represents an appreciable information compression. Any correlation matrix $R$ ($p \times p$) can be analyzed and decomposed with PCA. And then we could choose the eigenvectors to make the new variables or factors, namely PCs, according to the larger eigenvalues of the correlation matrix.

Here the calculation process of this method is described in detail as follows.

Step 1: get some data sets; take 3 water quality parameters as example, $X_1$*, *$X_2$*, and* $X_3$. Step 2: subtract the mean; each parameter $X_i$ minus $\bar{X}_i$ and then these are used as new vectors for matrix *A*. For PCA to work properly，step 2 is necessary and this produces a data set whose mean is zero. Step 3: calculate the covariance matrix. Through this step we could get the covariance matrix *R* from the matrix *A* through covariance formula, as follows:

$$\mathrm{var}(x) = \frac{\sum_{i=1}^{n}(x_i - \overline{x_i})(x_i - \overline{x_i})}{(n-1)}$$

$$\mathrm{cov}(x, y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x_i})(y_i - \overline{y_i})}{(n-1)}$$

Where,

$x_i$, $y_i$: vectors of matrix *A*;

$\bar{x}_i$, $\bar{y}_i$: mean values of $x_i$, $y_i$ ;

And the covariance matrix is available in the following type

$$R = \begin{pmatrix} \mathrm{cov}(x,x) & \mathrm{cov}(x,y) & \mathrm{cov}(x,z) \\ \mathrm{cov}(y,x) & \mathrm{cov}(y,y) & \mathrm{cov}(y,z) \\ \mathrm{cov}(z,x) & \mathrm{cov}(z,y) & \mathrm{cov}(z,z) \end{pmatrix}$$

Step 4: calculate the eigenvalues and eigenvectors of the covariance matrix; since the covariance matrix is square, we can calculate the eigenvalues and eigenvectors for this matrix. These are rather important as they tell us useful information about the data sets. Step 5: form a feature vector and derive new data sets; here is where the notion of data compression and reduced dimensionality comes into it. The eigenvector with the highest eigenvalue is the principal component of the data set. In this example, the eigenvectors with the larger eigenvalues are the ones that point down the middle of the data, which is the most important relationship between the data dimensions. Then choose the eigenvectors with the larger eigenvalues to form a feature vector according as the predefined 95% selection principal. At last, the new data set is available by taking the transpose of this feature vector and multiplying it on the left of the original mean-adjusted data set transposed. This new data sets are just the PCs that we need to train and evaluate artificial neural network model. With the biggish eigenvalues principal, we ignore and left out some components of less significance. Then the final data sets will have fewer dimensions than the original not losing much information.

These PCs could capture the main trend of the data. And PCA can minimize the influence of noise and outlier, which will cause the over-fitting problem easily for ANN model. So the hybrid model ANN&PCA could get a better prediction theoretically.

## 2.4 Model Development Process Steps

A number of steps were carried out during the model development process. These are shown schematically in Fig. 2.3. The ANN model shares many of the attributes and steps relevant to other models of the non ANN types.

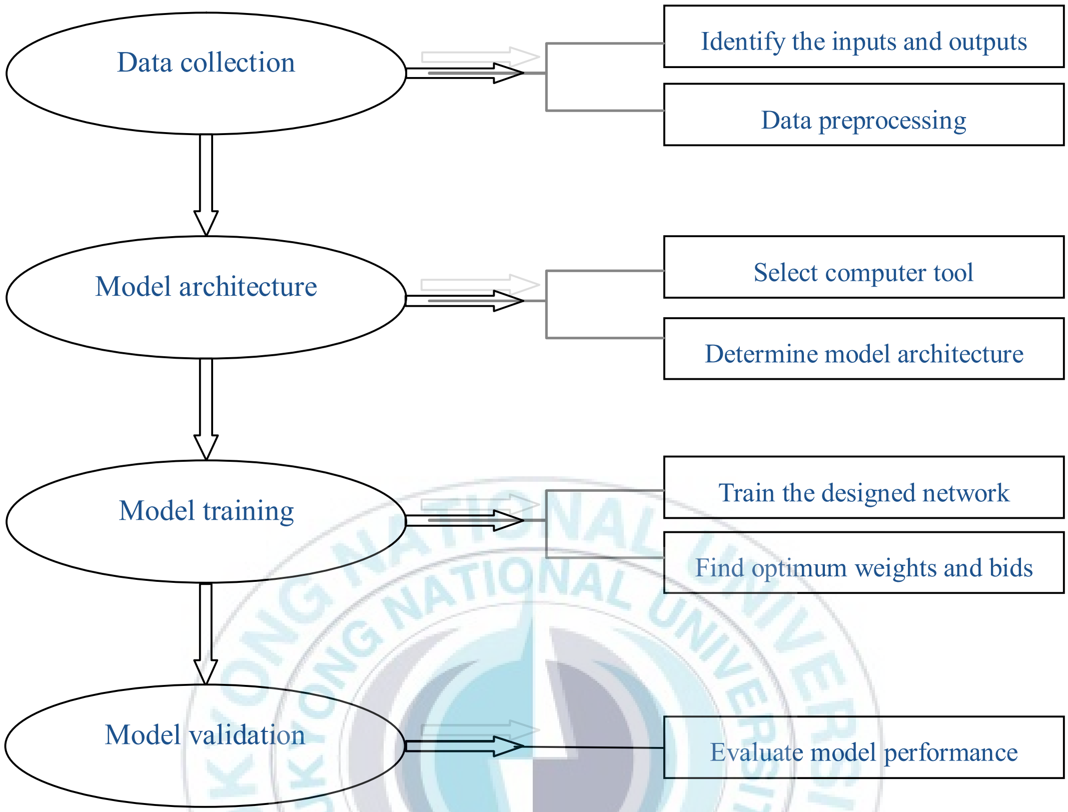


**Fig. 2. 3 Steps of the model development process**

### 2.4.1 Data collection

To develop successful ANN models of wastewater treatment processes, careful attention must be paid to the details of data collection and analysis, such as the format and reliability of the data. In collecting data, several factors need to be considered. First, the availability of the data must be ascertained. For data availability, the variables for which historical data exist, the time frame of historical measurements, and the frequency of data measurements must all be determined. The format and reliability of the data are also key considerations in data collection. Historical data can originate from grab-samples or real-time measurements, and measurements can be discrete or aggregated from a number of samples. The reliability of the data should be ascertained through an examination of quality assurance and quality control protocols. Finally it is of considerable importance to note any process changes that may have been implemented during the time frame for which data are available. With the above considerations in mind, it is important to delineate a number of guidelines for selecting data to be used in ANN modeling. First and foremost, data for each of the variables known or suspected to affect the process being modeled must be available. The quantity of data required to develop a model is site specific and is affected by seasonal fluctuations and the frequency of process upsets. As such, it is important to ensure that the data are fully representative of the range of conditions that can be expected during periods of routine and upset operations. As a general guideline, at least one full cycle of data must ne available to ensure a representative data set. With respect to the format of the data, the variability of the process as well as data availability will dictate whether to use hourly data, daily average, or some other frequency for each of the variables. Successful process models can often be made using the daily average or some daily percentile value of each of the model variables. With respect to the effect of major process change on data collection, data collected prior to major process conditions. Finally to maintain the integrity of the data set, appropriate quality assurance and quality control protocols for the collection of each model variable must be in place.

The available data for the South Wastewater Treatment Plant were carefully investigated. It was decided to relate the outputs of the primary settlement tank (PST) effluent to the inputs of the influent and PST influent stream, and the outputs of the secondary settlement tank (SST) effluent quality to the aeration tank (AT) conditions. Therefore, measurements of the Q, PH, T, BOD, COD, SS, TN, TP in the influent, PST, AT and SST stream were collected over a one-year period. This period was satisfactory as it covers all probable seasonal variations in the studied variables. The schematic diagram of the plant is shown in Fig. 2.1.

### 2.4.2 Data preprocessing

Once an appropriate historical data set has been selected, it should be fully characterized and subjected to a comprehensive statistical analysis. Data characterization involves a qualitative assessment of hourly, daily, and seasonal trends of each potential model variable. The statistical analysis involves the determination of measures of central tendency, measures of variation, and a percentile analysis, as well as the identification of outliers, erroneous entries, and non-entries for each data available. In combination, the data characterization and statistical analysis help to identify the boundaries of the study domain as well as potential deficiencies in the data set.

ANN modeling also requires data of good quality that reflect the dynamic of target system accurately. However, it's often hard to get from real wastewater treatment processes, so noise and measurement error is the main obstacle in setting up an ANN model based on the raw measured data. Therefore, data preprocessing is necessary for ANN models to get a better prediction. Each data pattern should initially be examined for erroneous entries, outliers and blank entries because of the transcription or transposition or experimental errors or human errors. Data refining was performed on the raw experimental data by excluding all outliers which were unusual points.

First, clip noise data by removing measurements that were not within the range of $\pm 3\sigma$, namely Upper Limit Control (ULC) and Low Limit Control (LCL):

$$UCL = X + A\sigma$$

$$LCL = X - A\sigma$$

Where,

$X$: mean value;

$\sigma$: standard deviation;

$A$: constant (here 3);

Table 2.1~Table 2.4 shows the water parameters' statistical property of Influent, Primary Settlement Tank, Aeration Tank and Secondary Settlement Tank:

**Table 2. 1 Statistical properties of influent composition**

| | Q | Tem | PH | BOD | COD | SS | TN | TP |
|---|---|---|---|---|---|---|---|---|
| | $m^3/d$ | ℃ | | mg/l | mg/l | mg/l | mg/l | mg/l |
| Min | 241920 | 5.9 | 6.9 | 45.7 | 15.7 | 37 | 11 | 0.8 |
| Max | 440780 | 27 | 7.9 | 123.6 | 74.3 | 142 | 45 | 4.8 |
| Mean | 326084 | 17.7 | 7.3 | 93 | 45.5 | 93 | 26 | 2.7 |
| Std | 37007 | 5.9 | 0.13 | 12 | 8.0 | 17 | 4.8 | 0.4 |

**Table 2. 2 Statistical properties of PST composition**

| | Tem | PH | BOD | COD | SS | TN | TP |
|---|---|---|---|---|---|---|---|
| | ℃ | | mg/l | mg/l | mg/l | mg/l | mg/l |
| Min | 2.7 | 6.7 | 48.6 | 22.5 | 51 | 12.5 | 12.5 |
| Max | 26.8 | 7.9 | 158 | 105.4 | 192 | 48.3 | 48.3 |
| Mean | 17.7 | 7.3 | 109 | 58 | 118 | 28.8 | 28.8 |
| Std | 5.8 | 0.1 | 18.7 | 13 | 25.8 | 5.7 | 5.7 |

**Table 2. 3 Statistical properties of AT composition**

| | Tem | PH | DO | MLSS | SVI | SV |
|---|---|---|---|---|---|---|
| | ℃ | | mg/l | mg/l | | |
| Min | 9 | 6.3 | 0.7 | 1184 | 77.4 | 10.6 |
| Max | 27.2 | 7.2 | 6.4 | 2156 | 164.8 | 31.2 |
| Mean | 18.7 | 6.5 | 2.3 | 1640 | 127 | 20.8 |
| Std | 5.7 | 0.1 | 0.8 | 164 | 16 | 3.8 |

**Table 2. 4 Statistical properties of SST composition**

| | Tem | PH | BOD | COD | SS | TN | TP |
|---|---|---|---|---|---|---|---|
| | ℃ | | mg/l | mg/l | mg/l | mg/l | mg/l |
| Min | 6.8 | 6 | 1.1 | 3.5 | 0.4 | 7.3 | 0.9 |
| Max | 27 | 7.4 | 10.8 | 14.8 | 8.7 | 20 | 1.8 |
| Mean | 18.4 | 6.7 | 5.9 | 8.9 | 2.5 | 15.2 | 1.5 |
| Std | 5.7 | 0.2 | 2.6 | 2.1 | 1.1 | 2.6 | 0.2 |

Where,

Tem: temperature;

Min: minimum value;

Max: maximum value;

Std: standard deviation;

Secondly, various statistical manipulations can be performed in order to decipher trends in the data series. These are known as smoothing techniques and are designed to reduce or eliminate short-term volatility in the data. A smoothed series is preferred to a non-smoothed one because it can capture changes in the direction of the time series better than the unadjusted series; in addition, data smoothing eliminates the undesirable effect of possible noise in the process data. The conventional moving average technique was calculated for certain time series by consolidating the available data points into longer units of time; namely an average of several historical data points. The formula of 5-day Moving Average Technology is

$$X_i = (X_{i-2} + X_{i-1} + X_i + X_{i+1} + X_{i+2})/5$$

Where,

$X_i$: measured variable;

$i$: current time period;

5: number of time periods in the average;

In most cases, researchers use three-, four- or five-point moving averages. Here the five degree moving average, which is proved to be an appropriate value, was used to generate smoothed data from the raw data.

Next, it is the data scaling in the data preparation procedure, namely zero-mean normalizing. This is a standard procedure for the neural networks data preparation. The main objective here is to ensure that the statistical distribution of the values for each net input and output is roughly uniform. In addition, the values should be scaled to match the range of the input neurons. The data sets usually scaled so that they always fall within a specified range or they are normalized so that they have zero mean and unity standard deviation, using the following formula:

$$X_i = (X_i - X_{i,\min})/(X_{i,\max} - X_{i,\min})$$

Where,

$X_i$: measured value;

$X_{i,max}$: maximum value;

$X_{i,min}$: minimum value;

### 2.4.3 Data division

The data distribution between learning and validation phases is significant. Because it is not sufficient to evaluate a model by testing the prediction capability for the data used for model construction. Verification of prediction capability for the data that is not used for setting up the model has to be followed. The training set is the largest and is used to train the model. The validation set is used as an independent validation of the model following training. Without the validation set, the model would simply memorize the interactions present in each of the training patterns and would not be able to provide accurate prediction on data from outside the training set. If the majority of the available data is used during the learning phase, the quantity of data usable to test its effective will be tiny and consequently non-representative of the whole distribution. Then, the performance will not characterize the learning smoothness. On the other hand, if the majority of data is used during the validation phase, the learning will be certainly very rough and the network generalization capability limited.

Most ANN software packages periodically process the validation set through the model during the training to ensure that memorization does not occur. The trained model is applied to the validation data set patterns, to which the model has not been exposed, and an assessment of the accuracy of prediction is made. In our study, the data sets were divided into two subsets in the ANN and the hybrid model. That is, 200 data sets and 164 data sets were used for model learning and model evaluation, respectively.

# III. Results and Discussions

## 3.1 Prediction of Primary Settlement Tank Effluent

### 3.1.1 Prediction by an artificial neural network model

The modeling and simulation of physical and chemical process have been developed using ever more complex deterministic models, due to the recent evolution of personal computer. However, some circumstance makes application of these models impossible. For example, this is the case when some of the data to be used in the model is too difficult to obtain or when the model is very complex and requires a lot of simplification. Some studies that use neural networks to solve these kinds of problems have been published. Neural networks, a statistical tool for data analysis, could be applied to establish a relation between variables describing a process state and different measured quantities. This depends, in a not always obvious way, on the predictive variables used. The principal characteristic of neural networks is their capability to automatically establish relations between variables by means of a mechanism called training or learning. Neural networks are designed for a specific application and, after a training phase, able to generate a prediction, applying the relationship developed during the training period. Artificial neural network models make it impossible to develop non-linear empirical correlations. It is thus possible to connect a set of input variables, $P_i$ ($1 \leq i \leq I$), with a set of output variables, $Y_k$ ($1 \leq k \leq K$), assuming that we have $N$ relevant experimental values for the couples $[P_i, Y_k]_n$ ($1 \leq n \leq N$).

A neural network used here is FBNN that is a feed-forward back-propagation neural network with LM algorithm. Network architecture determines the number of connection weights (free parameters) and the way information flows through the network. Determination of appropriate network architecture is one of the most important but also one of the most difficult tasks in the model building process. Traditionally, optimal network architecture, including the optimum number of hidden layers, the optimum number nodes in each layer, the optimum number of epochs, and the optimum value of the internal parameters and so on, has been found by trial and error. In view of the multiple factors and parameters affecting complete WWTP process, a trial-and-error procedure is commonly used until each model architecture design has been found with minimal error.

As in all empirical models, the users must bear in mind that the regression models quality $Y=f\ (P)$, will depend on the relevance and the quality of the available experimental data used during the training phase. Moreover, getting a good prediction over a some given points with a regression model, does not guarantee an important generalization capability or a good prediction for a new number of couples $[P_i,\ Y_k]_m$ ($N+1 \leq m \leq M$). Keeping this in mind, a neural network must establish general mechanisms and be able to adapt them continuously to new and unknown situations by means of recalibration procedures. Therefore, it is not sufficient to evaluate a model by testing the prediction capability for the data used for model construction. Firstly, all 200 data sets are used for modeling and then the prediction capability of the model is evaluated by root mean square error (RMSE), R-square ($R^2$) and average relative difference (ARD). As a second step, the remaining 164 data sets are used for verifying the prediction capability of the model, in the same way evaluating by RMSE, $R^2$ and ARD. Fig. 3.1.1 represents the whole used ANN procedure.

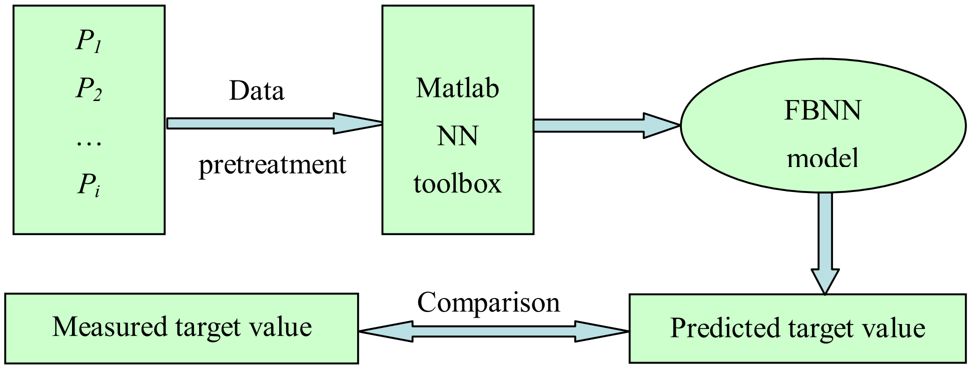


**Fig. 3.1. 1 General scheme of the prediction ANN stages**

To reach the suitable network architecture, several trials for each parameter have been conducted until the suitable number of hidden layer, number of neurons and epochs were determined, as shown in Table 3.1.1. The suitable architecture is the one which produced the minimal error term in both training and validating data.

**Table 3.1. 1 Model architectures and prediction results by ANN**

| Target | Architecture | | $R^2$ | | RMSE | | ARD | |
|---|---|---|---|---|---|---|---|---|
| | I-H1-H2-O-E | MSE | train | validation | train | validation | train | validation |
| BOD | 8-5-0-1-300 | 0.002 | 0.96 | 0.86 | 2.12 | 6.74 | 0.03 | 0.114 |
| COD | 8-5-0-1-100 | 0.001 | 0.99 | 0.92 | 1.17 | 2.94 | 0.03 | 0.086 |
| SS | 8-5-0-1-300 | 0.002 | 0.98 | 0.78 | 2.38 | 8.86 | 0.03 | 0.132 |
| TN | 8-3-3-1-100 | 0.001 | 0.99 | 0.90 | 0.91 | 3.01 | 0.02 | 0.092 |
| TP | 8-5-0-1-300 | 0.002 | 0.96 | 0.89 | 0.16 | 0.36 | 0.04 | 0.091 |

Where,

I-H1-H2-O-E: Input layer--Hidden layer1--Hidden layer2--Output layer--Epoch;

MSE: Mean Square Error (to show the model performance);

$R^2$: Correlation Coefficient (to show the correlativity between the normalized targets and the model outputs);

RMSE: Root Mean Square Error (to show the difference between the measured data and the model predicted data);

$$RMSE = \sqrt{\sum_{t=1}^{N} \left(Y_M - Y_P\right)^2 / N}$$

ARD: Average Relative Difference;

$$ARD = \frac{1}{N} \sum_{n=1}^{N} \frac{Y_M - Y_P}{Y_M}$$

Where,

$Y_M$: measured value;

$Y_P$: predicted value;

$N$: data point;

As the Table 3.1.1 said, we applied the single and multi-layer feed-forward back-propagation neural network model for the prediction of BOD, COD, SS, TP and TN respectively. The situation is quite different if non-linear hidden neurons are inserted between the input and output layers. In this case, it seems natural to assume that the more layers are used, the greater power the networks processes. However, it is not the case in practice. According to our trials, an excessive number of hidden layers often proves to be unproductive. It causes slower convergence in the back-propagation learning because intermediate neurons not directly connected to output neurons have very small weight changes and learn very slowly. The error signals are numerically degraded when propagating across too many layers; extra layers tend to create additional local minima. Thus, it is essential to identify the proper number of layers. According to our tests, the single and two-hidden-layer were used for the prediction of BOD, COD, SS, TP and TN respectively in the present study. And the number of neurons placed in the network hidden layer and the number of iterations completed during the training phase were also considered during every trial, due to the influence on the learning phase. The number of neurons placed in the network hidden layer and the number of iterations completed during the training phase are very important user-definable parameters. Although the relationship between the network performance and its hidden layer size is not well understood, a principle can be used as a guide: the principle of generalization versus convergence. Generalization means the network ability to produce good results with a data set that has not been used during the network learning phase. Convergence is the ability to learn the training data. As mentioned before, the trial-and-error method was used to find the proper network model architecture for each predicted target in our study.

Fig. 3.1.2~Fig. 3.1.6 represent the training and validation results of the proposed neural network for these water quality parameters, respectively. Thereinto, the training and validation prediction results are separated by the line.

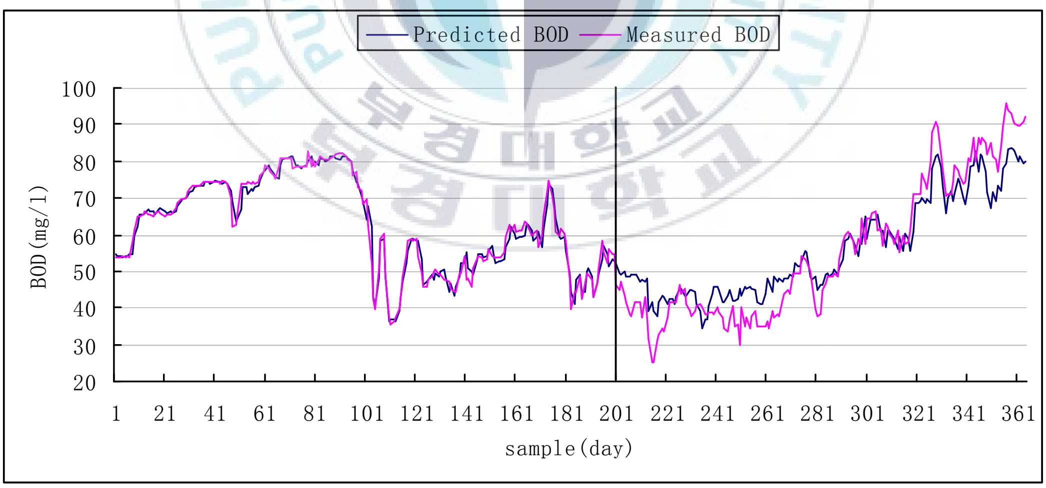


**Fig. 3.1. 2 Prediction of BOD with ANN in the primary settlement tank**

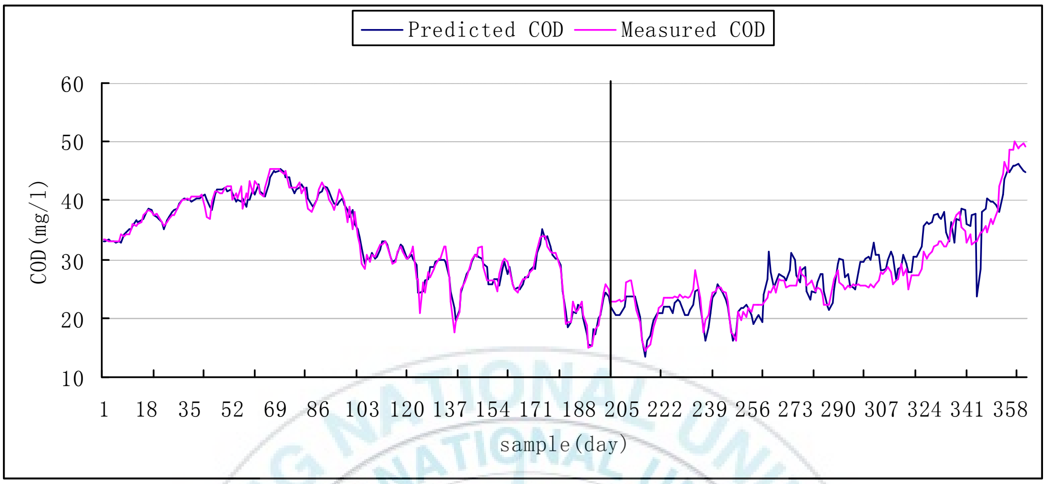


**Fig. 3.1.3 Prediction of COD with ANN in the primary settlement tank**

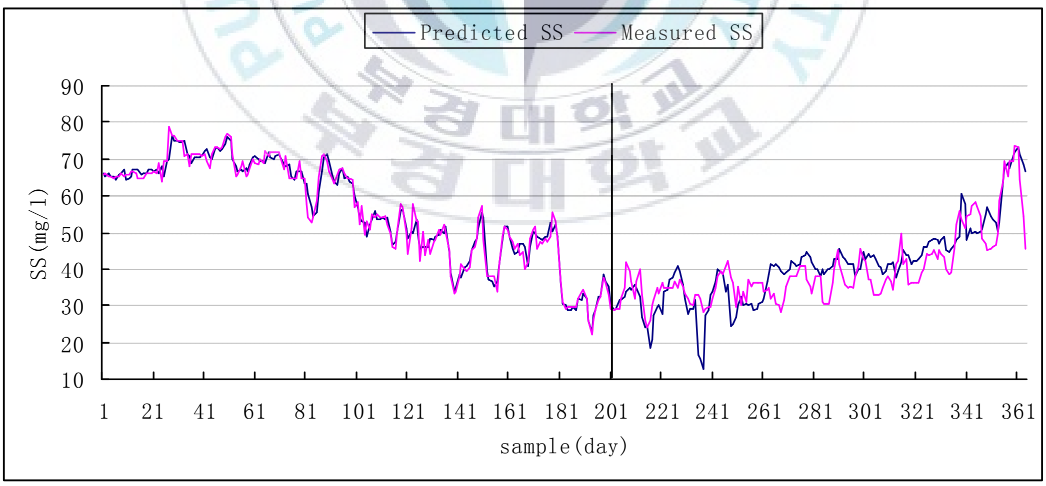


**Fig. 3.1.4 Prediction of SS with ANN in the primary settlement tank**

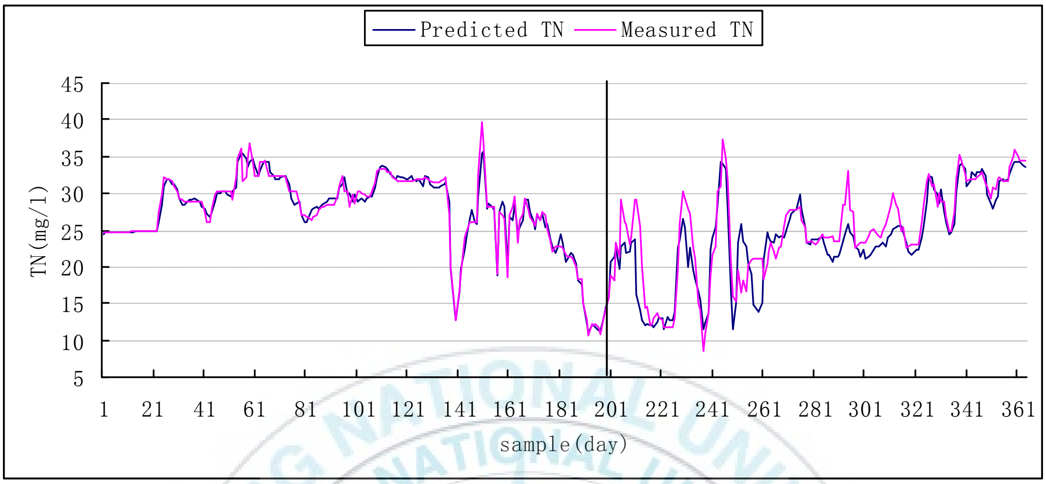


**Fig. 3.1.5 Prediction of TN with ANN in the primary settlement tank**

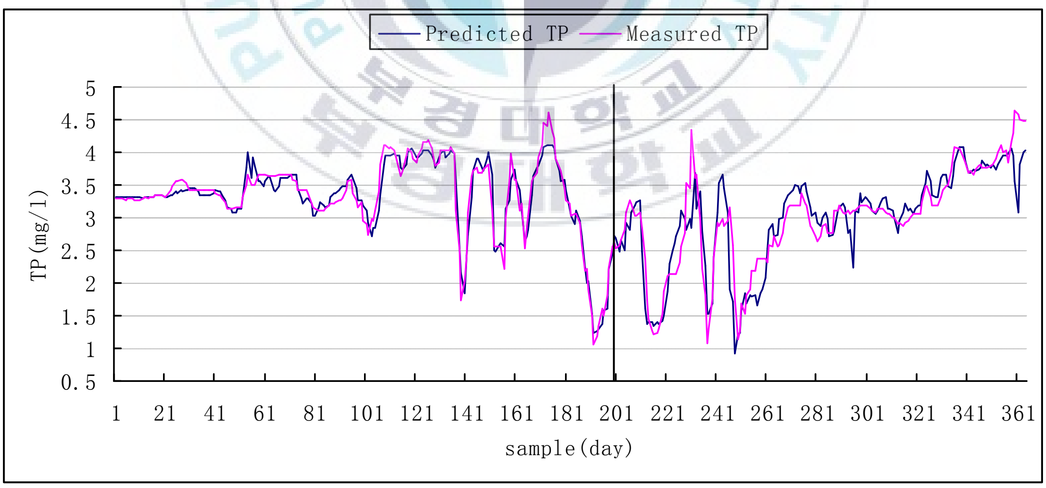


**Fig. 3.1.6 Prediction of TP with ANN in the primary settlement tank**

As shown in Table 3.1.1, during the training phase, the RMSE ranged from 0.16 to 2.38, and the R-square ranged from 96% to 99%. The narrow bound of error measures throughout all training groups and for the five modeled parameters is an indication of the ANN's robustness in model training. However, the RMSE and the R-square for the testing phase ranged from 0.36 to 8.86 and 78% to 92%, respectively. The model prediction for validation is not satisfying, especially SS and BOD with 8.86, 78% and 6.74 and 86%, respectively. And also the average relative difference of this two modeled parameters, SS and BOD, has exceeded 10%, but not exceed 15%, which is acceptable within the allowable range in the practical, while others are all lower than 10%, as shown in Table 3.1.1. It is considered that for BOD and SS, this not good validation prediction capability of ANN is probably due to its over-fitting problem. The over-fitting problem that can occur during the learning phase has an extremely negative effect on the network generalization capability, which becomes unable to predict good behaviors. Over-fitting usually arises due to too much training for noise data and outliers. Pretreatment cannot remove the noise within the data completely. This enhancement of prediction capability can be accepted as a result of data preprocessing with principal component analysis (PCA). It is thought that PCA can minimize the influence of noise and outlier, by capturing the main trend of the data set. The prediction procedure in the next section thus utilizes a PCA in order to optimize the neural network training.

### 3.1.2 Prediction by a hybrid artificial neural network model

As mentioned above, the data treatment by principal component analysis (PCA) is carried out to improve the necessary effectiveness of the neural network learning phase. The hybrid artificial neural network just means an artificial neural network combining principal component analysis. As part of the prediction procedure, PCA is used in order to reduce the number of network inputs. The set of variables is used for its predictive character and the PCA enables to reduce its dimension in order to support the network learning process without notable information loss. The technique was applied to each of data group, used to predict the five effluent parameters in primary settlement tank. The data file is composed of influent (k), primary settlement tank influent (k), primary settlement tank effluent (k-1). These quantitative variables have a high degree of correlation and thus a redundant character. The PCA allows us to extract from this set of correlated variables a non-redundant list of new synthetic variables. These synthetic variables are linear combinations of the initial quantitative variables. They make it possible to minimize the risks of over-fitting which often and easily happened in ANN model. The prediction result by ANN model in the preceding section also exemplified it.

In the same way, all 200 data sets are used for model constructing and the remaining 164 data sets are used for model validation, and also the performance function used for training and validation is evaluated in terms of error measurement root mean square error (RMSE), R-square and average relative difference (ARD). Fig.3.1.7 represents the whole used ANN&PCA procedure.

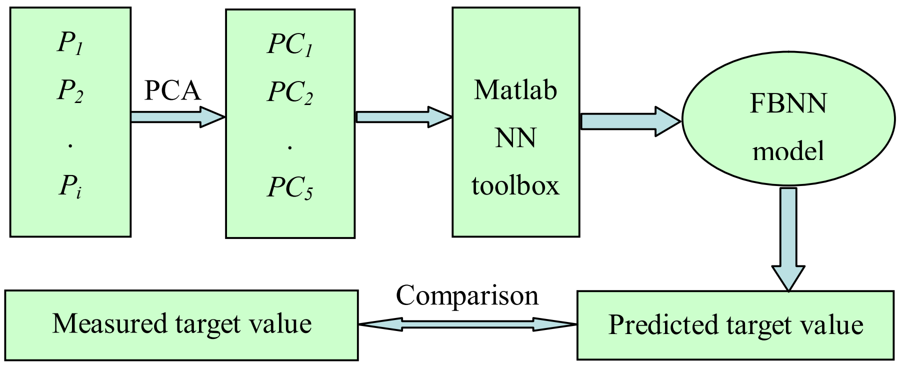


**Fig. 3.1. 7 General scheme of the prediction ANN&PCA stages**

The 8 wastewater quality parameters, the same inputs to the 3.1.1 section, are reduced to five principal components (PCs), which become inputs of ANN. According to the obtained results, shown in Table 3.1.2 and Table 3.1.3, the number of preserved synthetic variables can vary but must always be representative of the information expressed by the set of initial variables. For Table 3.1.2 used for prediction of BOD, COD and SS, the first five principal components which express 96.19% of the system total variance are preserved. That means 96.19% of the variation within the data can be explained by this five PCs. It is important to emphasize that more than 60% of this total variance is expressed by the first principal component alone. The three principal components not selected to form part of the new synthetic data file express 2.02%, 1.02% and 0.77% of the original variance, respectively. Their contribution is thus very low. The five new definite synthetic variables allow us to reduce considerably the data file dimension used for its predictive character, while preserving the information large majority. Table 3.1.3 presents the variance expressed by each of the 8 principal components using data sets II for TN, TP. We also preserve for data sets II the first five principal components which express 95.11% of the system total variance. The first principal component express almost 70.25% of the total variance, whereas not selected components express only 2.79%, 1.56% and 0.94% of this variance. As for data sets I, the five synthetic variables obtained by PCA make it possible to significantly reduce the data file dimension.

**Table 3.1. 2 Eigenvalues of the correlation matrix I for BOD, COD, SS**

| PC | Eigenvalue | Contribution (%) | Cumulative (%) | Used |
|---|---|---|---|---|
| $PC_1$ | 95.80 | 75.27 | 75.27 | Yes |
| $PC_2$ | 9.51 | 7.47 | 82.74 | Yes |
| $PC_3$ | 8.19 | 6.45 | 89.18 | Yes |
| $PC_4$ | 5.91 | 4.64 | 93.83 | Yes |
| $PC_5$ | 3.00 | 2.36 | 96.19 | Yes |
| $PC_6$ | 2.57 | 2.02 | 98.20 | No |
| $PC_7$ | 1.30 | 1.02 | 99.23 | No |
| $PC_8$ | 0.98 | 0.77 | 100 | No |

**Table 3.1. 3 Eigenvalues of the correlation matrix II for TN, TP**

| PC | Eigenvalue | Contribution (%) | Cumulative (%) | Used |
|---|---|---|---|---|
| $PC_1$ | 75.91 | 70.25 | 70.25 | Yes |
| $PC_2$ | 11.16 | 10.32 | 80.58 | Yes |
| $PC_3$ | 6.93 | 6.41 | 86.99 | Yes |
| $PC_4$ | 5.14 | 4.76 | 91.75 | Yes |
| $PC_5$ | 3.63 | 3.36 | 95.11 | Yes |
| $PC_6$ | 2.79 | 2.58 | 97.69 | No |
| $PC_7$ | 1.56 | 1.44 | 99.13 | No |
| $PC_8$ | 0.94 | 0.87 | 100 | No |

Fig. 3.1.8~Fig. 3.1.12 represent the training and validation results of the ANN&PCA hybrid neural network for these wastewater quality parameters, respectively. Thereinto, the training and validation phases are separated by line.

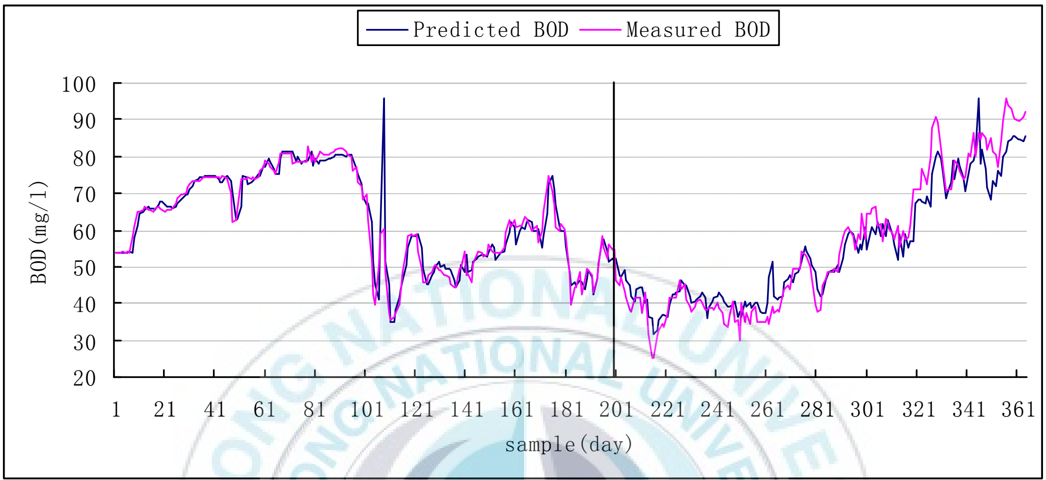


**Fig. 3.1. 8 Prediction of BOD with ANN&PCA in the primary settlement tank**

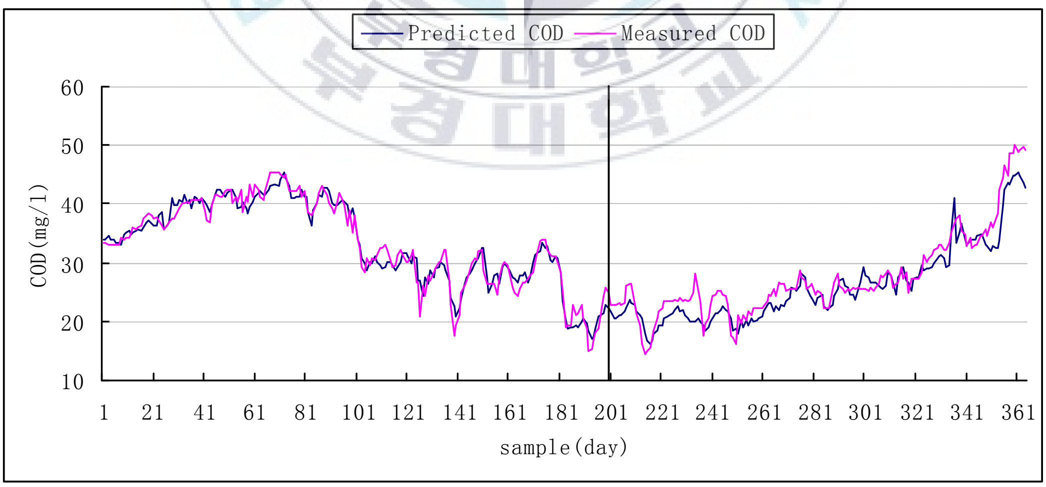


**Fig. 3.1. 9 Prediction of COD with ANN&PCA in the primary settlement tank**

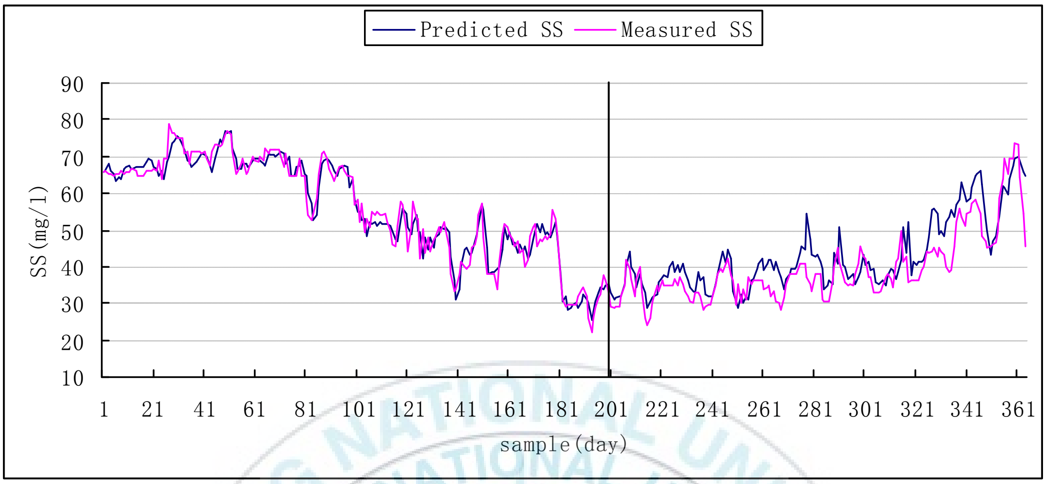


**Fig. 3.1.10 Prediction of SS with ANN&PCA in the primary settlement tank**

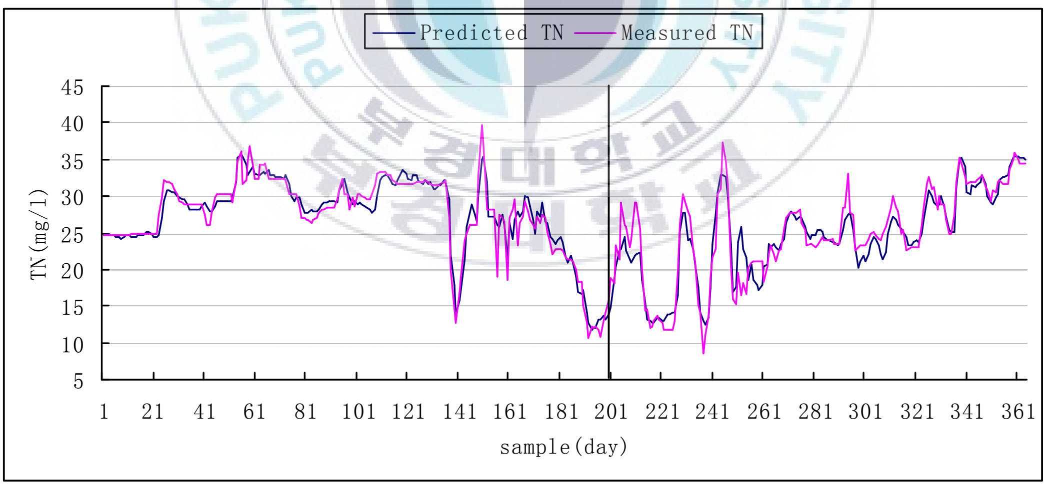


**Fig. 3.1.11 Prediction of TN with ANN&PCA in the primary settlement tank**

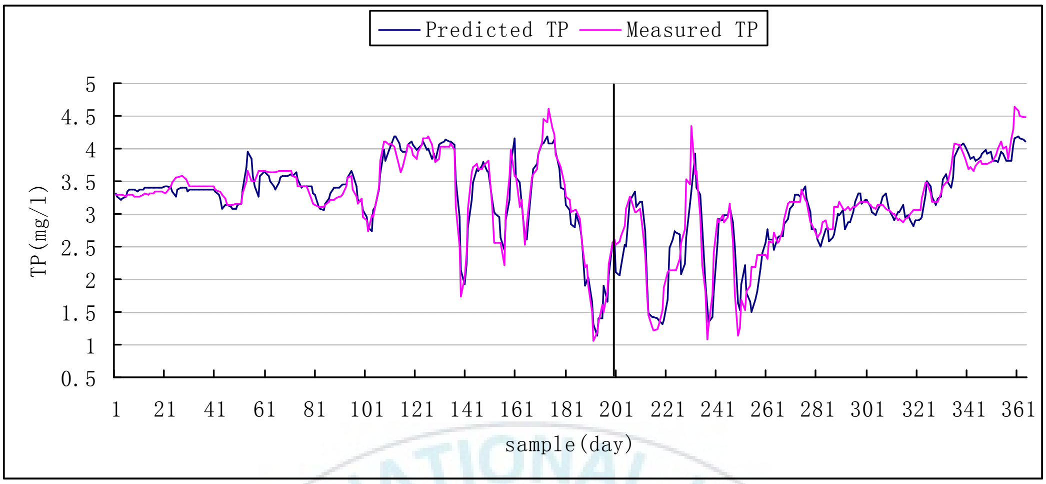


**Fig. 3.1. 12 Prediction of TP with ANN&PCA in the primary settlement tank**

**Table 3.1. 4 Model architectures and prediction results by ANN&PCA**

| Target | Architecture | | $R^2$ | | RMSE | | ARD | |
|---|---|---|---|---|---|---|---|---|
| | I-H1-H2-O-E | MSE | train | validation | train | validation | train | validation |
| BOD | 8-5-0-1-300 | 0.003 | 0.96 | 0.97 | 3.66 | 5.33 | 0.04 | 0.078 |
| COD | 8-5-0-1-100 | 0.002 | 0.97 | 0.95 | 1.67 | 2.62 | 0.04 | 0.077 |
| SS | 8-5-0-1-300 | 0.003 | 0.98 | 0.88 | 2.96 | 5.86 | 0.05 | 0.119 |
| TN | 8-3-3-1-100 | 0.002 | 0.96 | 0.94 | 1.48 | 2.18 | 0.04 | 0.070 |
| TP | 8-5-0-1-300 | 0.002 | 0.96 | 0.93 | 0.17 | 0.28 | 0.04 | 0.083 |

As shown in Table 3.1.4, during the training phase, the RMSE ranged from 0.17 to 3.66, and the R-square ranged from 96% to 98%, similar to the training phase by ANN. The RMSE and the R-square for the testing phase ranged from 0.28 to 5.86 and 88% to 97%, respectively. The model prediction of training phase by ANN&PCA is similar to the one by ANN. However, the model prediction of validation phase by ANN&PCA is much better than the one by ANN, especially BOD and SS. With the same model architecture to ANN, the hybrid ANN model that integrates the PCA reduced RMSE to 5.33 from 6.74 for BOD and from 8.86 to 5.86 for SS. A point must be noted that all the average relative difference is lower than 9%, expect for SS with the value of 11.9%, a little more than 10%. The results obtained from this ANN&PCA hybrid model indicate that it is capable of producing accurate predictions over the range of the data used for model calibration.

## 3.2 Prediction of Secondary Settlement Tank Effluent

### 3.2.1 Prediction by a hybrid artificial neural network model

The prediction tests carried out showed that a too large number of variables used as network inputs, combined with a too restricted number of data sets has a negative impact on the network training process and favors over-fitting. The reduction of the data file dimension used for the training phase could also be obtained by eliminating several variables but this would be to the detriment of the information they express.

As mentioned before, the prediction results indicate that the hybrid ANN is valuable among the various techniques. In this section, the coupling of the hybrid neural network model to the activated sludge systems that are the most extensively used in wastewater treatment plants was used. In an activated sludge process, the wastewater, which contains organic matter, suspended solids and nutrients, enters an aerated tank where it is mixed with biological floc particles. After a sufficient contact time, this mixture is discharged into a settler that separates the suspended biomass from the treated water. Most of the biomass is recirculated to the aeration tank, while a little amount is purged daily (see Fig. 2.1).

Biological wastewater treatment processes are commonly used to treat municipal and industrial wastewaters, in despite that it is really difficult to understand, and thus difficult to be correctly operated and controlled. Special attention has been paid to biological processes modeling, both for wastewater treatment and sludge stabilization processes. Operational difficulties in wastewater treatment plants are often encountered. Increased regulation and operational difficulties of wastewater treatment plant has resulted in increased need for tools to evaluate the organic matter and nutrient-removal capabilities of wastewater treatment processes. As a consequence, a task group pointed by the International Water Association developed a mathematical model of the activated sludge process, Activated Sludge Model No.1, considered the processes of organic substrate removal, nitrification and denitrification. The task force further added phosphorus removal to the model and published the result as Activated Sludge Model No.2. Activated Sludge Model No.1 was later modified to develop Activated Sludge Model No.3 in 1999. Reliable performance evaluation treatment plants can be done by simulating the plant behavior over a wide range of influent disturbances, including series of rain events with different intensity and duration, seasonal temperature variables, holiday effects, etc. Such simulation-based WWTP performance evaluations are in practice limited by long simulation time of the mechanistic WWTP models.

Activated sludge process is difficult to operate and control because of its complex operational behavior and usual significant process disturbances. And little work has been reported on addressing this operational problem because of its difficulty for acquiring operational knowledge that is well described and formulated for an activated sludge process. It is known that most of the problems of poor activated sludge effluent quality result from the inability of the secondary settler to efficiently remove the suspended biomass from the treated water. Owing to the complex interaction caused by the recycle of sludge from the secondary settlement tank to the aerator, the operational behavior of activated sludge process is usually very complicated. To increase safety and improve operating performance of this biological wastewater treatment process, it is important to develop computer operational decision support systems. Operation, control and supervision of WWTP have been approached from many different points of view, including classical control methods, mechanistic models, knowledge-based systems, case-based reasoning, neural nets and hybrid approaches. However, a direct cause-effect relationship for WWTP performance has been established only in s few cases and even in those, experimental results could lead to contradictory conclusions, avoiding the formulation of deterministic cause-effect relationship that could be used as prediction models. The identification of a model to predict in real-time with reasonable accuracy for the effluent quality parameters is therefore of great practical important. The intelligent computing system is able to assist ordinary operators to work at the level of domain expert in daily operation. The method used here is the artificial neural network, which is powerful because it can learn to represent complicated data patterns or data relationship between input and output variables of the system being studied. Nevertheless, it has limitations in performing heuristic reasoning of the domain problem. That means ANN are able to learn complex nonlinear between inputs and outputs of the activated sludge process to capture the knowledge, but are not able to help improve the heuristic understanding of the operational problems. In order to strength the extendibility of ANN model prediction, here some data values, getting from the ASM model were used to train the artificial neural network, together with the raw data from the wastewater treatment plant. As an alternative to activated sludge models, a hybrid neural network was used to model this strongly non-linear nature of the biological process. The flow of coupling of the artificial neural network to the activated sludge model used in this section is shown in Fig. 3.2.1.

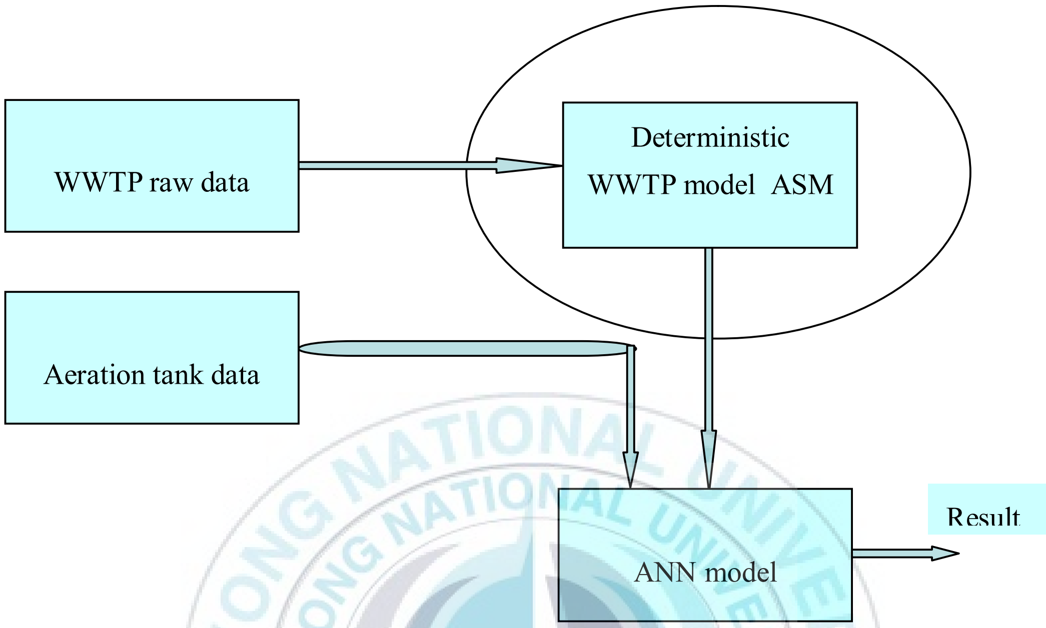


**Fig. 3.2. 1 Scheme of the proposed hybrid ANN model procedure**

Here some dissolved and particulate components, together with some raw material data, are used to characterize the influent wastewater or sludge and the active biomass. Table 3.2.1 gives us some detail information of the input variables for each wastewater quality parameters' prediction.

**Table 3.2. 1 Inputs and output targets for the ANN**

| Targets | Inputs |
|---|---|
| BOD(k) | T(k), BOD(k-1), COD(k-1), SS(k-1), S_S, X_S, X_H, X_A |
| COD(k) | T(k), BOD(k-1), COD(k-1), SS(k-1), S_S, X_S, X_H, X_A |
| SS(k) | MLSS(k), BOD(k-1), COD(k-1), SS(k-1), S_S, X_S, X_H, X_A, X_I |
| TN(k) | COD(k-1), TN(k-1), S_NO, X_ND, X_H, X_A, X_P |
| TP(k) | MLSS(k), TN(k-1), TP(k-1), S_NO, X_P, S_ND |

Where,

S_S: ready biological substrate;

X_S: slowly biological substrate;

X_H: heterotrophic organisms;

X_A: nitrifying organisms;

X_I: inert particular organic material;

S_NO: nitrate and nitrite nitrogen;

S_ND: soluble biodegradable organic nitrogen;

X_ND: particular biodegradable organic nitrogen;

X_P: particulate products arising from biomass decay;

k: present;

k-1: one day before;

Input-target training data are usually pretreated as explained in the above section in order to improve the numerical condition for the optimization problem and for better behavior of the training process. And also the data sets are normally divided into two subsets; training and validation subsets. The training subsets data are used to accomplish the network learning and fit the network weights by minimizing an appropriate error function. Back-propagation is the training technique usually used for this purpose. It refers to the method for computing the gradient of the case-wise error function with respect to the weights for a feed-forward network. The performance of the networks is then compared by evaluating the error function. The validation subset data are then used to measure the generalization of the network (i.e. how accurately the network predicts targets for inputs that are not in the training subset data).

Selecting network structure is a critical step in the overall design of ANN, although there is no widely accepted best method of developing ANN models.

When all the possible options in building the ANN model architecture are considered, an almost infinite number of distinct architectures are possible. As such, each model developer may use a different protocol to reduce the number of architectures that are evaluated. The structure must be optimized to reduce computer processing, achieve good performance and avoid over-fitting. Table 3.2.2 shows the ANN architectures of all the five quality parameters and the training and validation prediction results.

**Table 3.2. 2 ANN architecture and prediction result**

| Target | Architecture | | $R^2$ | | RMSE | | ARD | |
|---|---|---|---|---|---|---|---|---|
| | I-H1-H2-O-E | MSE | train | validation | train | validation | train | validation |
| BOD | 8-3-3-1-300 | 0.001 | 0.99 | 0.987 | 0.25 | 0.68 | 0.03 | 0.112 |
| COD | 8-3-3-1-300 | 0.001 | 0.99 | 0.982 | 0.23 | 0.42 | 0.02 | 0.046 |
| SS | 9-3-3-1-100 | 0.001 | 0.99 | 0.95 | 0.27 | 0.66 | 0.08 | 0.186 |
| TN | 7-3-3-1-300 | 0.001 | 0.99 | 0.927 | 0.46 | 0.88 | 0.02 | 0.054 |
| TP | 6-3-3-1-300 | 0.002 | 0.97 | 0.94 | 0.04 | 0.06 | 0.02 | 0.033 |

The selection of the best number of hidden units depends on many factors. The size of the training set, amount of noise in the targets, complexity of the sought function to be modeled, type of activation functions used and the training algorithm all have interacting effects on the sizes of the hidden layers. There is no way to determine the best number of hidden units without training several networks and estimating the generalization error of each. Here a common method, trial and error method, was used to select the best ANN architecture for each wastewater quality parameter. As Table 3.2.2 shown, ANN architectures were selected according to the principal of minimum prediction error.

Fig. 3.2.2~Fig. 3.2.6 gives us the prediction result for the five wastewater quality parameters, respectively.

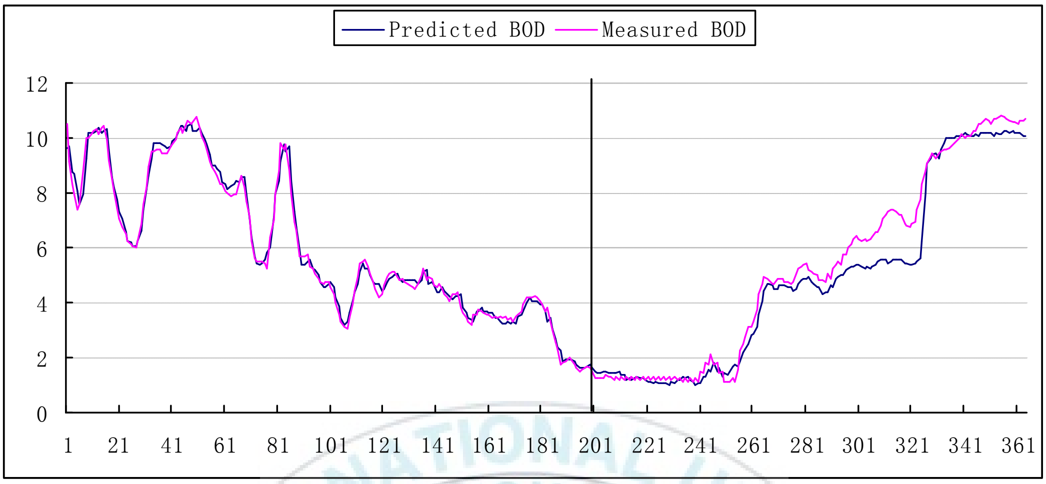


**Fig. 3.2. 2 Prediction of BOD with hybrid ANN in the secondary settlement tank**

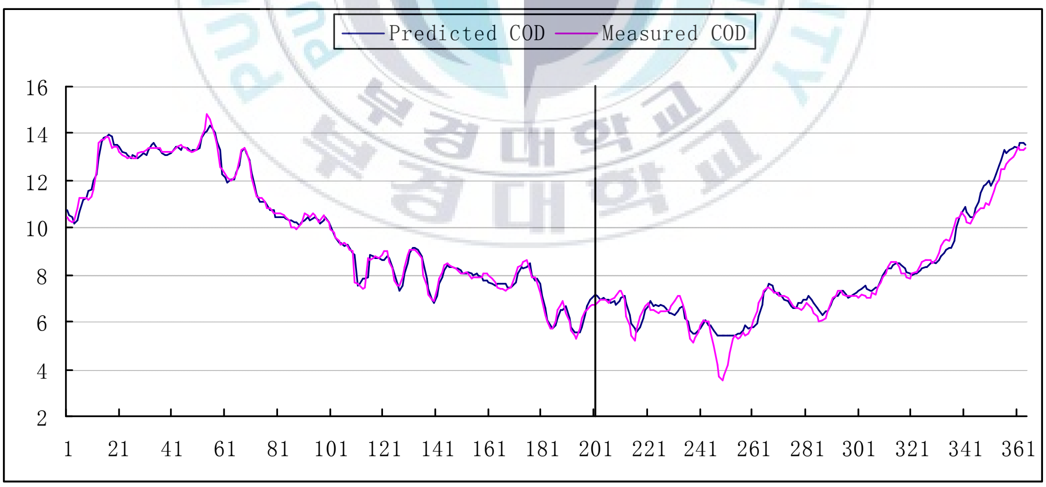


**Fig. 3.2. 3 Prediction of COD with hybrid ANN in the secondary settlement tank**

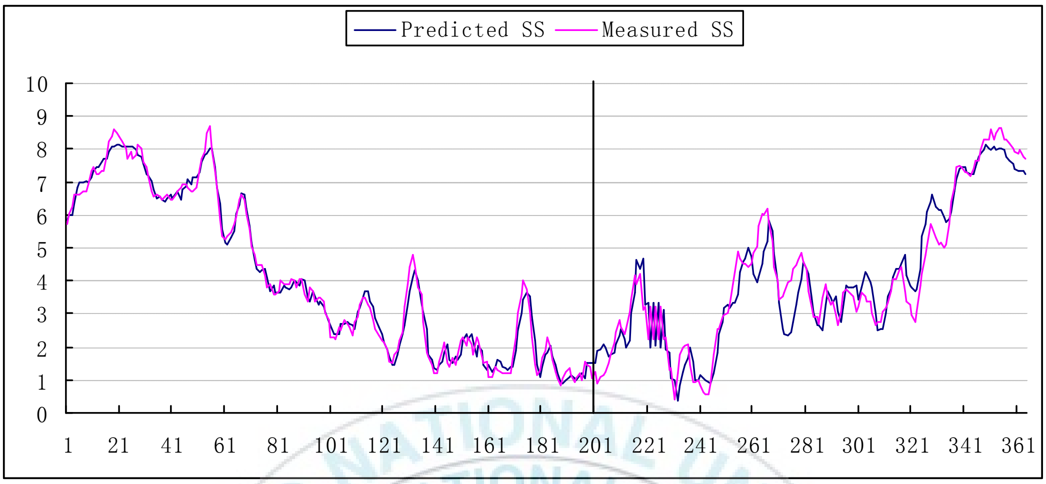


Fig. 3.2. 4 Prediction of SS with hybrid ANN in the secondary settlement tank

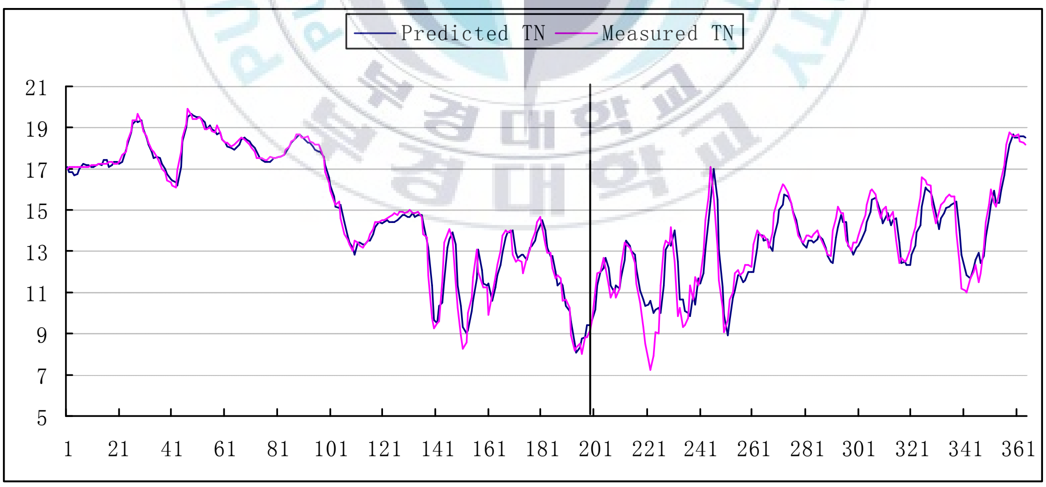


Fig. 3.2. 5 Prediction of TN with hybrid ANN in the secondary settlement tank

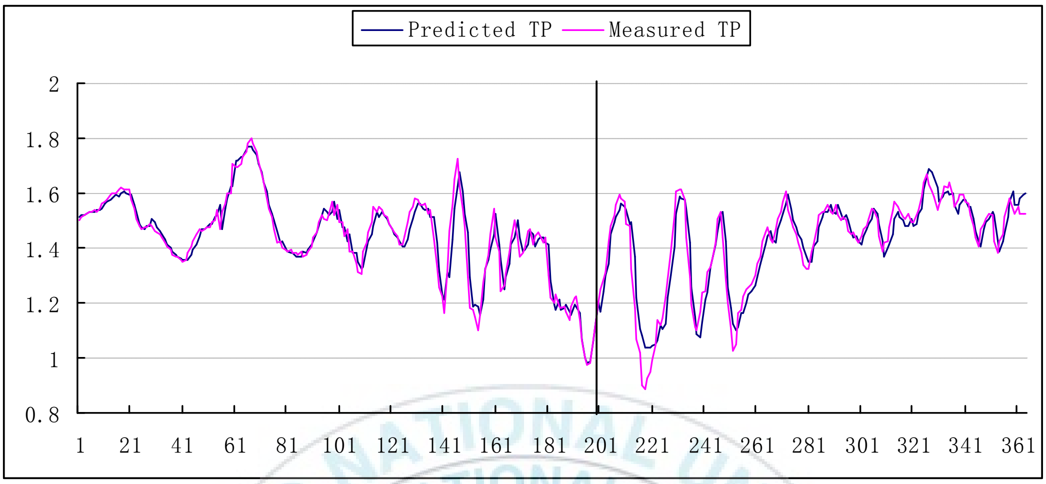


**Fig. 3.2. 6 Prediction of TP with hybrid ANN in the secondary settlement tank**

Fig 3.2.2 to Fig. 3.2.6 shows the training and validation prediction performance for the five modeled targets. From these figures, we could know that the modeled values were followed well in the direction of the measured values. As shown in Table 3.2.2, all of the R-squares, including training and validation phases, are above 93%. And the root mean square error values are also lower. Especially for COD, TN and TP, the average relative difference values are lower than 5%, although the value for BOD is a little more than 10%. But in evidence, the average relative difference value for SS is much more than 10% indubitably, even exceeding our allowable practical limitation 15%. This is probably because of the noise problem within the raw data from the WWTP.

Looking over the result Table 3.1.1, Table 3.1.4 and Table 3.2.2, it is possible to know that effluent parameters COD, TN and TP, the average relative error made during learning and validation phases is very weak, lower than 10% for the primary settlement tank and for the secondary settlement tank it is even lower than 5%. These parameters have weak variations which make easier their estimations. And also for the BOD in our study, it is a little more than 10% including the primary and secondary settlement tank. However, without a doubt, the error levels show that SS prediction is more difficult than others. Investing this reason, this is probably not because of possible shortcomings in the modeling technique, but because of possible shortcoming in the data used for model development, or the erroneous data used.

# IV. Conclusions

Modeling a WWTP is difficult to accomplish due to the high nonlinearity of the plant and the non-uniformity and variability of the crude supply as well as the nature of the strongly complex biological treatment. An ANN modeling approach was implemented to solve this problem and to discover the interdependence of input-output variables. The plant input-output data were used to predict the plant behavior. The modeling approaches used in this study, namely artificial neural network model and hybrid artificial neural network model, gave comparable predictions of the wastewater treatment plant performance.

In the first section, a prediction procedure was presented based on a 12-month-data set, to obtain BOD, COD, SS, TN, TP estimation for WWTP primary settlement tank effluent. Some parameters are not directly measurable and can not be evaluated by laboratory analysis which can be relatively long. Their estimation by neural network and hybrid neural network was carried out using past real data obtained from the South Wastewater Treatment Plant, Busan. Although the first neural network approach gave not very good predictions because of the over-fitting problem, the error calculated is not exceeded 15%, the limited allowable practical range. During training of ANN, some degradation of prediction capability caused by over-fitting is frequently observed. Over-fitting usually arises due to too much training for the noise data. Pretreatment cannot remove the noise within the data completely. The following hybrid neural network approach that preprocesses inputs through PCA to ANN shows the most accurate prediction capability, as shown in Table 3.1.1 and Table 3.1.4, and root mean square error and average relative difference of the hybrid artificial neural network model are all lower than that of artificial neural network during the training and validation phase, in the prediction of the primary settlement tank effluent. This indicates that the hybrid neural network is more capable of producing accurate predictions over the range of the data used for model calibration. It is confirmed that in modeling of the correlated noisy data, the preprocessing using PCA that enables to reduce the dimensionality of input variables is effective.

In the complex biological treatment section, a hybrid neural network, combined with activated sludge models that are the most extensively used in wastewater biological treatment processes, was used to predict the secondary settlement tank effluent. From the result table, most of the five parameters are modeled well, except for SS with the average relative difference error exceeding out allowable practical limitation in this study. Determination of an appropriate model structure for biological treatment systems of industrial wastewater is a formidable task. It is not always clear whether the poor fit to the data owes to the structure of the model or to the estimation of model parameters or the input variables.

When it is difficult to model a system by a separate approach due to nonlinearity and noise within data, the hybrid method that integrates the merits of different approaches can be a better alternative. Section 3.1 and section 3.2 showed characteristics of hybrid artificial neural network and their applicability for the wastewater treatment system. The hybrid models provided a robust tool for the prediction in that the prediction error is acceptable in an allowable range. The limitation in data, however, should be highlighted. If more data were collected, and if the data were less noisy, this would have resulted in an improved predictive capability of the neural network. The main advantage of neural network is their ability to extract the underlying phenomena directly from historical plant data whereas other artificial intelligence systems such as expert systems need human intervention to encode knowledge about the process. However, if the database is not sufficient, neural model could lead to erroneous interpolations, or restricted to a narrow range of operating conditions, neural model could lead to erroneous extrapolations. In either of these conditions, the using of neural network approach will be limited. The hybrid neural network models, combined with other approaches, are used mainly to improve the prediction quality of neural network and also weaken the influence of over-fitting. Nevertheless, the neural network and hybrid neural network are tools that are worth considerations in the prediction of WWTP.

# References


[1] Maged M. Hamed, Mona G. Khalafallah, Ezzat A. Hassanien, 2004. Prediction of wastewater treatment plant performance using artificial neural networks. Environmental Modeling & software 19, 919-928.

[2] Dong-Jin Choi, Heekyung Park, 2001. A hybrid artificial neural network as software for optimal control of a wastewater treatment process. Wat. Res. Vol.35, No.16, 3959-3967.

[3] Holger R. Maier, Nicolas Morgan, Christopher W.K. Chow, 2004. Use of artificial neural networks for predicting optimal alum does and treated water quality parameters. Environmental Modeling & software 19, 485-494.

[4] Sung Hun Hong, Min Woo Lee, Dae Sung Lee, Jong Moon Park, 2007. Monitoring of sequencing batch reactor for nitrogen and phosphorus removal using neural networks. Biochemical Engineering Journal 35, 365-370.

[5] Stephane Grieu, Adama Traore, Monique Polit, Jesus Colprim, 2005. Prediction of parameters characterizing the state of a pollution removal biologic process. Engineering Applications of Artificial Intelligence 18, 559-573.

[6] Farouq S. Mjalli, S. Al-Asheh, H.E. Alfadala, 2007. Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. Journal of Engineering Management 83, 329-338.

[7] B. Raduly, K.V. Gernaey, A.G. Capodaglio, P.S. Mikkelsen, M. Henze, 2007. Artificial neural networks for rapid WWTP performance evaluation: Methodology and case study. Environmental Modeling & software 22, 1208-

1216.

[8] D. Aguado, A. Ferrer, A. Seco, J. Ferrer, 2006. Comparison of different predictive models for nutrient estimation in a sequencing batch reactor for wastewater treatment. Chemometrics and Intelligent Laboratory Systems 84, 75-81.

[9] Haejin Ha, Michael K. Stenstorm, 2003. Identification of land use with water quality data in storming using a neural network. Water Research 37, 4222-4230.

[10] Karla Patricia Oliveira-Esquerre, Dale E. Seborg, Roy E. Bruns, Milton Mori, 2004. Application of a steady-state and dynamic modeling for the prediction of the BOD of an aeration lagoon at a pulp and paper mill; Part I. linear approaches. Chemical Engineering Journal 104, 73-81.

[11] Glenda M. Pigram, 2001. Use of neural network models to predict industrial bioreactor effluent quality. Environ. Sci. Technol 35, 157-162.

[12] C.W. Baxter, S.J. Stanley, Q. Zhang, D.W. Smith, 2002. Developing artificial neural network models of water treatment processes: a guide for utilities. J. Environ. Eng. Sci. 1, 201-211.

[13] Holger R. Maier, Graeme C. Dandy, 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modeling issue and applications. Environmental Modeling & software 15, 101-124.

[14] Qing Zhang, Stephen J. Stanley, 1997. Forecasting raw-water quality parameters for the North Saskatchewan River by neural network modeling. Wat. Res. Vol. 31, No. 9, 2340-2350.

[15] Dae Sung Lee, Jong Mon Park, 1999. Neural network modeling for on-line estimation of nutrient dynamics in a sequentially-operated batch reactor. Journal of Biotechnology 75, 229-239.

[16] Junko Kambe, Tomoo Aoyama, Aiko Yamauchi, Umpei Nagashima, 2007.

Extraction of a parameter as an index to asses water quality of the Tamagawa, Tokyo, Japan, by using neural networks and multivariate analysis. J. Comput. Chem. Jpn., Vol.6, No.1, 19-26.

[17] Mohamed F. Hamoda, Ibrahim A. Al-Ghusain, Ahmed H. Hassan, 1999. Integrated wastewater treatment plant performance evaluation using artificial neural networks. Wat. Sci. Tech. Vol.40, No.7, 55-65.

[18] Nanqi Ren, Zhaobo Chen, Xiangjing Wang, Dongxue Hu, Aijie Wang, 2005. Optimized operational parameters of a pilot scale membrane bioreactor for high-strength organic wastewater treatment. International Biodeterioration & Biodegradation 56, 216-223.

[19] F. Gumrah, B. Oz, B. Guler, S. Evin, 2000. The application of artificial neural networks for the prediction of water quality of polluted aquifer. Water, Air, and Soil Pollution 119, 275-294.

[20] T.Y. Pai, Y.P. Tsai, H.M. Lo, C.H. Tsai, C.Y. Lin, 2007. Grey and neural network prediction of suspended solids and chemical oxygen demand in hospital wastewater treatment plant effluent. Computers and Chemical Engineering 31, 1272-1281.

[21] I. Machon, H. Lopez, J. Rodriguez-Iglesias, E. Maranon, I. Vazquez, 2007. Simulation of a coke wastewater nitrification process using a feed-forward neuronal net. Environmental Modeling & software 22, 1382-1387.

[22] G.M. Zeng, X.S. Qin, L. He, G.H. Huang, H.L. Liu, Y.P. Lin, 2003. A neural network predictive control system for paper mill wastewater treatment. Engineering Application of Artificial Intelligence 16, 121-129.

[23] Lee, D.S., Jeon, C.O., Park, J.M., Chang, K.S., 2002. Hybrid neural network modeling of a full-scale industrial wastewater treatment process. Biotechnol. Bioeng. 78, 670-682.

[24] Psichogios, D.C., Ungar, L.H., 1992. A hybrid neural network first principal

approach to process modeling. AIChE J. 38, 1499-1511.

[25] Thompson, M., Kramer, M.A., 1994. Modeling chemical processes using prior knowledge and neural network. AIChE J 40, 1328-1340.

[26] Martin Cote, Bernard P. A.Grandjean, Paul Lessard, Jules Thibault, 1995. Dynamic modeling of the activated sludge process: improving prediction using neural networks. Wat. Res. Vol. 29, No.4, 995-1004.

[27] Sung Hun Hong, Min Woo Lee, Dae Sung Lee, Jong Moon Park, 2007. Monitoring of a sequencing batch reactor for nitrogen and phosphorus removal neural networks. Biological Engineering Journal 35, 365-370.

[28] C.A. Gontarski, P.R. Rodrigues, M. Mori, L.F. Prenem, 2000. Simulation of an industrial wastewater treatment plant using artificial neural ntworks. Computers and Chenmical Engineering 24, 1719-1723.

[29] Dae Sung Lee, Peter A. Vanrolleghem, Jong Moon Park, 2005. Parrallel hybrid modeling methods for a full-scale cokes wastewater treatment plant. Journal of Biotechnology 115, 317-328.

[30] Ozer Cinar, 2005. New tool for evaluation of performance of wastewater treatment plant: Artificial neural network. Process Biochemistry 40, 2980-2984.

[31] Chien-Hsien Wen, C.A. Vassiliadis, 1998. Applying hybrid artificial intelligence techniques in wastewater treatment. Engineering Application of Artificial Intelligence 11, 685-705.

[32] Ahmed Gamal El-Din, Daniel W. Smith, 2002. A neural network model to predict the wastewater inflow incorporating rainfall events. Water Research 36, 1115-1126.

[33] Lluis A. Belanche, Julio J. Valdes, Joaquim Comas, Ignasi R. Roda, Manel Poch, 1999. Towards a model of input-output behavior of wastewater treatment plants using soft computing techniques. Environmental Modeling

& Software 14, 409-419.

[34] Carmen Gabaldon, Jose Ferrer, Aurora Seco, Paula Marzal, 1998. A software for the integrated design of wastewater treatment plants. Environmental Modeling & Software 13, 31-44.

[35] Taesup Moon, 2008. Prediction of influent flow rate and influent components using artificial neural network (ANN). Journal of Korean Society on Water Quality, Vol.24, No.1, 91-98.

# 신경망 모델을 이용한 하수처리장 공정효율 예측

조 영 나

환경공학과 물환경시스템

부경대학교 대학원

## 요 약

환경 규제가 점차적으로 강화됨에 따라, 하수처리장의 안정적이고 경제적인 운전을 위한 ICA (Instrumentation Control and Automation) 기술들은 빠르게 발전해 왔다. 개발된 대부분의 다른 모델들은 시간에 변화하는 많은 변수들을 가짐으로써 보정하기 어렵다는 단점을 가지는 결정론적 모델들이었다.

본 연구에서는 유출수 예측 모델의 필요성에 따라 공정으로부터 확보된 측정 데이터를 기반으로 주어진 입력변수와 목표 변수 간의 변화 패턴만을 고려하여 생성되는 Black-box 모델인 인공신경망을 사용하여, 실제 하수처리장의 1 차 침전지 유출수질 뿐만 아니라 2 차 침전지 유출수질을 예측 할 수 있었다. 인공신경망은 그 구조에 따라 Single Layer Feed-forward Network Multi-layer, Feed-back network recurrent network, self-organized network 으로 분류된다. 사용목적에 따라 적절한 형태의 신경망 구조를 선택하게 된다. 유출수 예측이 목적이기 때문에 모델링과 예측 분야에서 가장 일반적으로 사용되고 있는 feed-forward back-propagation network 을 사용하였다. 본 연구에서는 신호전달을 위한 전달함수로써 tan-sigmoid function 을 사용하였다. 그리고 역전파 알고리즘의 단점을 극복하고 더 빠른 학습이 가능한 Levenberg-Marquart (LM) 알고리즘을 학습 알고리즘으로 사용하였다. 대상 하수처리장에서 실제 측정된 2005 년 유입 수질 데이터들과 다른 데이터들을

수집하였고 전처리를 거친 200 개의 데이터을 예측 모델 개발을 위해 사용하였다. 만들어진 신경망의 예측 성능을 검증하기 위해 나머지 164 개의 데이터를 사용하였다. 1 차 침전지 유출수 예측의 경우에는 인공 신경망뿐만 아니라 결합성 신경망 모델을 하였다. 2 차 침전지 유출수 예측에서는 IWA task group 에 의해 제안된 활성 슬러지 모델 데이터와 포기조 안에 데이터들에 의해 만들어진 신경망을 사용하였다.

따라서, 시행착오법 (Trial and Error Method)에 의해 적절한 신경망 의 구조와 epoch 수를 바탕으로 만들어진 신경망을 일반화하기에 충분하게 가중치 값들을 결정하였으며 상용소프트웨어인 MATLAB 에서 제공되는 neural network toolbox 를 이용해서 모든 시뮬레이션 작업을 수행하였다.

# Acknowledgments

Time goes fast; I am going to finish my master degree. During the past two years, I have studied hard and acquired much knowledge in the area of environmental engineering. It is accepted that I am very interested in this area so that I can find many interests from it. I am very lucky to meet a good advisor, Byung-Hun Lee, an outstanding professor. He guided me into the door of science research and gave me many chances to extend my ability. His diligent attitude, profound knowledge, foresight and father-like care magnetize me strongly. And also I want to show my appreciations to the chairman, Prof. Seok-Jun Yoa, and the members, Prof. Il-Kyu Kim and Prof. Byung-Hun Lee, who gave me some precious suggestion to improve and deepen my thesis and my study.

I appreciate PhD Eun-Joo, Hwang, the leader of our lab, who gave me so much kind help for my studying and life after my coming and imparted me some basic techniques of environmental systems. I also express my thanks to Hyeon Bae and his wife, PhD of Busan National University. They gave me the cares and supports during my researches with kindness. Also I have to thank all of my lab members, Kwang-Jae Kim, Chang-Wook Jung, Hirofumi Takahashi, Jae-Sin Ha, Kwang-Hoon Choi, Eun-Myong Song, and Sang-Yeob Kim. They gave me a lot of help, happiness and cares that are important for an international student studying in Korea, especially Eun-Myong Song helped me do some experiment during my thesis period. In a word, whenever I need some help, they try their best to help me kindly.

Here I must express my thanks to a special friend, Qing Wang, who is

sometimes my teacher guiding my positive life direction and sometimes acts as my closed brother giving meticulous care to me. Especially he accompanied with me to spend the most difficult beginning periods. I will remember this overseas time forever to thank my good friend.

And it is my pleasure to thank my parents for their endless encouragement. Their remote cares accompany me all the time.

At last, please allow me to express my wishes. I bless the people of disaster areas, Sichuan Province, China, wishing the god confer well-being on these areas. Bless the lucky living get a better future life, bless the still buried people rescued quickly and bless the dead rest in peace in heaven!!

Yong-Na Zhao