Prediction of wastewater treatment plant performance based on wavelet packet decomposition and neural networks

Davut Hanbay a,*, Ibrahim Turkoglu a, Yakup Demir b

a Firat University, Technical Education Faculty, Department of Electronics and Computer Science, 23119, Elazig, Turkey
b Firat University, Department of Electrical-Electronics Engineering, 23119, Elazig, Turkey

Abstract

In this paper, an intelligent wastewater treatment plant model is developed to predict the performance of a wastewater treatment plant (WWTP). The developed model is based on wavelet packet decomposition, entropy and neural network. The data used in this work were obtained from a WWTP in Malatya, Turkey. Daily records of these WWTP parameters over a year were obtained from the plant laboratory. Wavelet packet decomposition was used to reduce the input vectors dimensions of intelligent model. The suitable architecture of the neural network model is determined after several trial and error steps. Total suspended solid is one of the measures of overall plant performance so the developed model is used to predict the total suspended solid concentration in plant effluent. According to test results, the developed model performance is at desirable level. This model is an efficient and a robust tool to predict WWTP performance.

Keywords: Wavelet packet decomposition; Entropy; Wastewater treatment plant; Total suspended solid

1. Introduction

In recent years, intelligent modeling studies related to WWTP have become popular because of the rising concern about environment. The developments in intelligent methods make them possible to use in complex systems modeling. Intelligent modeling was firstly used to increase the robustness of existing models but now it is used to obtain new models.

Complex mathematical models are including many biochemical processes. Yet, due to the scarcity of measured data, it is almost impossible to obtain reliable estimates of unknown dynamical parameters. Therefore, either simpler, manageable models are needed. So considerable efforts have been expended to change or modify the traditional wastewater treatment processes (Müller, Noykova, Gyllenberg, & Timmer, 2002). Improper operation of a WWTP may bring about serious environmental and public health problems, as its effluent to a receiving water body can cause or spread various diseases to human beings. A better control of a WWTP can be achieved by developing robust models for predicting the plant performance based on past observation so certain key parameters. However, modeling a WWTP is a difficult task due to the complexity of the treatment processes. The complex physical, biological and chemical processes involved in wastewater treatment process exhibit nonlinear behaviors which are difficult to describe by linear mathematical models (Hamed, Khalafallah, & Hassanien, 2004).

In the last decade, many studies were realized in wastewater treatment based on intelligent methods. These researches are related to modeling WWTP. These researches are about predictions of WWTP parameters, process control of WWTP, estimating WWTP output parameters characteristics. Some of these studies based on intelligent methods are as follows. A novel approach on the basis of ANN model that was designed to provide better predictions of nitrogen contents in treated effluents was reported by...
Chen, Chang, and Shieh (2003). Total suspended solid (TSS) is an indication of plant performance. A simple prediction models based on neural network for TSS was demonstrated in Belanche, Valdés, Comas, Roda, and Poch (2000). To develop a neural network model to predict long term fouling of nanofiltration membranes that are used to purify contaminated water supplies was demonstrated in Shetty and Chellam (2003). ANN-based models for prediction of biological oxygen demand (BOD) and suspended solid (SS) concentrations in plant effluent were presented in Hamed et al. (2004). A model based on ANN was developed for evolution of the pollutant concentration during irradiation time under various conditions was presented in Göb et al. (1999). The coagulation-flocculation is a major step in the drinkable water treatment process allowing the removal of colloidal particles. ANN predictor of coagulant dosage in order to facilitate process operation was reported in Gagnon, Grandjean, and Thibault (1997).

A simplified hybrid neural net approach was applied for the modeling and subsequent analysis of a chemical waste water treatment plant in Mizushima, Japan to reduce the occurrences of overflow in the clarifier caused by filamentous bulking and thereby increase waste water treatment capacity was presented in Miller, Itoyama, Uda, Takada, and Bhat (1997). The soft-sensing method based on neural networks was proposed in order to detect on-line wastewater treatment quality parameters. In that study WWT technique was analyzed systematically. The parameters which can not be detected on-line were taken as the secondary variables. The parameters which can not be detected on-line were taken as the primary variables. Back propagation (BP) NN for soft-sensing is proposed and trained using the testing data of practical treatment process. The simulation results were showed that the soft-sensing system of wastewater treatment based on BPNN can correctly estimates the quality parameters on real time (Wan-liang & Min, 2002). In another study, neural network models were used to model alum dosing of southern Australian surface waters was presented in Maier, Morgan, and Chow (2004).

In this paper, wavelet packet analysis, entropy and NN are combined to extract the features from processed data for estimating TSS as output quality parameter. The paper is organized as follows. In Section 2, we review some basic properties of used methods, wavelet packet decomposition, NNs. The intelligent model and entropy is described in Section 3. This method enables reduction of the data size and make model efficient. The effectiveness of the proposed model is demonstrated in Section 4. Finally Section 5 presents discussion and conclusions.

2. Preliminaries

In this section, the theoretical foundations for the intelligent modeling used in the presented study are given in the following subsections.

2.1. Wavelet transform

Wavelet transforms are finding inverse use in fields as diverse as telecommunications and biology. Because of their suitability for analyzing non-stationary signals, they have become a powerful alternative to Fourier methods in many medical applications, where such signals abound (Daubechies, 1998). The main advantages of wavelets is that they have a varying window size, being wide for slow frequencies and narrow for the fast ones, thus leading to an optimal time-frequency resolution in all the frequency ranges. Furthermore, owing to the fact that windows are adapted to the transients of each scale, wavelets lack the requirement of stationary.

A wavelet expansion is Fourier series expansion, but is defined by a two-parameter family of functions. It can be defined as follows:

$$f(x) = \sum_{i,j} c_{ij} \psi_{ij}(x),$$

(1)

where \(i\) and \(j\) are integers, the functions \(\psi_{ij}(x)\) are the wavelet expansion functions and the two-parameter expansion coefficients \(c_{ij}\) are called the discrete wavelet transform (DWT) coefficients of \(f(x)\). The coefficients are given by:

$$c_{ij} = \int_{-\infty}^{+\infty} f(x) \psi_{ij}(x).$$

(2)

The wavelet basis functions can be computed from a function \(\psi(x)\) called the generating or mother wavelet through translation and dilation:

$$\psi_{ij}(x) = 2^{-i/2} \psi(2^{-j}x - j),$$

(3)

where \(j\) is the translation and \(i\) the dilation parameter. Mother wavelet function is not unique, but it must satisfy a small set of conditions. One of them is multi-resolution condition and related to the two-scale difference equation:

$$\phi(x) = \sqrt{2} \sum_{k} h(k) \phi(2x - k),$$

(4)

where \(\phi(x)\) is scaling function and \(h(k)\) must satisfy several conditions to make basis wavelet functions unique, orthogonal and have a certain degree of regularity. The mother wavelet is related to the scaling function as follows,

$$\psi(x) = \sqrt{2} \sum_{k} g(k) \phi(2x - k),$$

(5)

where \(g(k) = (-1)^k h(1 - k)\). At this point, if valid \(h(x)\) is available, one can obtain \(g(x)\). Note that \(h\) and \(g\) can be viewed as filter coefficients of half band low pass and high pass filters, respectively. \(J\)-level wavelet decomposition can be computed with Eq. (6) as follows:

$$f_0(x) = \sum_{k} c_{0,k} \phi_{0,k}(k)$$

$$= \sum_{k} \left( c_{j_{1}k} \phi_{j_{1}k}(x) + \sum_{j=0}^{J} d_{j+1,k} \psi_{j+1,k}(x) \right).$$

(6)
where coefficient $c_{0,k}$ are given and coefficients and coefficient $c_{j+1,n}$ and $d_{j+1,n}$ at scale $j + 1$ and they can be obtained if coefficient at scale $j$ is available:

$$c_{j+1,n} = \sum_{k} c_{j,k} h(k - 2n)$$

$$d_{j+1,n} = \sum_{k} c_{j,k} g(k - 2n).$$

2.2. Wavelet packet decomposition

As an extension of the standard wavelets, wavelet packet represent a generalization of multi-resolution analysis and use the entire family of subband decomposition to generate an overcomplete representation of signals (Wang, Teo, & Lin, 2001). Wavelet decomposition uses the fact that it is possible to resolve high frequency components within a small time window, while only low frequency components need large time windows. This is because a low frequency component completes a cycle in a large time interval whereas a high frequency component completes a cycle in a much shorter interval. Therefore, slow varying components can only be identified over long time intervals but fast varying components can be identified over short time intervals. Wavelet decomposition can be regarded as a continuous time wavelet decomposition sampled at different frequencies at every level or scale. The wavelet decomposition functions at level $m$ and time location $t_m$ can be expressed as:

$$d_m(t_m) = x(t)\Psi_m\left(\frac{t - t_m}{2^m}\right),$$

where $\Psi_m$ is the decomposition filter at frequency level $m$. The effect of the decomposition filter is scaled by the factor $2^m$ at stage $m$, but otherwise the shape is the same at all scales (Devasahayam, 2000). Wavelet packet analysis is an extension of the discrete wavelet transform (DWT) (Burrus, Gopinath, & Guo, 1998) and it turns out that the DWT is only one of the much possible decomposition that could be performed on the signal. Instead of just decomposing the low frequency component, it is therefore possible to subdivide the whole time-frequency plane into different time-frequency pieces. The advantage of wavelet packet analysis is that it is possible to combine the different levels of decomposition in order to achieve the optimum time-frequency representation of the original (Turkoglu, Arslan, & Ilkay, 2003).

2.3. Neural networks

Neural networks (NNs) are biologically inspired and mimic the human brain. They are consisting of a large number of simple processing elements called as neurons. A schematic diagram for an artificial neuron model is shown in Fig. 1. Let $X = (X_1, X_2, \ldots, X_m)$ represent the $m$ input applied to the neuron. Where $W_i$ represent the weight for input $X_i$ and $b$ is a bias then the output of the neuron is given by Eq. (9). These neurons are connected with connection link. Each link has a weight that multiplied with transmitted signal in network. Each neuron has an activation function to determine the output. There are many kind of activation function. Usually nonlinear activation functions such as sigmoid, step are used. NNs are trained by experience, when applied an unknown input to the network it can generalize from past experiences and product a new result (Haykin, 1994).

$$u = \sum_{i=0}^{m} x_i w_i - b$$

and $$V = f(u).$$

Neural networks are systems that are deliberately constructed to make use of some organizational principles resembling those of the human brain (Haykin, 1994). They represent the promising new generation of information processing systems.

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. They represent the promising new generation of information processing systems. Neural networks are good at tasks such as pattern matching and classification, function approximation, optimization and data clustering (Bishop, 1996).

When designing a NN model, a number of considerations must be taken into account. First of all the suitable structure of the NN model must be chosen, after this the activation function and the activation values need to be determined. The number of layers and the number of units in each layer must be chosen. Generally desired model consist of a number of layers. The most general model assumes complete interconnections between all units. These connections can be bidirectional or unidirectional. We can sort the advantages of NN as follows:

- They can be implemented electrically, optically, or can be modeled on general purpose computer.
- They are fault tolerant and robust.
- They work in parallel and special hardware devices are being designed and manufactured which take advantage of this capability.
- Many learning paradigm or algorithms are available in practice.
- An ability to learn how to do tasks based on the data given for training or initial experience.
NN can create its own organization or representation of the information it receives during learning time.

There are many kinds of NN structures. One of these is multilayer feedforward NN and is shown in Fig. 2.

3. Procedure

The realization steps are as follows:

Step 1: First of all, parameters database is formed. The parameters data which have missing value are ignored. Some of these data are shown in Table 1. in this table, $T$ is temperature, COD is chemical oxygen demand, TN total nitrate. These data are normalized by

$$s(i) = \frac{s(i) - \min(s)}{\max(s) - \min(s)}.$$  \hspace{1cm} (10)

Step 2: This step is related to feature extracting and classification. Fig. 3 shows the wavelet packet and NN structure for intelligent modeling. Feature extraction is the key process for intelligent methods. So that it is arguably the most important component of modeling based on intelligent. A feature extractor should reduce the input vector (i.e., the original waveform) to a lower dimension, which contains most of the useful information from the original vector. The goal of the feature extraction is to extract features from these data for reliable intelligent modeling. For feature extraction, the wavelet packet and NN structure was used.

The wavelet packet and NN structure is composed of two layers. These are wavelet packet layer and Multilayer perceptrons layer, respectively.

Wavelet packet layer: This layer is responsible for feature extraction from input data. The feature extraction process has two stages:

Stage 1–Wavelet packet decomposition: For wavelet packet decomposition of the input data, the decomposition structure at level 3 was realized and shown in Fig. 4.

![Multilayer feed forward neural network structure.](image)

**Fig. 2. Multilayer feed forward neural network structure.**
Wavelet packet decomposition was applied to the input data using the Symlet-1 wavelet decomposition filters, \( w \).

**Stage 2—Wavelet entropy:** An Entropy-based criterion describes information-related properties for an accurate representation of a given signal. Entropy is a common concept in many fields, mainly in signal processing (Quiroga, Roso, & Basar, 1999). A method for measuring the entropy appears as an ideal tool for quantifying the ordering of non-stationary signals. We next calculated the sure entropy of the wavelet packet coefficients as defined in

\[
E(s) = \sum \min(s_i^2, p^2) \quad |s_i| \leq p,
\]

where, the wavelet entropy \( E \) is a real number, \( s \) is the terminal node signal and \( (s_i) \) is the waveform of terminal node signals. In sure entropy, \( P \) is the threshold and must be a positive number. At the WPNN training process, while the \( P \) parameter is updated by 0.1 increasing steps, the weights of the NN is updated randomly. Thus, feature vectors which have the length of four are obtained.

**Multi-layer perception (MLP) layer:** This layer is realized the classification by using features from wavelet packet layer. The training parameters and the structure of the MLP are shown in Table 2. These were selected for the best performance after several trial and error stages, such as the number of hidden layers, the size of the hidden layers, value of the moment constant and learning rate, and type

<table>
<thead>
<tr>
<th>Architecture</th>
<th>MLP architecture and training parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of layers</td>
<td>3</td>
</tr>
<tr>
<td>The number of neuron on the layers</td>
<td>Input: 4</td>
</tr>
<tr>
<td></td>
<td>Hidden: 25</td>
</tr>
<tr>
<td></td>
<td>Output: 1</td>
</tr>
<tr>
<td>The initial weights and biases</td>
<td>Random</td>
</tr>
<tr>
<td>Activation functions</td>
<td>Tangent-sigmoid</td>
</tr>
<tr>
<td>Tangent-sigmoid</td>
<td>Linear</td>
</tr>
<tr>
<td>Training parameters</td>
<td>Levenberg–Marquardt</td>
</tr>
<tr>
<td>Learning rule</td>
<td>Back-propagation</td>
</tr>
<tr>
<td>Sum-squared error</td>
<td>0.0000001</td>
</tr>
</tbody>
</table>

![Fig. 3. The structure of intelligent modeling.](image)

![Fig. 4. Total decomposition tree of wavelet packet analysis.](image)

![Fig. 5. The intelligent model training.](image)

![Fig. 6. The intelligent model training performance.](image)
of the activation functions. WPNN training performance is shown in Figs. 5 and 6.

4. Experimental results

We performed WPNN model training using 50 different input data. Another 12 input data is used for testing. Test results which are shown in Fig. 7 are showed that WWTP output characteristics are modeled correctly by WPNN. It clearly indicates the effectiveness and the reliability of the proposed approach for extracting features from input data.

5. Discussion and conclusion

This work indicates the use of wavelet packet and NN for feature extracting and classification in intelligent modeling. The application of the wavelet packet entropy and adaptive feature extraction from input data are shown. Wavelet entropy proved to be a very useful tool for characterizing the input data. This means that with this method, new information can be accessed with an approach different from the traditional analysis methods.

The most important aspect of the intelligent model is the ability of self-organization of the WPNN without requirements of programming and the immediate response of a trained net during real-time applications. These features make the intelligent model suitable for complex systems. These results point out the ability of design of a new intelligence model.

The test results of the realized model are showed the advantages of intelligent modeling; it is rapid, easy to operate, non-invasive, and not expensive.

Acknowledgement

We thank, the Malatya Municipality, Malatya, Turkey for providing the WWTP process data to us.

References


