Prediction of treated water turbidity and its parametric behavior for the coagulation process of a water treatment plant using a joint extreme learning machine-genetic algorithm

C D Jayaweera¹ and N Aziz¹

¹ School of Chemical Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Pulau Pinang, Malaysia

E-mail: chnaziz@usm.my

Abstract. A new genetic algorithm was developed to improve the performance of the extreme learning machine radial basis function (ELM-RBF) neural network model to predict treated water (TW) turbidity for the coagulation process of water treatment. The genetic algorithm was constructed such that a notable improvement in the model could be achieved, within a short period of time. Two sets of models were developed based on high and low turbidity values. The genetic algorithm improved the correlation coefficient (R) and the mean squared error (MSE) of the low turbidity model from 0.76 and 2.16 × 10⁻⁴ to 0.81 and 4.85 × 10⁻⁵ respectively, in 15 minutes; whilst improving R and MSE of the high turbidity model from 0.62 and 0.0011 to 0.93 to 2.89 × 10⁻⁴ respectively, in 2 minutes. The ability of the model to capture the reported variation of TW turbidity with gradually increasing coagulant dosage was also tested. It was noted that ELM-RBF was more capable of capturing such physical and chemical phenomena better than multilayer perceptron and ELM-single layer feed-forward models. The genetic algorithm improved the ability of the model to capture the parametric behavior of TW turbidity too.

1. Introduction
Coagulation is an essential stage of the water treatment process primarily for the removal of turbidity. It ensures the agglomeration of suspended particles via electrostatic forces generated by the ions of the coagulant, which helps improve the filterability of the particles. The size of the agglomerates increases with increasing coagulant dosage. The treated water turbidity continues to rise with increasing coagulant dosage until the agglomerates attains a filterable size. Once the particles had flocculated to form agglomerates large enough to be filtered out, the treated water turbidity begins to decrease. However, hydrophilic components of natural organic matter and extremely small particles contributing to turbidity are resistant to the coagulant content regardless of the dosage [1]. Increasing the coagulant dosage beyond the optimum value restabilizes the particles and causes the treated water (TW) turbidity to increase [2]. Thus, continuous supply of coagulant does not reduce the turbidity further. Instead, it increases the cost of treatment whilst increasing both turbidity and residual aluminum content.
In order to optimize the coagulant dosage, the coagulation process had been modeled using multilayer perceptron neural networks, radial basis function neural networks and general regression neural networks [3-7]. In [8], parametric analysis was carried out by developing TW turbidity models and observing the response of TW turbidity to changes in the coagulant dosage. Nevertheless, the response of the predicted TW turbidity lacked sensitivity to gradually increasing coagulant dosage. It was clear that employing Multi-layer Perceptron neural network (MLP) or General Regression neural network (GRNN) were ineffective in capturing such chemical and physical phenomena related to the coagulation process.

Another form of neural network known as the ‘extreme learning machine’ [9] single layer feed-forward neural network (ELM-SLFN) had also been used to predict the optimum coagulant dosage [10]. It was observed that the performance of the ELM-SLFN was not only close to that of the MLP but the former also took only a fraction of the time taken by the MLP to process the data. It was also reported that the ELM-RBF not only demonstrated a generalization ability similar to support vector machines but also consumed minimum processing time [11]. Based on this development, [12] further compared the performance of ELM-RBF, ELM-SLFN and MLP models and found that the ELM-RBF outperformed the MLP model. In this work, a genetic algorithm was developed to optimize the input weights of an ELM-RBF model developed to predict the TW turbidity using coagulation data from the Segama water treatment plant. A genetic algorithm (GA) is an adaptation derived from Darwin’s theory of evolution [13-14]. It is a well-established and commonly used method of optimization [15]. GAs are most commonly used to improve MLP neural networks by ensuring the solution reaches the global optimum by avoiding chances of it being trapped at a local minimum [16-18]. It has been argued that GAs consume time to converge to the final solution [16]. However, the time consumed could be greatly reduced when utilizing GAs with appropriate genetic operators [16] to optimize GRNNs and ELM neural networks. [19] demonstrated that genetic algorithms could be used to improve ELM ANNs as well. In the current study, a genetic algorithm was developed to improve the correlation coefficient, mean squared error and the ability to capture physical and chemical phenomena of the process.

2. Methodology

2.1. Genetic algorithm
The non-generational genetic algorithm shown in Figure 2 was constructed to maximize the performance of an ELM-RBF by optimizing the input weights (or radial basis centers) of the ELM-RBF through the following 5 steps.

1. Generating an initial population
   In this scenario, the initial population consisted of random numbers between 0 and 1, each of size n x m, m being the number of input variables and n being the number of radial basis centers. The rate of convergence decreased as the number of members in the initial population increased.

2. Chromosome encoding
   Real coding was used in this study, where real numbers (weights connecting the inputs to the radial basis layer) were used to represent positions of a chromosome.

3. Determination of a suitable fitness function
   The fitness function is generally accepted to be a maximizing function. MSE being a cost function, the inverse of MSE calculated by validation data was taken as the fitness function i.e.
   \[ f(x) = \frac{1}{MSE} \]  \hspace{1cm} (1)
4. Selection of parent chromosomes
   Parent selection was carried out based on positive assortive fitness [16], as it reduces the effective size of the population, reduces the diversity of the population and enables faster convergence.

5. Recombination and evolution
   Recombination and evolution was carried out using ‘crossover’ and ‘mutation’ which are called mating operators [16, 20]. The k-point cross over, where two chromosomes are divided at k number of points and interchanged with each other, was used. The 2-point crossover carried out in this study is demonstrated in Figure 1.

   ![2 point crossover](image)

   **Figure 1.** 2 point crossover.

   The chromosome is divided into three sections. The crossover is considered biased as only one section of the second fittest member is passed on to the offspring. The fittest two of the three offspring are used to replace the weakest of the current population. Mutation operators make a random change at a random location in a chromosome. The termination criterion used in this study is the number of matings.

   The ability of the genetic algorithm to improve the following aspects was tested.
   1. Improve the correlation coefficient and mean squared error.
   2. Ability to handle extrapolated data
   3. Ability to demonstrate the proper response of treated water turbidity to increasing coagulant dosage (hereafter referred to as the parametric analysis)

2.2 Input parameter selection
   Input parameters were selected from 12 variables (RW turbidity, RW color, RW TDS, RW Alkalinity, RW pH, TW turbidity(t-1), TW color, TW pH, TW alkalinity, TW TDS, Residual Aluminium and Alum dosage) using the global search algorithm employed by [21]. Alum dosage, TW turbidity (t-1) and RW turbidity were fixed as essential variables, when implementing the search algorithm. Combinations of input parameters that produced models with a correlation coefficient higher than 0.8 were used to carry out a parametric analysis to test how well the model captured reported physical and chemical phenomena occurring during the coagulation process. The phenomena employed for the parametric analysis was the response of treated water turbidity to gradually increasing coagulant dosage. The optimum combination of input parameters was selected based on the parametric analysis. The neural network model utilized for the search algorithm is the extreme learning machine radial basis function neural network (ELM-RBF), due to its high accuracy and less training time required [12].
2.3 **ANN model development**

In order to compare predictive performance and ability to capture reported parametric behavior, multilayer perceptron (MLP), ELM single layer feed-forward neural network (ELM-SLFN) and ELM-RBF models were developed using the selected set of input parameters. The MLP neural network was trained using the Levenberg Marquardt algorithm. The network consisted of one hidden layer with 15 hidden neurons. The number of hidden neurons was selected based on a theory employed [3] where the maximum number of neurons considered was 2I+1. ‘I’ represents the number of hidden neurons. The hidden layer activation function used for the MLP model and the ELM-SLFN was the tansigmoidal. Number of hidden nodes in the ELM-SLFN neural network was 20. The activation function used for ELM-RBF model was the radbas function ($e^{-x^2}$). The number of radial basis centers used was 20.

2.4 **Data preprocessing and data division**

Data normalization for preprocessing purposes was carried out using Equation 2.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

(2)

All values were adjusted between 0 and 1 using the above equation. Data division for model development was carried out as follows:

1. **Training** – 64% of the data set was used for training. Of the training data, 84% of raw water turbidity values were less than 0.1 (after normalizing). The training data set was utilized for calculation of output weights.

2. **Validation** – 24% of the data set was used for model validation. The validation data set was used to calculate the fitness of chromosomes i.e. radial basis centers. Of the validation data used, 77% had RW turbidity values less than 0.1.

3. **Testing** – After training the network using the genetic algorithm in Figure 2, the performance of the model was tested using unseen data which constituted 12% of the entire data set. Of the test data set used, 87% consisted of RW turbidity values less than 0.1.

Initially models were developed using the entire range of data (0 to 1). When the performance was shown to be inadequate, two sets of models were developed using data with RW turbidity values less than 0.1 and greater than 0.1 respectively. The performance (correlation coefficient and mean squared error), regression and parametric analysis were carried out under each instance and compared with a corresponding ELM-RBF that wasn’t trained with the genetic algorithm.
Start

Generate 50 chromosomes each of size (nxm)

Calculate fitness of all 50

Choose the two with highest fitness

Carry out 2-point biased crossover (3 offspring)

Choose the fittest two of the offspring

Is the fitness of either or both offspring higher than the weakest two or one?

Yes

Replace the weakest with the offspring

Apply mutation to a copy of the fittest

Is the mutated chromosome fitter than the weakest?

Yes

Replace weakest with mutated chromosome

Matings=N?

Yes

Stop

No

No

No

No

No

Figure 2. The genetic algorithm
2.5 **Extrapolation ability**

Neural networks with MLP architecture construct global approximations while RBFs construct local approximations of input-output mappings due to their exponentially decaying localized nonlinearities [22]. Therefore RBFs have very limited extrapolation ability. However the extent to which the models developed could extrapolate was evaluated.

Water quality data are given in Table 1.

| Table 1. Water quality ranges [6] |
|----------------------------------|
| Raw water                        |
| Treated water                    |
| pH, T turbidity (NTU), Col (HU), TDS, Alkalinity (mg/l) | pH, T turbidity (NTU), Col (HU), TDS, Alkalinity (mg/l), Al residual (mg/l), Coagulant dosage |
| Min                              |
| 6.6                              |
| 11                               |
| 36                               |
| 60                               |
| 50                               |
| 6.5                              |
| 0.19                             |
| 0                                |
| 90                               |
| 20                               |
| 0.01                             |
| 20                               |
| Max                              |
| 7.9                              |
| 3405                             |
| 656                              |
| 160                              |
| 170                              |
| 7.5                              |
| 5.24                             |
| 6                                |
| 170                              |
| 100                              |
| 0.20                             |
| 170                              |

3. Results and discussion

3.1. **Input selection**

Models developed using combinations of input variables that had R values higher than 0.8 are shown in Table 2.

| Table 2. Optimum combinations of input parameters |
|--------------------------------------------------|
| Model | Input parameters                                |
|-------|-------------------------------------------------|
| 1     | TW turbidity (t-1), RW turbidity, Alum dosage, TW color |
| 2     | TW turbidity (t-1), RW turbidity, Alum dosage, TW pH, TW color |
| 3     | TW turbidity (t-1), RW turbidity, Alum dosage, RW alkalinity, TW color, Residual Aluminium |
| 4     | TW turbidity (t-1), RW turbidity, Alum dosage, TW color, RW color, RW TDS (Total Dissolved Solids), TW TDS |

The Alum dosage vs. TW turbidity plots of the four models given in table 2 are shown in Figure 3.
Figure 3. Alum dosage (x) vs. TW turbidity (y) plots of Models 1(a), 2(b), 3(c) and 4(d)

The curve is expected to initially increase, in response to increasing size of agglomerates which have not yet attained a filterable size. The curve subsequently attains a maximum and gradually decrease as the particles continue to be filtered away. However, as the alum dosage is increased the remaining particles restabilize and cause the TW turbidity to increase. Therefore, the curve reaches a minimum and begins to increase up to some extent. The optimum point would be the alum dosage at the minimum point. It is evident that models 1 and 3 have failed to capture the expected variation of TW turbidity with alum dosage. Model 2 shows the expected shape, although the initial part of the curve is irregular. The shape of the curve of model 4 is smoother and meets all the expectations. Thus, model 4 was selected to proceed with the following stages.

Figure 4 demonstrates the plots for MLP and ELM-SLFN models developed using the input parameters employed in model 4.

Figure 4. MLP (left) and ELM-SLFN(right) plots of alum dosage (x) vs. TW turbidity
The correlation coefficients of all three models i.e. MLP, ELM-SLFN and ELM-RBF, were approximately equal to 0.84. However, according to Figure 4, MLP and ELM-SLFN models have not adequately captured the parametric behavior of TW turbidity. The plot representing ELM-RBF model’s ability to capture parametric behavior of TW turbidity, shown in Figure 3 (d), is smoother and more accurate than Figure 4. Therefore, ELM-RBF model architecture was chosen to proceed with the following stages of the study.

3.2. Model development

3.2.1. Model development using the entire range of data. The genetic algorithm in Figure 2 was employed to improve the prediction accuracy of the model. The termination criterion used in the algorithm is the number of matings. The model demonstrated a tendency to overfit the data in the validation data set when a high number of matings was used. Therefore selection of an appropriate number of matings was essential. The number of matings (N) employed to train models developed using the entire range of data is 500. The time taken for training the model was 30 minutes. Table 3 shows a comparison of the performance of the models developed with and without the genetic algorithm (A-GA and A-WGA respectively).

| Models  | Training          | Validation        | Test     |
|---------|-------------------|--------------------|----------|
|         | MSE   | CC     | MSE   | CC     | MSE   | CC  |
| A-GA    | 4.45×10^{-4}   | 0.8832 | 6.85×10^{-5} | 0.9613 | 1.32×10^{-4} | 0.8268 |
| A-WGA   | 3.84×10^{-4}   | 0.8998 | 0.1352 | 0.8587 | 4.59×10^{-4} | 0.8688 |

It is evident that the GA model has overfitted to the validation data set. The test correlation coefficient of A-GA is lesser than that of A-WGA. However, the test MSE of A-GA has improved. The regression plots in Figure 5 demonstrate the distribution of predicted test data relative to actual values.

![Figure 5](image_url)

**Figure 5.** A-GA(left) and A-WGA (right) plots of predicted TW turbidity (y) vs. actual TW turbidity (x)

It is apparent that the majority of the data points lie below 0.05, the reason being RW turbidity of majority of the data points being under 0.1. The regression of A-GA is evidently poorer than A-WGA.

As majority of turbidity values are less than 0.1, it was speculated that the data set could be divided to low and high turbidity data. The low turbidity data set would consist of RW turbidity values less than 0.1, and the high turbidity data set would consist of values greater than 0.1. ‘N’ was also minimized in order to prevent overfitting and to reduce computational time.
3.2.2. \textit{RW turbidity < 0.1}. The performance of models developed using the low turbidity data set is shown in Table 4.

\textbf{Table 4. Performance of models developed using data with RW turbidity less than 0.1}

| Models   | Training      | Validation    | Test         |
|----------|---------------|---------------|--------------|
|          | MSE           | CC            | MSE          | CC            | MSE          | CC            |
| B-GA     | $6.41 \times 10^{-5}$ | 0.9378       | $2.66 \times 10^{-5}$ | 0.8205       | $4.85 \times 10^{-5}$ | 0.8109       |
| B-WGA    | $1.14 \times 10^{-4}$ | 0.8867       | $1.05 \times 10^{-4}$ | 0.8095       | $2.16 \times 10^{-4}$ | 0.7640       |

The test performance, in terms of both MSE and CC, has been significantly improved by the genetic algorithm. The training and validation performances of B-GA are also better than B-WGA. The value for ‘N’ used in this instance was 200. The time taken to train the model was 15 minutes. The regression plots of B-GA and B-WGA is shown in Figure 6.

Values predicted by the model trained with GA has converged significantly closer to the y=x line. A significant decrease in the dispersion of values in the GA regression plot could observed in x values less than 0.02. The reduction in the dispersion of data in the GA regression plot supports the significant reduction in the mean squared error that was noted in Table 4.

\textbf{Figure 6.} B-GA(left) and B-WGA (right) plots of predicted TW turbidity (y) vs. actual TW turbidity (x).

Parametric analysis was carried out by gradually increasing the alum dosage, while keeping other input variables constant. Three random points from training, validation and test data sets were selected and parametric plots corresponding to each data point were drawn.
The numerical accuracy of all parametric plots is questionable. However, the shapes of the curves obtained were fairly accurate. It is evident that the shapes of the curves are different from one another. However, the genetic algorithm has moderated the shape of the curves as noted in Figure 7 (c) and (d). The expected variation of TW turbidity with alum dosage is sufficiently represented by Figure 7 (b), (d) and (f).

3.2.3. **RW turbidity > 0.1**. The performance of models developed using high turbidity data is shown in Table 5.
Table 5. Performance of models developed with data with RW turbidity greater than 0.1

| Models | Training | Validation | Test  |
|--------|----------|------------|-------|
|        | MSE      | CC         | MSE   | CC   | MSE   | CC   |
| C-GA   | 0.0015   | 0.9126     | 1.36×10^{-4} | 0.9760 | 2.89×10^{-4} | 0.9267 |
| C-WGA  | 0.0015   | 0.9120     | 0.0013 | 0.9419 | 0.0011 | 0.6173 |

A significant improvement in the test performance has occurred from the use of the genetic algorithm. Training the model with a wider range of data had helped improve the correlation coefficient of the all models. However, it can be noted that the training MSE of the GA model has not improved. The GA model was fitted to the validation data set. Therefore, the input weights may have overgeneralized output values of the training data set. The time taken for training the model with the genetic algorithm is only 2 minutes, due to lesser number of data points. The value for ‘N’ employed is 200. The regression plots of C-GA and C-WGA are shown in Figure 8.

![Figure 8. C-GA(left) and C-WGA (right) plots of predicted TW turbidity (y) vs. actual TW turbidity (x)](image)

Data predicted by C-GA model are more aligned with the y=x line than data predicted by the C-WGA model. The decrease in the dispersion of data on the C-GA regression plot demonstrates the decrease in MSE values of predicted test data.

Plots of the parametric analysis carried out with three data points from training, validation and testing, are shown in Figure 9.
The shapes of the curves have improved when data with a wider range of RW turbidity values had been used. The genetic algorithm has improved the shapes of the curves in all three plots more successfully, in this case.

3.3. Extrapolation

The extrapolation ability of the models was tested using the models developed in section 3.2.2. B-GA and B-WGA were developed using data with RW turbidity less than 0.1. Therefore, in this section, the performance of the models in predicting values between 0.1 and 0.2, and, 0.2 and 0.3, was tested. Results of this analysis are given in Table 6.

Figure 9. C-WGA training (a), C-GA training (b), C – WGA validation (c), C-GA validation (d), C-WGA test (e), C-GA test (f) plots of alum dosage (x) vs. TW turbidity
Table 6. Performance of models when tested with extrapolated data

| RW turbidity values | Models  | Test       |
|---------------------|---------|------------|
|                     |         | MSE        | CC         |
| 0.1 – 0.2           | B-GA    | 4.79×10⁻⁴ | 0.6438     |
|                     | B-WGA   | 0.0561     | 0.7476     |
| 0.2 – 0.3           | B-GA    | 0.0812     | 0.3582     |
|                     | B-WGA   | 0.0068     | 0.4669     |

B-GA model has demonstrated relatively satisfactory results for RW turbidity values between 0.1 and 0.2. However, its performance has dropped significantly, compared to B-WGA model, for predicting TW turbidity values for data with RW turbidity values greater than 0.2. Therefore the genetic algorithm may have caused a limitation in the extrapolation ability.

4. Conclusion
The optimum combination of input parameters that could capture reported physical and chemical phenomena, while demonstrating acceptable performance, was TW turbidity (t-1), Alum dosage, RW turbidity, RW color, TW color, RW TDS and TW TDS. It was apparent that ELM-RBF model was more capable of capturing the parametric behavior of TW turbidity, than MLP or ELM-SLFN models. It was shown that the genetic algorithm improved the model performance and the parametric behavior more effectively when two separate sets of models were developed, based on RW turbidity values. Thus two models, for low and high turbidity water, were developed. Moderating the value of ‘N’, which is the number of matings, prevented overfitting the model to the data, while reducing the time taken for training. The genetic algorithm employed in this research was capable of improving the model performance (correlation coefficient and the mean squared error) and the ability to capture physical and chemical phenomena, of both low and high turbidity models. The extrapolation ability of RBFs is generally limited. Both models (GA and WGA) poorly predicted extrapolated data. However, the genetic algorithm constructed in this research is capable of notably improving the performance of ELM-RBF model developed to predict treated water turbidity, with satisfactory generalization. This would enable the development of a hardware sensor as part of an early warning system which would reduce process upsets.

Acknowledgements
The cooperation of Sabah Water Supply Department and LDWS for supplying Segama Water Treatment Plant data is greatly acknowledged.

References
[1] Matilainen A, Vepsäläinen M and Sillanpää M 2010. Natural organic matter removal by coagulation during drinking water treatment: A review Advances in Colloid and Interface Science 159(2) 189-97.
[2] Malik Q H 2018 Performance of alum and assorted coagulants in turbidity removal of muddy water Applied Water Science, 8(1), 1-4.
[3] Maier H 2004 Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters Environmental Modelling & Software 19(5) 485-94.
[4] Wu G and Lo S 2008 Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system", Engineering Applications of Artificial Intelligence 21 1189-95.
[5] Wu G and Lo S 2010 Effects of data normalization and inherent-factor on decision of optimal coagulant dosage in water treatment by artificial neural network Expert Systems with Applications 37(7) 4974-83.
[6] Robenson A, Shukor S and Aziz N 2009 Development of Process Inverse Neural Network Model to Determine the Required Alum Dosage at Segama Water Treatment Plant Sabah, Malaysia Computer Aided Chemical Engineering 525-30.

[7] Griffiths K and Andrews R 2011 The application of artificial neural networks for the optimization of coagulant dosage Water Science & Technology: Water Supply 11(5) 605-11.

[8] Kennedy M, Gandomi A and Miller C 2015 Coagulation modeling using artificial neural networks to predict both turbidity and DOM-PARAFAC component removal", Journal of Environmental Chemical Engineering 3(4) 2829-38.

[9] Huang G, Zhu Q and Siew C 2006 Extreme learning machine: Theory and applications Neurocomputing, 70(1-3) 489-501.

[10] Deng X and Lin C 2016 Application of ELM to predict the coagulant dosing in water treatment plant Water Supply 17(4) 1053-61.

[11] Guang-Bin H and Chee-Kheong S 2004 Extreme learning machine: RBF network case. ICARCV 2004 8th Control, Automation, Robotics and Vision Conference.

[12] Jayaweera C and Aziz N 2018. Development and comparison of extreme learning machine and multi-layer perceptron neural network models for predicting optimum coagulant dosage for water treatment Journal of Physics: Conference Series 1123 012032.

[13] Miller G, Todd P and Hegde S 1989 Designing neural networks using genetic algorithms.

[14] Whitley D 1995 Genetic algorithms and neural networks Genetic algorithms in engineering and computer science.

[15] Anglada E, Martinez-Jimenez L and Garmendia I 2017 Performance of Gradient-Based Solutions versus Genetic Algorithms in the Correlation of Thermal Mathematical Models of Spacecrafts International Journal of Aerospace Engineering 1-12.

[16] Leardi R 2003 Nature-inspired methods in chemometrics Amsterdam: Elsevier.

[17] Bowden G J, Dandy G C and Maier H R 2005 Input determination for neural network models in water resources applications Part 1—background and methodology Journal of Hydrology 301(1-4) 75-92.

[18] Bowden G J, Maier H R and Dandy G C 2005 Input determination for neural network models in water resources applications Part 2 Case study: forecasting salinity in a river Journal of Hydrology 301(1-4) 93-107.

[19] Qingfeng X 2016 Improvement on ELM-RBF Neural Network using genetic algorithm. ILUR Transactions On Science And Engineering, 1-6.

[20] Biglarjoo N, Mirbagheri S, Bagheri M and Ehteshami M 2017 Assessment of effective parameters in landfill leachate treatment and optimization of the process using neural network, genetic algorithm and response surface methodology Process Safety and Environmental Protection 106 89-103.

[21] Jayaweera C, Othman M and Aziz N 2019 Improved predictive capability of coagulation process by extreme learning machine with radial basis function Journal of Water Process Engineering 32 100977.

[22] Neuman K 2013 Reliability of Extreme Learning Machines (PhD thesis). Bielefeld University. Germany.

[23] Goonatilake S and Khebbal S 1995 Intelligent hybrid systems Chichester: Wiley.

[24] Ruan J, Zhang C, Li Y, Li P, Yang Z, Chen X, Huang M and Zhang T 2017 Improving the efficiency of dissolved oxygen control using an on-line control system based on a genetic algorithm evolving FWNN software sensor Journal of Environmental Management 187 550-59.

[25] Bagheri M, Mirbagheri S, Bagheri Z and Kamarkhani A 2015Modeling and optimization of activated sludge bulking for a real wastewater treatment plant using hybrid artificial neural networks-genetic algorithm approach Process Safety and Environmental Protection 95 12-25.
[26] Valentin N, Denoeux T and Fotoohi F 1999 An hybrid neural network based system for optimization of coagulant dosing in a water treatment plant International Joint Conference on Neural Networks Proceedings 5 3380-5.

[27] Kim C and Parnichkun M 2016 MLP, ANFIS, and GRNN based real-time coagulant dosage determination and accuracy comparison using full-scale data of a water treatment plant Journal of Water Supply: Research and Technology - Aqua 66(1) 49-61.

[28] Demuth H, Beale M and Hagan 2007 Neural network toolbox MathWorks.

[29] Zangooei H, Delnavaz M and Asadollahfardi G 2016 Prediction of coagulation and flocculation processes using ANN models and fuzzy regression. Water Science and Technology 74(6), 1296-311.