Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2019.23357

140 (2019) 118–131
February

Applying GMDH-type neural network for modeling and prediction of turbidity and free residual aluminium in drinking water

Alahyar Daghbandan, Saba Khalatbari*, Mohammad Mahdi Abbasi

Department of Chemical Engineering, University of Guilan, Guilan, Iran, Tel. +989111960853; email: khalatbari.saba92@gmail.com (S. Khalatbari), Tel. +989111318785; email: daghbandan@guilan.ac.ir (A. Daghbandan), Tel. +989357019923; email: m_a2510@yahoo.com (M.M. Abbasi)

Received 13 April 2018; Accepted 26 October 2018

ABSTRACT

Water has a fundamental role in human life. The proper quality and quantity of water is therefore an important issue. Coagulation and flocculation are essential processes for turbidity removal from drinking water. Metals such as aluminium have been implicated in the pathogenesis of Alzheimer's disease. In this study, group method of data handling (GMDH)-type neural networks have been used for modeling and prediction of turbidity and free residual aluminium in drinking water. To validate the proposed model, a case study was carried out based on the data sets obtained from Guilan WTP. For modeling, the experimental data were divided into train and test sections (70% for training and 30% for testing). Eventually, the results of modeling were compared with experimental data and demonstrated good data compliance, with the coefficient of determination (R^2) in GMDH-type network was 0.8239 and 0.9138 for residual turbidity and residual aluminium, respectively. Moreover, the results of error analysis showed good performance of the proposed models, in this regard can be referred to the mean square error results which obtained 0.0248 for residual turbidity and 0.00000438 for residual aluminium.

Keywords: Water treatment; Turbidity; Free residual aluminium; Coagulation and flocculation; Modeling; Group method of data handling

1. Introduction

Water quality is becoming an ever more important issue, as water of low quality causes many significant problems. In particular, there is a wide range of microbial and chemical constituents of drinking water that can cause either acute or chronic detrimental health effects, and the detection of these constituents in treated water is often time-consuming, complex, and expensive [1]. However, water of poor quality can also be harmful from an economic perspective, as resources have to be directed toward improving the water supply system every time a problem arises. For these reasons, there is growing pressure to improve water treatment and water quality management in order to ensure safe drinking water at reasonable costs. Systematic assessments of raw water, treatment processes, and operational monitoring issues are needed to meet these challenges.

There are many parameters, which can be used to measure the quality of water, of which turbidity is a common one, the purpose being to measure impurities in the water. In a physical sense, turbidity is a reduction in the clarity of water due to the presence of suspended or colloidal particles, and it is commonly used as an indicator of the general condition of drinking water [1]. In addition, turbidity has been used for many decades as an indicator of the efficiency of drinking water coagulation and filtration processes, so that it is an important operational parameter for this reason, too. High turbidity values refer to poor disinfection and possibly to fouling problems in the distribution network, so that turbidity should be minimized [2]. However, turbidity is a quite sensible and faulty measurement, and many variables

* Corresponding author.

and phenomena are influencing it. This makes turbidity challenging for modeling purposes [1,2].

Another important quality parameter for treated water is residual aluminium, especially when aluminium flocculants are used in the treatment process [2]. Residual aluminium causes turbidity in water networks, resulting in acceptability problems for consumers. Usually the phenomenon can be seen when residual aluminium exceeds 0.1–0.2 mg/L, which are the usual guideline levels for residual aluminium [1]. In addition, aluminium has been suggested as a causal factor in Alzheimer's disease, in part because of reports showing the toxicity of aluminium, the elevation of aluminium concentrations in the brains of patients with Alzheimer's disease, and an association between aluminium concentrations in water and the prevalence of Alzheimer's disease [3]. Some epidemiological studies show that there can be a correlation between neural disorders and Al concentrations of 0.1 mg/L in the drinking water [4].

Many chemical and physical features of raw water affect the water treatment process. As an example of physical parameters, the water temperature has a remarkable influence on the flocculation in water treatment processes [5–7]. Naturally, process conditions also have a great effect. The dose of the flocculation chemical is naturally the key parameter, as is the adjusted pH value [2]. Turbidity can affect or be affected by the physical, microbiological, chemical and radiological characteristics of water [8]. Electrolytic conductivity, which is also called specific conductance, is a useful test in raw water for quick determination of minerals. Other important parameters that influence coagulant dosage in the water treatment process are colour and suspended solids [9].

In recent years, modeling and optimization have received great importance in most areas. Optimization allows for a better understanding of the needed system and in predicting the system behavior. In modeling, obtaining a non-linear relationship between input and output is the most important part that in classical way is often expensive and time-consuming and has little flexibility against sudden changes [10]. To overcome such problems, researchers decided to use neural networks to solve these problems.

The group method of data handling (GMDH) algorithm introduced by Ivakhnenko [11] is a heuristic self-organization process that establishes an input–output relationship within a complex system. It utilizes a multi-layered conceptual structure, similar to a feed forward multilayer neural network [11]. Ikeda et al. [12] added a recursive procedure to the GMDH algorithm to utilize updated observation data and to modify parameters within the nodes of each layer, enabling time-variable modeling. They subsequently applied the enhanced model to the prediction of daily river flows. Tamura and Kondo [13] utilized the prediction of sum of squares or Akaike's information criterion as parameter selection indicators. Because the algorithm can easily generate high-level non-linear terms, this non-linear dynamic system can be well defined; however, its practicality would be seriously reduced [13]. In response, Yoshimura et al. [14] improved the model with a stepwise regressive procedure, returning the complex final system to a low-level non-linear system, thereby increasing its applicability. The GMDH algorithm enables the automatic selection of input variables during model construction, as well as a hierarchical polynomial regression of necessary complexity [15]. Specific functional dependence between the input and output variables is unnecessary, as the dependence has been incorporated into the modeling structure. The GMDH algorithm has been applied in various fields, for example, weather modeling, pattern recognition, physiological experiments, cybernetics, medical science, education, ecology, safety science, economics and hydraulic field engineering systems [16–31].

Nowadays many studies are being performed on prediction of residual turbidity and aluminium. For example, Chaves and Kojiri [32] developed the method, named stochastic fuzzy neural network, and successfully applied to the optimization of the monthly operational strategies considering maximum water utilization and improvements on water quality in Barra Bonita reservoir located in the Sao Paulo State of Brazil. In 2012, Juntunen et al. [33] employed a multivariate linear regression method and a non-linear modeling method (multi-layer perceptron [MLP]) to model turbidity and residual aluminium in a water treatment in Finland. The results showed no significant difference in the turbidity case and only a small difference in the residual aluminium case between the linear and non-linear method [33]. Rak [34] developed a flexible Bayesian model of neural network, Gaussian processes and mixtures that demonstrate Markov chains of Monte Carlo Methods to model water turbidity during the water treatment process in Sosnowka reservoir located in Finland. Kennedy et al. [35] evaluated four different types' neural network models, including two hybrid MLP models, a generalized regression neural network model and radial basis function model for predicting both turbidity and dissolved organic matter removal at the Akron water treatment plant in Ohio, USA. The results showed two hybridizations of MLP were selected as the best model for predicting both outputs [35].

Although the above methods may be used to predict residual turbidity and aluminium, the disadvantages should be taken into consideration. For instance, artificial neural networks have a poor reproducibility because the weight and bias between neurons are given randomly and therefore are easy to fall into a local optimal solution, thus they cannot provide an optimal condition for the engineering application in water treatment [36]. In this research, multi-objective uniform-diversity genetic algorithm (MUGA) [37] is used for Pareto optimal design of a GMDH-type neural network. This method is developed according to a series of steps, including solutions encoding, fitness computation based on the objective function, selection of the best chromosomes and the genetic propagation of chosen parent chromosomes by genetic operators, such as crossover and mutation. Both crossover and mutation are implemented to produce the new and better populations of chromosomes. Also, this method would be practical when the number of input variables is large while training dataset is relatively small. Furthermore, the outputs are formed as a mathematical equation which is used in optimizing, simulating and developing integrated software.

In this study, modeling was conducted with six inputs including pH, temperature, initial turbidity, electrolytic conductivity and also the amount of chemical injection, coagulant (polyaluminium chloride [PAC]) and coagulant assistant (polyelectrolyte) that the effect of input parameters on the residual turbidity and aluminium was studied. To ensure

the accuracy of the proposed approach toward behavior of experimental data, coefficient of determination (R^2) and mean square error (MSE) were used.

2. Materials and methods

Clean output water from treatment plants is one of the highly important parameters with regard to the inlet water to treatment plant from natural sources such as rivers. A schematic diagram of experimental plant is shown in Fig. 1. Raw water is supplied from bed of river and surface waters, and a screen filter is installed to remove debris that might cause problems with pumps. In order to provide proper height of raw water, a pumping station where multiple centrifugal pumps are installed is considered. Output water from primary pumping station enters to divider of primary sedimentation ponds. In this unit, chemicals (polyelectrolyte) are added to water in order to accelerate the settling process; and copper sulfate and lime milk are used to control the growth of rooted aquatic plants. In the primary sedimentation, after mixing with chemicals, the raw water is directed to six circular ponds, and then in the next stage, after injecting polyaluminium chloride as coagulant and polyelectrolyte as coagulant assistant into water, it first enters to coagulation and flocculation unit and then settling basin. After coagulation, filtration is used to remove suspended solids in the liquid, which is the last stage of treatment to remove suspended solids in the water. In water filtration output, disinfectants (secondary chlorine) are used to destroy or deactivate the pathogenic microorganisms (pathogens) including bacteria, seaweeds, viruses, etc.

In order to measure pH, the pH meter device (SENTIX41 model, WTW company, Germany) and conductivity meter and digital thermometer (WTW company, and cond 330i/tetrocon325 model) turbidity meter (N2100.model, HACH company, USA) were used. To evaluate the effect of residual turbidity, jar test was used. This experimental study was performed in laboratory scale (80 series experiments in jar container) over a period of a year (from July 2016 to July 2017) in chemical laboratory of Guilan water treatment plant located in Rasht, Guilan, Iran. The plant uses mainly surface water from Sepidrood River or Shahrebijar River and provides drinking water of cities and villages in central regions and parts of East and West regions of Guilan province. In this period polyaluminium chloride applied as coagulant and polyelectrolyte operated as assistant coagulant. At the beginning of each series of tests, four parameters of raw water such as initial turbidity, electrical conductivity, pH and temperature were assigned as inputs. The jar test was used to identify the most adapted mix of chemical compounds and concentrations for coagulation–flocculation. It was a batch test consisting of using several identical jars containing the same volume and concentration of feed, which were charged simultaneously with six different doses of a potentially effective coagulant. The six jars could be stirred simultaneously at known speeds. The treated feed samples were mixed rapidly and then slowly and then allowed to settle. These three stages were an approximation of the sequences based on the large-scale plants of rapid mix, coagulation flocculation and settling basins. At the end of the settling period, test samples were drawn from the jars and turbidity of *supernatant* liquid were measured. A plot of *turbidity* against coagulant dose gave an indication of the optimum dosage (i.e., the minimum amount required to give acceptable clarification). So, a water sample with the lowest turbidity was identified and the spectrophotometer (DR 6000 model, HACH company) showed the amount of free residual aluminium. The lower, upper and average bounds of the data are shown in Table 1 and samples of experimental data including obtained input and output from Guilan water treatment plant are shown in Table 2.

The results of this experiment are used to provide a non-linear model to predict water quality. The model consists of six parameters as input and two outputs that is amount of residual turbidity and aluminium.

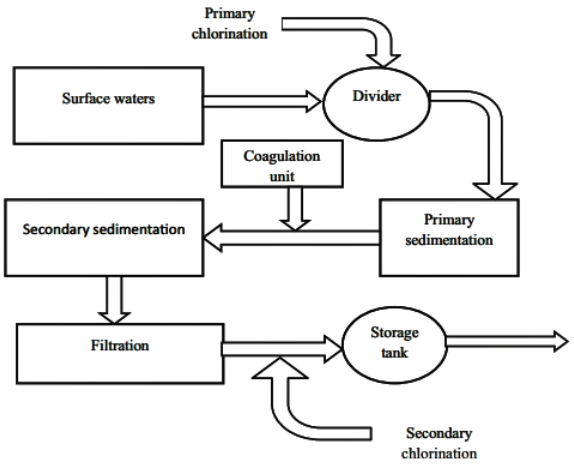

Fig. 1. Water treatment stages in treatment plant.

3. GMDH type – neural network

The GMDH is a basic technique for self-organized learning. It enables the researchers to control the process of the complex model from the input set to the output data and to determine the model parameters [38].

3.1. Basic theory

Imagine for a given input vector $X = (x_1, x_2, x_3, ..., x_N)$ we want to predict the output ($\hat{y}_i$) as close as possible to the actual output (y_i). Therefore, given M observations of multi-input-single-output data pairs, we have:

Table 1
Range of collected data from tapping unit

Inputs	Minimum	Maximum	Average
Temperature (°C)	12.7	23.4	16.59
pH	7.33	8.6	8.05
Electrolytic conductivity (μs/cm)	140.8	290	200.5
Initial turbidity (NTU)	1.92	197	46.04
Polyelectrolyte (ppm)	0.05	0.32	0.104
Polyaluminium chloride (ppm)	1	11.5	3.62

Table 2
Sample of experimental data including obtained input and output from Guilan water treatment plant

Test no.	Model inputs						Model outputs	
	Polyelectrolyte (ppm)	pH	Initial turbidity (NTU)	Polyaluminium chloride (ppm)	Temperature (°C)	Electrolytic conductivity (ms/cm)	Residual turbidity (NTU)	Free residual aluminium (ppm)
1	0.06	7.78	7.23	1.5	15.7	182.2	0.45	0.053
2	0.11	8	43.6	4	19	174	2.2	0.13
–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–
79	0.18	8.34	127	7.5	15.9	213	1.4	0.08
80	0.15	7.75	118	6.5	13.5	176	0.88	0.066

$$y_i = f(x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in}) \quad (i = 1, 2, 3, ..., M) \tag{1}$$

To train the GMDH network to predict ($\hat{y}_i$) with a given input vector $X = (x_1, x_2, x_3, ..., x_N)$, we have:

$$\hat{y}_i = \hat{f}(x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in}) \quad (i = 1, 2, 3, ..., M) \tag{2}$$

Now the GMDH neural network should be determined such that the square difference between the actual output and predicted value is minimized:

$$E = \sum_{i=1}^{M}\left(\hat{f}(x_{i1}, x_{i2}, x_{i3}, ...x_{in}) - y_i\right)^2 \longrightarrow \text{Min} \tag{3}$$

The GMDH neural network establishes a general mapping between the input and output variables as a function of the non-linear Volterra equation as follows [39]:

$$\hat{y} = a_0 + \sum_i^m a_i x_i + \sum_{i=1}^m \sum_{j=1}^m a_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m x_i a_{ijk} x_i x_j x_k + ... \tag{4}$$

This equation is represented as a Kolmogorov–Gabor polynomial [15]. This mathematical equation can also be represented by a system of partial quadratic polynomials containing only two variables (neurons) in the form of:

$$\hat{y} = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i^2 + a_4 x_j^2 + a_5 x_i x_j \tag{5}$$

The purpose of GMDH is to determine coefficient a_i in Eq. (5) using regression in order to minimize the difference between the actual output (y) and predicted output ($\hat{y}$) for each pair (x_i, x_j) of input variables [40,41]. Therefore, to optimize the coefficients of each quadratic function, the principle of least-squares error is employed as follows:

$$E = \frac{\sum_{i-1}^m (\hat{y}_i - y_i)^2}{M} \rightarrow \text{Min} \tag{6}$$

In the basic GMDH algorithm form, all possibilities of two independent variables out of a total of n input variables are used to construct the regression polynomial in the form of Eq. (5) that best fits the dependent observation (y_i, $i = 1, 2, 3, ..., M$) with a least squares sense. Consequently $\binom{n}{2} = \frac{n(n-1)}{2}$ neurons will be built in the first hidden layer of the feedforward neural network from the observations $\{(y_i, x_{ip}, x_{iq}); (i = 1, 2, 3, ..., M)\}$ for different ($p, q \in \{1, 2, 3, ..., n\}$). Hence, for each M data triples the following matrix is presented:

$$Y = Aa \tag{7}$$

where $Y = \{y_1, y_2, y_3, ..., y_M\}$ is the observed output vector, $a = (a_1, a_2, a_3, a_4, a_5)$ is the vector coefficient of the quadratic polynomial, and A is computed as follows:

$$A = \begin{bmatrix} 1 & x_{1p} & x_{1q} & x_{1p}x_{1q} & x_{1p}^2 & x_{1q}^2 \\ 1 & x_{2p} & x_{2q} & x_{2p}x_{2q} & x_{2p}^2 & x_{2q}^2 \\ 1 & x_{3p} & x_{3q} & x_{3p}x_{3q} & x_{3p}^2 & x_{3q}^2 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{Mp} & x_{Mq} & x_{Mp}x_{Mq} & x_{Mp}^2 & x_{Mp}^2 \end{bmatrix} \tag{8}$$

The least squares process from multiple regression analysis generates a normal equation solution as follows:

$$a = (A^T A)^{-1} A^T Y \tag{9}$$

which calculates the vector of the best coefficients of quadratic equation (Eq. (5)) for the entire set of M data triples. It should be noted that this process is repeated for each neuron in the next hidden layer according to the network's connectivity topology.

3.2. Applying the genetic algorithm to the generalized structure GMDH (GS-GMDH)

Genetic algorithms (GAs) are more efficient than traditional gradient methods [42] and are utilized to train neural networks with coefficients and associated weights. The classical GMDH algorithm can be in the form of a set of neurons, whereby in each layer various neuron pairs are connected and associated with a quadratic polynomial, thus creating new neurons in the next layer.

Therefore, it is possible to produce a simple and novel encoding scheme applicable to the evolutionary design of the generalized structure GMDH (GS-GMDH) where the connectivity configuration is not limited to adjacent layers, as specified initially by Nariman-Zadeh and Jamali [43].

The GS-GMDH encoding scheme involves GA and two objective functions, that is, training error (TE) and prediction error (PE) and presents accurate solutions. This kind of GMDH must exhibit the ability to specify different sizes and lengths of such neural networks [44]. GS-GMDH is summarized below.

Neuron 14 in the first hidden layer is connected to the output layer directly and passes to the second layer. Hence, the output layer neuron is denoted by 12231414 (with 14 twice). In fact, a neuron 1223 in the second hidden layer is used to induce output neuron 12231414 by constructing a virtual neuron 1414 created in the second hidden layer (Fig. 2). If a neuron traverses many adjacent layers and connects to another neuron in the next second, third, fourth or following hidden layers, the above iteration takes place. The number of neuron iterations depends on the number of hidden layers traversed, $n^{\sim}$, which is computed as $2n^{\sim}$. It is also noted that chromosome 1212 2323 (not the same as chromosome 1212 1323) is not valid and should be rewritten as 1223.

The crossover and mutation genetic operators are utilized to induce two offspring from the parents. The crossover of two chosen individuals, substitutes two chromosomes' tails by a random point selected according to Fig. 3. The building block information of GS-GMDH can be substituted by crossovers, as seen in Fig. 4. Different lengths of chromosomes created via such crossover operation lead to varying GS-GMDH network structure sizes. The population diversity is related to the mutation operation, which is easily substituted by different chromosome genes to other possible symbols, for example, 12231414 to 12233414. These operations are repeated until a valid chromosome is created. The flowchart of the GS-GMDH is shown in Fig. 5.

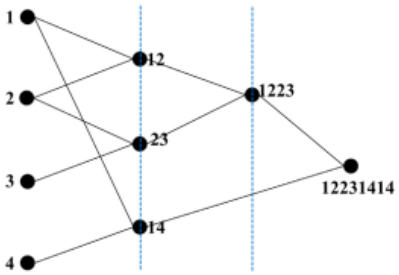

Fig. 2. GS-GMDH network chromosome structure.

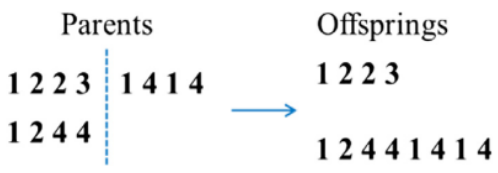

Fig. 3. Crossover operation of two individuals in the GS-GMDH model.

3.3. Multi-objective Pareto optimization

Multi-objective optimization which is also called multi-criteria optimization or vector optimization has been defined as finding a vector of decision variables satisfying constraints to give optimal values to all objective functions. In multi-objective optimization problems (MOPs), there are several objective or cost functions (a vector of objectives) to be optimized (minimized or maximized) simultaneously. These objectives often conflict with each other so that improving one of them will deteriorate another. Therefore, there is no single optimal solution as the best with respect to all the objective functions. Instead, there is a set of optimal

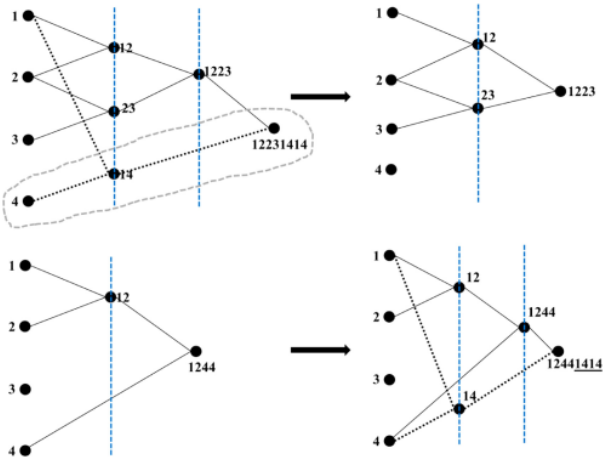

Fig. 4. Crossover operation in two GS-GMDH networks.

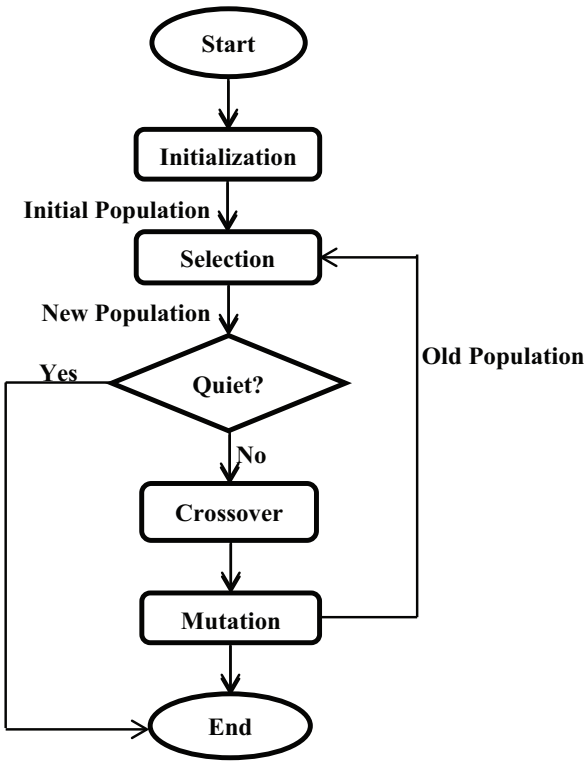

Fig. 5. Flowchart of GS-GMDH neural network.

solutions, known as Pareto optimal solutions or Pareto front for MOPs. Such multi-objective minimization based on the Pareto approach explained in a study by Jamali et al. [37].

3.3.1. Multi-objective uniform-diversity genetic algorithm

The MUGA uses non-dominated sorting mechanism together with an ε-elimination diversity preserving algorithm to get Pareto optimal solutions of multi-objective optimization more precisely and uniformly. A sorting procedure to constitute a front could be simply accomplished by comparing all the individuals of the population and including the non-dominated individuals in the front [37].

4. Results

Genetic algorithm is used to create optimum chromosome in which an initial population is considered for optimization, and new chromosomes are produced with mutation probability functions and the new chromosome crossover. Genetic stage parameters are shown in Table 3. Chromosome with lowest experimental and training error is selected as the top chromosome that is shown by Pareto curve in Figs. 6 and 7. Considering the graph, point A has the greatest training error and the least experimental error. In this paper, the purpose is to optimize the training and experimental error that point M is the optimum due to the high correlation coefficient point. The lowest training and testing error according to Pareto's figure is given in Tables 4 and 5. The final premier chromosomes to predict the amount of residual aluminium and turbidity in water is shown in Tables 6 and 7. It should be noted that each number

Table 3
Structural parameters of GMDH model

Population size	400	Crossover	0.98
Number of iteration	1,300	Number of objective function	2
Mutation	0.1	Number of hidden layer	6

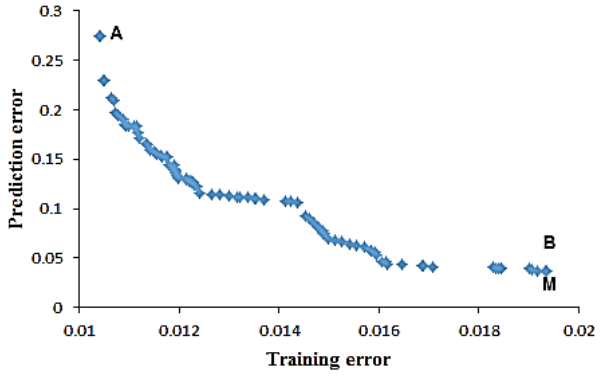

Fig. 6. Pareto points of training and testing errors of GMDH model for residual turbidity.

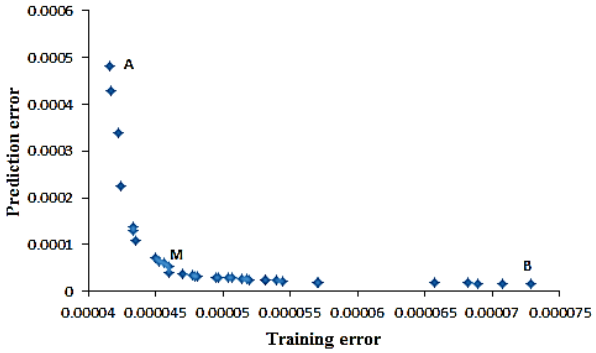

Fig. 7. Pareto points of training and testing errors of GMDH model for residual aluminium.

Table 4
Train and test error of Pareto curve for residual turbidity

Design points	Training error	Testing error
A	0.01041	0.27523
B	0.019359	0.036913
M	0.019359	0.036913

Table 5
Train and test error of Pareto curve for residual aluminium

Design points	Training error	Testing error
A	0.0000416411	0.0004807126
B	0.0000728757	0.0000158282
M	0.0000459905	0.0000390073

Table 6
Premier chromosomes to predict the amount of residual aluminium

The best of chromosomes	Training error	Testing error
11453436163526331535243 616343656122555562335344613461436233634551223245133513435563561356145614344655222226562534354626443345265	0.0000459905	0.00000390073

Table 7
Premier chromosomes to predict the amount of residual turbidity

The best of chromosomes	Training error	Testing error
1135163515451545141424461646243614662526236265623344455264435461224262635363566141622341424164613331423144523451423232515362556	0.019359	0.036913

allocated to one input variable, which are listed as: (1) polyelectrolyte, (2) pH, (3) turbidity, (4) polyaluminium chloride, (5) temperature, (6) electrical conductivity.

The proposed structure of GMDH neural network which applied to predict residual turbidity and aluminium is revealed in Figs. 8 and 9.

In this modeling, for better network training, laboratory data were divided into two categories (70% for training and 30% for testing). After obtaining the proposed model, results of the model were compared with experimental results. As can be seen in Figs. 10 and 11, the results of the model are in good agreement with the obtained results in operation unit of water treatment plant.

To assess accuracy of the model and error calculation, mean squared error (MSE; Eq. (10)), mean absolute RMSE (Eq. (11)) and coefficient of determination (R^2; Eq. (12)) relations were used. Where, N is the number of samples used for modeling, $Y_{(i,\text{exp})}$ is the experimental value, and $Y_{(i,\text{model})}$ is the networks' predicted value. Table 8 indicate the reliability of model by presenting the coefficient of determination values (R^2), mean square error (MSE), normal root mean square error (NRMSE).

$$\text{MSE} = \frac{1}{N}\sum_{i}^{N}(Y_{(i,\text{exp})} - Y_{(i,\text{model})})^2 \tag{10}$$

$$\text{RMSE} = \left[\frac{\sum_{i}^{N}(Y_{i,\text{exp}} - Y_{i,\text{model}})^2}{N}\right]^{0.5} \tag{11}$$

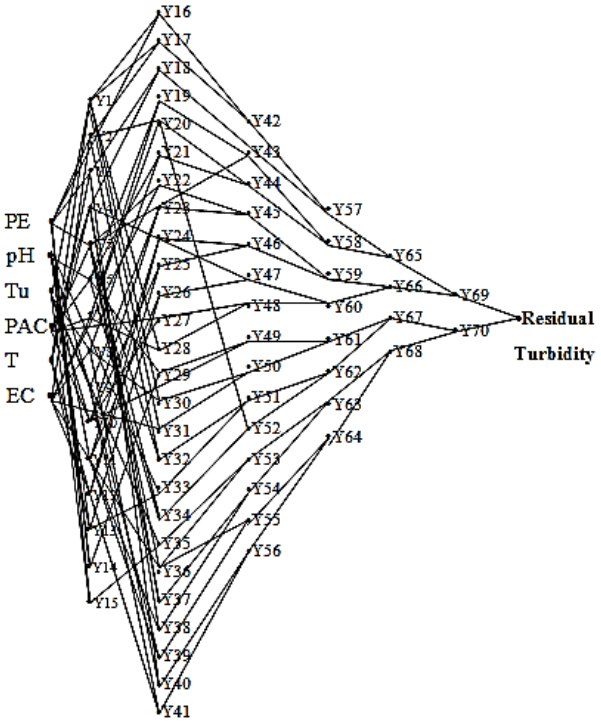

Fig. 8. Structure of the proposed GMDH neural network for residual turbidity.

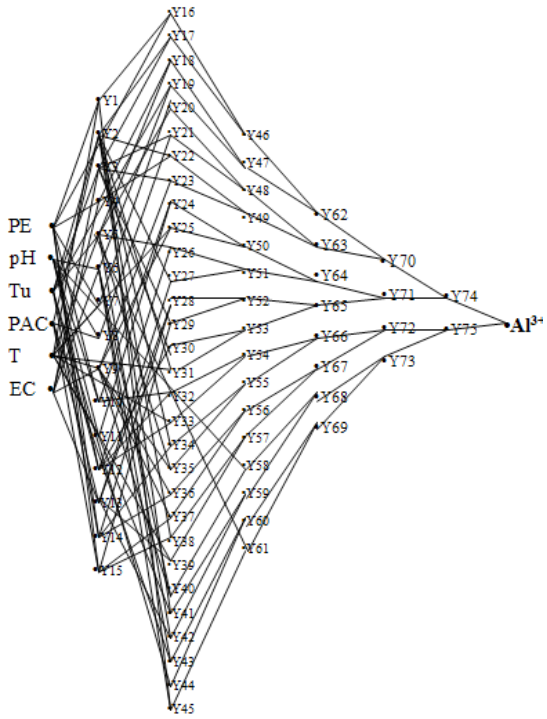

Fig. 9. Structure of the proposed GMDH neural network for residual aluminium.

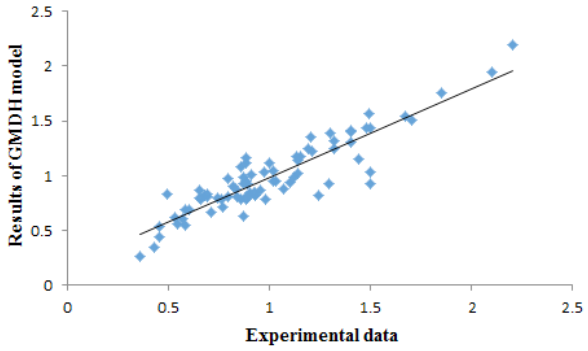

Fig. 10. Real output vs. modeled output for residual turbidity.

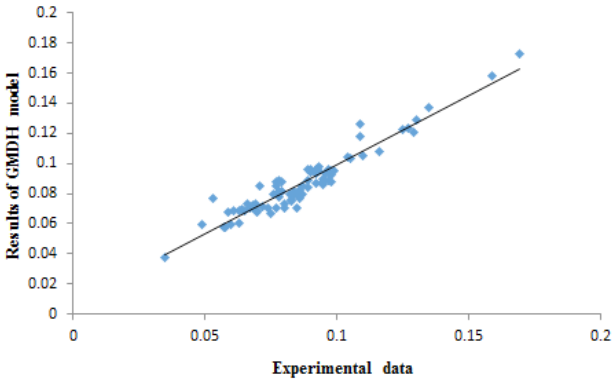

Fig. 11. Real output vs. modeled output for residual aluminium.

Table 8
GMDH results for training, testing and total data

	Residual turbidity			Residual aluminium		
	Total data	Training set	Testing set	Total data	Training set	Testing set
MSE	0.0246	0.0193	0.0369	0.00000438	0.00000425	0.00000469
NRMSE	0.0852	0.0755	0.01043	0.04944	0.04869	0.05115
R^2	0.8219	0.8488	0.7782	0.9138	0.9209	0.8953

$$R^2 = 1 - \left[\frac{\sum_{i=1}^{N} (Y_{i,\exp} - Y_{i,\text{model}})^2}{\sum_{i=1}^{N} (Y_{i,\exp})^2} \right] \quad (12)$$

Prediction of residual turbidity and aluminium using the GMDH, parameters values were obtained by non-linear Volterra equation (Appendices 1 and 2).

5. Discussion

Artificial intelligence (AI) provides an effective tool for many complicated engineering problems in various fields. As a result, in this study, AI-based techniques of GMDH were employed for prediction on residual turbidity and residual aluminium. The models were presented using GMDH network, which indicated the relation between pH, temperature, initial turbidity, electrolytic conductivity and also the amount of chemical injection, coagulant (PAC) and coagulant assistant (polyelectrolyte) with residual turbidity and residual aluminium.

According to premier chromosomes (Tables 6 and 7), it can be concluded that amount of polyelectrolyte injection, pH, initial turbidity, amount of polyaluminium chloride injection, temperature, electrical conductivity have effect on residual turbidity (14.84%, 16.40%, 14.84%, 20.31%, 15.62%, 17.96%) and also residual aluminium (10.15%, 13.28%, 23.43%, 15.62%, 21.09%, 16.40%), respectively. The results of modeling were compared with experimental data that demonstrates good data compliance, with the coefficient of determination (R^2) in (GMDH)-type network of 0.8239 and 0.9138 for residual turbidity and for residual aluminium, respectively. Also, error analysis results showed good performance of the proposed models, with the mean square error (MSE) values of 0.0248 for residual turbidity and 0.00000438 for residual aluminium.

6. Conclusions

As drinking water quality guidelines continue to become more stringent, modeling methods which utilize process histories will offer valuable tools for process modeling and control in water treatment plants and provide an alternative to conventional methodologies. Moreover, these modeling techniques allow such utilities to increase their process knowledge and, therefore, facilitate process control. These results provide further support for the wider use of the data-driven selection of variables and modeling in water treatment processes and warrant further investigation.

In this paper, genetic algorithms (GAs) are deployed for multi-objective Pareto optimal design of GMDH-type neural networks that have been used for modeling of a non-linear system. In this way, GAs with a specific encoding scheme is first presented to evolutionary design of the generalized GMDH-type neural networks in which the connectivity configurations in such networks are not limited to adjacent layers. Multi-objective GAs with a new diversity preserving mechanism are second used for Pareto optimization of such GMDH-type neural networks. The important conflicting objectives of GMDH-type neural networks that are considered in this work are, namely, TE, PE and number of neurons (N) of such neural networks. Moreover, an important trade-off can be discovered by such Pareto optimum approach to the design of GMDH-type neural networks which helps a designer to select a network compromisingly.

References

[1] World Health Organization, Guidelines for Drinking-Water Quality, Geneva, 2006.
[2] R.D. Letterman, Ed., Water Quality and Treatment: A Handbook of Community Water Supplies, American Water Works Association, Denver, CO, 1999.
[3] V. Rondeau, H. Jacqmin-Gadda, D. Commenges, C. Helmer, J. Dartigues, Aluminium and silica in drinking water and the risk of Alzheimer's disease or cognitive decline: findings from 15-year follow-up of the PAQUID cohort, Am. J. Epidemiol., 169 (2009) 489–496.
[4] A. Campell, The role of aluminium and copper on neuroinflammation and Alzheimer's disease, J. Alzheimers Dis., 10 (2006) 165–172.
[5] J.E. Van Benschoten, J.K. Edzwald, Chemical aspects of coagulation using aluminium salts - I. Hydrolytic reactions of alum and polyaluminium chloride, Water Res., 24 (1990) 1519–1526.
[6] J.E. Van Benschoten, J.K. Edzwald, Chemical aspects of coagulation using aluminium salts - II. Coagulation of fulvic acid using alum and polyaluminium chloride, Water Res., 24 (1990) 1527–1535.
[7] C. Huang, H. Shiu, Interactions between alum and organics in coagulation, Colloids Surf. A, 113 (1996) 155–163.
[8] M. Akbarizadeh, A. Daghbandan, M. Yaghoobi, Modeling and optimization of poly electrolyte dosage in water treatment process by GMDH type-NN and MOGA, IGCCE, 3 (2013) 94–106.
[9] S. Heddam, A. Bermad, N. Dechemi, ANFIS-based modeling for coagulant dosage in drinking water treatment plant: a case study, Environ. Monit. Assess., 184 (2012) 1953–1971.
[10] T.-W. Ha, K.-H. Choo, S.-J. Choi, Effect of chlorine on adsorption/ultrafiltration treatment for removing natural organic matter in drinking water, J. Colloid Interface Sci., 274 (2004) 587–593.
[11] A.G. Ivakhnenko, Group method of data handling-rival of method of stochastic approximation, Sov. Autom. Control, 13 (1968) 43–55.
[12] S. Ikeda, S. Fugishige, Y. Sawaragi, Nonlinear prediction model of river flow by self-organization method, Int. J. Syst. Sci., 7 (1976) 165–176.

[13] H. Tamura, T. Kondo, Heuristics free group method data handling algorithm of generating optimal partial polynomials with application to air pollution prediction, Int. J. Syst. Sci., 11 (1980) 1095–1111.
[14] T. Yoshimura, R. Kiyozumi, K. Nishino, T. Soeda, Prediction of air pollutant concentrations by revised GMDH algorithms in Tokushima Prefecture, IEEE Trans. Syst. Man Cybern. SMC, 12 (1982) 50–56.
[15] S.J. Farlow, Self-Organizing Methods in Modeling: GMDH-Type Algorithms, Marcel Dekker, New York, 1984.
[16] W.M. Lebow, R.K. Mehra, H. Rice, P.M. Tolgalagi, Forecasting Applications in Agricultural and Meteorological Time Series, In: S.J. Farrow, Ed., Self-organizing Methods in Modeling, GMDH Type Algorithms, Marcel Dekker, New York, 1984.
[17] A.G. Ivakhnenko, G.A. Ivakhnenko, J.A. Muller, Self-organization of the neural networks with active neurons, Pattern Recogn., 4 (1994) 177–188.
[18] T. Kondo, A.S. Pandya, J.M. Zurada, GMDH-type Neural Networks and Their Application to the Medical Image Recognition of the Lungs, Proc. 38th SICE Annual Conference, Vol. 3, 1999, pp. 1181–1186.
[19] F.J. Chang, Y.Y. Hwang, A self-organization algorithm for realtime flood forecast, Hydrol. Process, 13 (1999) 123–138.
[20] L. Sarycheva, Using GMDH in ecological and socio-economical monitoring problems, Syst. Anal. Model Simul. (SAMS), 43 (2003) 1409–1414.
[21] N. Pavel, S. Miroslav, Modeling of student's quality by means of GMDH algorithms, Syst. Anal. Model Simul. (SAMS), 43 (2003) 1415–1426.
[22] S.L. Hwang, G.F. Liang, J.T. Lin, Y.J. Yau, T.C. Yenn, C.C. Hsu, C.F. Chuang, A real-time warning model for teamwork performance and system safety in nuclear power plants, Saf. Sci., 47 (2009) 425–435.
[23] T.M. Tsai, P.H. Yen, T.J. Huang, Wave Height Forecasting Using Self-organization Algorithm Model, International Offshore and Polar Engineering Conference Osaka, Japan, 4 (2009) 806–812.
[24] M. Najafzadeh, Neurofuzzy-based GMDH-PSO to predict maximum scour depth at equilibrium at culvert outlets, J. Pipeline Syst. Eng. Pract., 7 (2015) 1–5.
[25] M. Najafzadeh, G.A. Barani, M.A. Hazi, GMDH to predict scour depth around a pier in cohesive soils, Appl. Ocean Res., 40 (2013) 35–41.
[26] M. Najafzadeh, G.A. Barani, M,R. Hessami Kermani, Group method of data handling to predict scour at downstream of a skijump bucket spillway, Earth Sci. Inform., 7 (2014) 231–248.
[27] M. Najafzadeh. G.A. Barani, M.R. Hessami-Kermani, Evaluation of GMDH networks for prediction of local scour depth at bridge abutments in coarse sediments with thinly armored beds, Ocean Eng., 104 (2015) 387–396.
[28] P.C. Verpoort, P. MacDonald, G.J. Conduit, Materials data validation and imputation with an artificial neural network, Comput. Mater. Sci., 147 (2018) 176–185.
[29] N. Maleki, S. Kashanian, E. Maleki, M. Nazari, A Novel Enzyme Based Biosensor for Catechol Detection in Water Samples Using Artificial Neural Network, Biochem. Eng. J., 128 (2015) 1–11.
[30] E. Maleki, O. Unal, K.R. Kashyzadeh, Fatigue behavior prediction and analysis of shot peened mild carbon steels, Int. J. Fatigue, 116 (2018) 48–67.
[31] E. Maleki, G.H. Farrahi, Modelling of conventional and severe shot peening influence on properties of high carbon steel via artificial neural network, Int. J. Eng. Sci., 31 (2018) 382–393.
[32] P. Chaves, T. Kojiri, Deriving reservoir operational strategies considering water quantity and quality objectives by stochastic fuzzy neural networks, Adv. Water Resour., 30 (2007) 1329–1341.
[33] P. Juntunen, M. Liukkonen, M. Pelo, M.J. Lehtola, Y. Hiltunen, Modelling of Water Quality: An Application to a Water Treatment Process, Appl. Comput. Intell. Soft Comput., 2 (2012) 1–9.
[34] A. Rak, Water Turbidity Modelling During Water Treatment Processes Using Artificial Neural Networks, Int. J. Water Sci., 29 (2013) 1–10.
[35] M.J. Kennedy, A.H. Gandomiab, C.M. Millera, Coagulation modeling using artificial neural networks to predict both turbidity and DOM-PARAFAC component removal, J. Environ. Chem. Eng., 3 (2015) 2829–2838.
[36] M. Fan, J. Hu, R. Cao, W. Ruan, X. Wei, A review on experimental design for pollutants removal in water treatment with the aid of artificial intelligence, Chemosphere, 200 (2018) 330–343.
[37] A. Jamali, A. Hajiloo, N. Nariman-zadeh, Reliability-based robust Pareto design of linear state feedback controllers using a multi-objective uniform-diversity genetic algorithm (MUGA), Expert Syst. Appl., 37 (2010) 401–413.
[38] L. Yang, H. Yang, H. Liu, GMDH-Based Semi-Supervised Feature Selection for Electricity Load Classification Forecasting, Sustainability, 10 (2018) 1–16.
[39] A.G. Ivakhnenko, J.A. Müller, Recent developments of self-organizing modeling in prediction and analysis of stock market, Microelectron. Reliab., 37 (1995) 1053–1072.
[40] A.G. Ivakhnenko, Polynomial theory of complex systems, IEEE Trans. Syst. Man Cybern., 4 (1971) 364–378.
[41] H. Iba, T. Sato, A numerical approach to genetic programming for system identification, Evolut. Comput., 3 (1995) 417–452.
[42] N. Nariman-Zadeh, A. Darvizeh, G.R. Ahmad-Zadeh, Hybrid genetic design of GMDH-type neural networks using singular value decomposition for modeling and prediction of the explosive cutting process, Proc. Inst. Mech. Eng. B: J. Eng. Manuf., 217 (2003) 779–790.
[43] N. Nariman Zadeh, A. Jamali, Pareto Genetic Design of GMDH-type Neural Networks for Nonlinear Systems, International Workshop on Inductive Modeling, Czech Technical University, Prague, Czech Republic, 2007, pp. 96–103.
[44] N. Nariman-Zadeh, A. Darvizeh, A. Jamali, A. Moeini, Evolutionary design of generalized polynomial neural networks for modeling and prediction of explosive forming process, J. Mater. Process. Technol., 164 (2005) 1561–1571.

Appendices

Appendix 1
Polynomial equation of the GMDH model for the prediction of residual turbidity

Y1=(-2.534249e+00)+(5.349917e-03*Tu)+(3.972091e-01*T)+(-4.958308e-05*(Tu)^2)+(-1.173815e-02*(T)^2)+(2.325866e-04*Tu*T);
Y2=(4.396963e+00)+(6.594475e+00*PE)+(-3.689741e-02*EC)+(-2.111863e+01*(PE)^2)+(8.346654e-05*(EC)^2)+(6.467581e-03*PE*EC);
Y3=(-3.171863e-01)+(-9.978223e+00*PE)+(1.715436e-01*T)+(-1.540328e+01*(PE)^2)+(-7.523384e-03*(T)^2)+(1.066683e+00*PE*T);
Y4=(-1.302972e+00)+(-2.347234e-02*PAC)+(2.467508e-01*T)+(-1.399330e-02*(PAC)^2)+(-8.069955e-03*(T)^2)+(1.353679e-02*PAC*T);
Y5=(1.403025e+00)+(-2.625819e+01*PE)+(5.287365e-01*PAC)+(-7.140982e+01*(PE)^2)+(-1.796670e-01*(PAC)^2)+(9.546558e+00*PE*PAC);
Y6=(3.021631e+01)+(-6.808825e+00*pH)+(-7.717788e-01*PAC)+(3.860603e-01*(pH)^2)+(-2.206140e-02*(PAC)^2)+(1.275657e01*pH*PAC);
Y7=(3.855509e+00)+(1.406572e-01*PAC)+(-3.061957e-02*EC)+(-1.996265e-02*(PAC)^2)+(6.734379e-05*(EC)^2)+(4.527770e-04*PAC*EC);
Y8=(3.858998e+00)+(8.082464e-03*Tu)+(-2.914066e-02*EC)+(-5.870398e-05*(Tu)^2)+(6.640365e-05*(EC)^2)+(8.225888e-06*Tu*EC);
Y9=(1.570817e+01)+(-5.886964e+00*pH)+(1.304018e+00*T)+(4.258881e-01*(pH)^2)+(-1.634969e-02*(T)^2)+(-8.960211e-02*pH*T);
Y10=(9.454767e+00)+(-6.431176e+00*pH)+(1.829721e-01*EC)+(7.174420e-01*(pH)^2)+(6.888317e-05*(EC)^2)+(-2.648328e-02*pH*EC);
Y11=(2.161046e+01)+(-4.810229e+00*pH)+(-2.040795e-02*Tu)+(2.752489e-01*(pH)^2)+(-6.079240e-05*(Tu)^2)+(3.802826e-03*pH*Tu);
Y12=(-4.172748e+00)+(6.226588e-01*T)+(3.451400e-05*EC)+(-8.963073e-03*(T)^2)+(5.668901e-05*(EC)^2)+(-1.497650e-03*T*EC);
Y13=(5.729705e-01)+(3.665203e-03*Tu)+(1.027654e-01*PAC)+(-1.264403e-04*(Tu)^2)+(5.054042e-03*(PAC)^2)+(9.753606e-04*Tu*PAC);
Y14=(1.738917e+01)+(-2.363417e+00*PE)+(-3.821225e+00*pH)+(-1.592304e+01*(PE)^2)+(2.135904e-01*(pH)^2)+(1.138403e+00*PE*pH);
Y15=(9.331941e-01)+(-6.419897e-01*PE)+(1.803878e-03*Tu)+(-5.374308e+01*(PE)^2)+(-2.882927e-04*(Tu)^2)+(2.781535e-01*PE*Tu);
Y16=(5.139762e-01)+(-2.080955e+00*PE)+(-9.858684e-02*Y1)+(9.006520e+00*(PE)^2)+(7.969901e-01*(Y1)^2)+(-1.251081e+00*PE*Y1);
Y17=(-3.018744e-01)+(3.131842e+00*Y2)+(-1.827276e+00*Y1)+(-2.676241e+00*(Y2)^2)+(-1.351629e-01*(Y1)^2)+(2.840377e+00*Y2*Y1);
Y18=(1.008218e+00)+(5.738668e+00*Y3)+(-6.779057e+00*Y4)+(-2.184193e+01*(Y3)^2)+(-1.397651e+01*(Y4)^2)+(3.683629e+01*Y3*Y4);
Y19=(3.600092e-02)+(1.309750e+00*Y6)+(-4.728705e-01*Y7)+(-8.149312e-01*(Y6)^2)+(7.742147e-02*(Y7)^2)+(8.657444e-01*Y6*Y7);
Y20=(-3.192379e-01)+(-1.966325e+00*Y2)+(3.768856e+00*Y7)+(2.321067e+00*(Y2)^2)+(4.421428e-01*(Y7)^2)+(-3.240565e+00*Y2*Y7);
Y21=(-1.220172e-01)+(2.192853e+00*Y6)+(-9.715260e-01*Y8)+(-8.083308e-02*(Y6)^2)+(1.389922e+00*(Y8)^2)+(-1.413312e+00*Y6*Y8);
Y22=(3.123993e+00)+(1.232929e+00*Y5)+(-3.036353e-02*EC)+(-5.062570e-01*(Y5)^2)+(6.427012e-05*(EC)^2)+(3.021726e-03*Y5*EC);
Y23=(8.532315e-01)+(7.236378e-01*Y9)+(-1.865073e+00*Y10)+(-2.032153e+00*(Y9)^2)+(-3.885186e-01*(Y10)^2)+(3.751105e+00*Y9*Y10);
Y24=(8.194861e-01)+(-6.309860e-01*Y11)+(-6.709227e-01*Y12)+(-3.253282e-01*(Y11)^2)+(-3.594373e-01*(Y12)^2)+(2.180767e+00*Y11*Y12);
Y25=(-1.872281e+00)+(1.955294e+00*Y10)+(2.354289e+00*Y12)+(1.505519e+00*(Y10)^2)+(1.404478e+00*(Y12)^2)+(-4.367517e+00*Y10*Y12);
Y26=(-2.740226e-01)+(1.129560e+00*Y11)+(4.130860e-01*Y13)+(6.234194e-01*(Y11)^2)+(1.193394e+00*(Y13)^2)+(-2.088544e+00*Y11*Y13);
Y27=(3.185215e-01)+(-6.712800e-02*Y10)+(1.276433e-01*PAC)+(3.539824e-01*(Y10)^2)+(-1.711278e-02*(PAC)^2)+(5.568889e-02*Y10*PAC);
Y28=(4.488382e-01)+(-1.108440e+00*Y1)+(1.102692e+00*Y7)+(-3.682313e+00*(Y1)^2)+(-4.723191e+00*(Y7)^2)+(9.009614e+00*Y1*Y7);
Y29=(9.630772e-01)+(-5.028623e+00*Y14)+(4.381553e+00*Y16)+(3.882025e+00*(Y14)^2)+(4.092315e-01*(Y16)^2)+(-3.629140e+00*Y14*Y16);
Y30=(4.445298e-01)+(-2.125319e+00*Y1)+(2.082777e+00*Y8)+(-2.603745e+00*(Y1)^2)+(-4.507046e+00*(Y8)^2)+(7.743488e+00*Y1*Y8);
Y31=(3.330239e+00)+(-9.219826e-02*Y1)+(-2.710335e-02*EC)+(4.188751e-01*(Y1)^2)+(5.990984e-05*(EC)^2)+(1.473905e-03*Y1*EC);
Y32=(-1.159549e+00)+(7.945073e-01*Y5)+(2.410819e+00*Y2)+(-9.978539e-01*(Y5)^2)+(-1.707158e+00*(Y2)^2)+(1.715868e+00*Y5*Y2);
Y33=(3.536761e+01)+(-7.705563e+00*pH)+(-6.399059e+00*Y13)+(4.085510e-01*(pH)^2)+(-2.374659e-01*(Y13)^2)+(9.707910e-01*pH*Y13);
Y34=(-5.121630e-02)+(6.335661e-01*Y5)+(4.097567e-01*Y6)+(-1.195428e+00*(Y5)^2)+(-1.105537e+00*(Y6)^2)+(2.322328e+00*Y5*Y6);
Y35=(-1.154447e+00)+(2.733729e+00*Y15)+(1.058104e-02*Tu)+(-4.329898e-01*(Y15)^2)+(-1.029354e-05*(Tu)^2)+(-1.205211e-02*Y15*Tu);
Y36=(-9.942429e-02)+(9.612879e-01*Y5)+(1.890964e-01*Y11)+(-8.540307e-01*(Y5)^2)+(-6.205909e-01*(Y11)^2)+(1.436177e+00*Y5*Y11);
Y37=(7.019803e-01)+(1.361868e+00*Y5)+(-1.835637e+00*Y4)+(-3.551945e+00*(Y5)^2)+(-1.892801e+00 *(Y4)^2)+(6.228997e+00*Y5*Y4);
Y38=(1.063678e-01)+(1.094391e+00*Y11)+(-4.711464e-01*Y4)+(-6.720486e-01*(Y11)^2)+(2.833633e-01*(Y4)^2)+(6.556707e-01*Y11*Y4);
Y39=(1.179561e+00)+(3.961439e-01*Y11)+(-2.056264e+00*Y9)+(-2.312680e+00*(Y11)^2)+(-1.102893e+00*(Y9)^2)+(4.955775e+00*Y11*Y9);

(*Continued*)

Appendix 1 (*Continued*)

Y40=(7.392900e-01)+(-1.515048e+00*Y3)+(9.608336e-01*Y8)+(-1.579244e+00*(Y3)^2)+(-2.622079e+00*(Y8)^2)+(5.035660e+00*Y3*Y8);
Y41=(2.639330e+00)+(-7.775909e+00*Y9)+(2.435152e+00*Y12)+(3.877738e+00*(Y9)^2)+(-9.913655e-01*(Y12)^2)+(7.657675e-01*Y9*Y12);
Y42=(2.779018e-01)+(3.220507e+00*Y16)+(-2.573860e+00*Y17)+(-1.225961e+01*(Y16)^2)+(-9.472484e+00*(Y17)^2)+(2.190167e+01*Y16*Y17);
Y43=(1.751429e-02)+(6.389429e-01*Y5)+(3.750539e-01*Y19)+(-4.050181e+00*(Y5)^2)+(-3.534269e+00*(Y19)^2)+(7.587393e+00*Y5*Y19);
Y44=(-4.952019e-03)+(-2.625304e+00*Y20)+(3.590192e+00*Y21)+(4.282066e+00*(Y20)^2)+(1.039065e+00*(Y21)^2)+(-5.288120e+00*Y20*Y21);
Y45=(-5.686924e-01)+(1.393357e+00*Y22)+(6.324175e-01*Y23)+(-2.127044e-01*(Y22)^2)+(2.907076e-01*(Y23)^2)+(-5.242944e-01*Y22*Y23);
Y46=(5.665858e-02)+(2.643864e+00*Y24)+(-1.888613e+00*Y25)+(-9.475337e-01*(Y24)^2)+(1.309524e+00*(Y25)^2)+(-1.892508e-01*Y24*Y25);
Y47=(1.731618e+00)+(2.357651e+00*Y26)+(-4.975441e+00*Y4)+(-1.027736e+01*(Y26)^2)+(-7.540594e+00*(Y4)^2)+(1.974631e+01*Y26*Y4);
Y48=(7.381382e-01)+(-4.099027e-01*Y27)+(-4.066926e-01*Y28)+(8.390097e-01*(Y27)^2)+(9.514665e-01*(Y28)^2)+(-7.697851e-01*Y27*Y28);
Y49=(2.322927e-01)+(8.453149e-01*Y29)+(-6.937675e-01*Y10)+(-2.776347e-01*(Y29)^2)+(5.017628e-01*(Y10)^2)+(3.826283e-01*Y29*Y10);
Y50=(2.150037e-01)+(-7.776868e-01*Y30)+(1.300145e+00*Y31)+(-6.256264e+00*(Y30)^2)+(-6.705575e+00*(Y31)^2)+(1.325905e+01*Y30*Y31);
Y51=(3.037004e-01)+(5.289203e-01*Y32)+(-3.144789e-01*Y33)+(5.400987e-01*(Y32)^2)+(7.692310e-01*(Y33)^2)+(-8.475286e-01*Y32*Y33);
Y52=(2.148329e-01)+(1.859845e+00*Y34)+(-1.288501e+00*Y20)+(-6.208213e+00*(Y34)^2)+(-4.115877e+00*(Y20)^2)+(1.056565e+01*Y34*Y20);
Y53=(-4.258709e-01)+(1.089225e+00*Y35)+(7.810581e-01*Y36)+(1.388191e-01*(Y35)^2)+(1.431551e-01*(Y36)^2)+(-7.131398e-01*Y35*Y36);
Y54=(-4.765462e-02)+(-9.960924e-01*Y37)+(2.096180e+00*Y38)+(9.175709e-02*(Y37)^2)+(-2.445221e+00*(Y38)^2)+(2.306137e+00*Y37*Y38);
Y55=(2.208790e-01)+(1.519750e+00*Y36)+(-1.143601e+00*Y39)+(-1.999338e+00*(Y36)^2)+(-6.114384e-01*(Y39)^2)+(3.032686e+00*Y36*Y39);
Y56=(-2.176225e-01)+(1.550833e+00*Y40)+(-3.664896e-01*Y41)+(-3.196257e-01*(Y40)^2)+(7.396902e-01*(Y41)^2)+(-3.954749e-01*Y40*Y41);
Y57=(-4.195049e-01)+(-1.290428e+00*Y42)+(3.146901e+00*Y18)+(2.188431e+00*(Y42)^2)+(-1.484755e+00*(Y18)^2)+(-1.161502e+00*Y42*Y18);
Y58=(1.040030e-01)+(3.630972e-01*Y43)+(3.989505e-01*Y44)+(-2.840879e+00*(Y43)^2)+(-2.519960e+00*(Y44)^2)+(5.507675e+00*Y43*Y44);
Y59=(2.834904e-01)+(1.000719e+00*Y45)+(-5.358068e-01*Y46)+(-3.337358e+00*(Y45)^2)+(-2.197664e+00*(Y46)^2)+(5.810000e+00*Y45*Y46);
Y60=(5.097583e-01)+(1.501204e+00*Y47)+(-1.678192e+00*Y48)+(5.357896e-01*(Y47)^2)+(1.913017e+00*(Y48)^2)+(-1.849728e+00*Y47*Y48);
Y61=(7.820931e-01)+(-1.299957e-01*Y49)+(-8.145208e-01*Y50)+(-6.867047e-01*(Y49)^2)+(-5.716942e-01*(Y50)^2)+(2.412836e+00*Y49*Y50);
Y62=(2.606641e-01)+(-1.239490e+00*Y51)+(1.601514e+00*Y52)+(4.182415e-01*(Y51)^2)+(-9.907003e-01*(Y52)^2)+(9.384172e-01*Y51*Y52);
Y63=(2.362268e-01)+(1.816781e+00*Y53)+(-1.272461e+00*Y54)+(-5.968370e+00*(Y53)^2)+(-4.532086e+00*(Y54)^2)+(1.078464e+01*Y53*Y54);
Y64=(-1.312742e-01)+(1.730468e+00*Y55)+(-5.129862e-01*Y56)+(-7.931690e-01*(Y55)^2)+(3.608799e-01*(Y56)^2)+(3.568982e-01*Y55*Y56);
Y65=(-5.471205e-01)+(-1.171959e+00*Y57)+(3.579326e+00*Y58)+(1.073377e+00*(Y57)^2)+(-1.961507e+00*(Y58)^2)+(3.088630e-02*Y57*Y58);
Y66=(-1.072116e-01)+(8.825674e-01*Y59)+(1.446823e-01*Y60)+(2.328192e+00*(Y59)^2)+(2.625686e+00*(Y60)^2)+(-4.931836e+00*Y59*Y60);
Y67=(-4.504153e-01)+(-4.010753e-01*Y61)+(2.518524e+00*Y62)+(-1.055767e-01*(Y61)^2)+(-2.085689e+00*(Y62)^2)+(1.555078e+00*Y61*Y62);

(*Continued*)

Appendix 1 (*Continued*)

Y68=(-2.845898e-01)+(1.637336e+00*Y63)+(-4.042879e-02*Y64)+(2.211240e-01*(Y63)^2)+(1.064879e+00*(Y64)^2)+(-1.604184e+00*Y63*Y64);
Y69=(-8.118940e-02)+(1.152379e+00*Y65)+(-4.807945e-02*Y66)+(5.285302e-01*(Y65)^2)+(1.047534e+00*(Y66)^2)+(-1.622464e+00*Y65*Y66);
Y70=(2.585155e-01)+(-1.492458e+00*Y67)+(1.941866e+00*Y68)+(2.897684e-01*(Y67)^2)+(-1.065277e+00*(Y68)^2)+(1.062598e+00*Y67*Y68);
Y71=(-1.346896e-01)+(-1.275198e+00*Y69)+(2.723254e+00*Y70)+(3.368519e-01*(Y69)^2)+(-2.510099e+00*(Y70)^2)+(1.886811e+00*Y69*Y70);

Appendix 2
Polynomial equation of the GMDH model for the prediction of residual aluminium

Y1=(-1.702106e-01)+(-7.626401e-03*PAC)+(2.772828e-02*T)+(8.764070e-05*(PAC)^2)+(-7.425876e-04*(T)^2)+(5.624936e-04*PAC*T);
Y2=(7.467993e-02)+(-1.193848e-04*Tu)+(2.578315e-03*PAC)+(1.165939e-05*(Tu)^2)+(4.991641e-03*(PAC)^2)+(-4.934242e-04*Tu*PAC);
Y3=(-1.397023e-01)+(1.296172e-03*Tu)+(2.017830e-03*EC)+(-2.076688e-06*(Tu)^2)+(-4.387130e-06*(EC)^2)+(-5.371943e-06*Tu*EC);
Y4=(-1.455833e-01)+(6.966782e-01*PE)+(1.954198e-03*EC)+(-5.603512e-01*(PE)^2)+(-4.062518e-06*(EC)^2)+(-2.740257e-03*PE*EC);
Y5=(-2.451797e-01)+(1.375095e-04*Tu)+(3.572820e-02*T)+(-1.591860e-07*(Tu)^2)+(-9.321519e-04*(T)^2)+(-4.779012e-06*Tu*T);
Y6=(1.161813e+00)+(-4.200895e-01*pH)+(4.844727e-03*EC)+(3.430198e-02*(pH)^2)+(-2.305933e-06*(EC)^2)+(-4.912461e-04*pH*EC);
Y7=(-2.051531e-01)+(-1.921306e-02*PE)+(3.066802e-02*T)+(-1.956042e-01*(PE)^2)+(-8.075459e-04*(T)^2)+(9.139996e-03*PE*T);
Y8=(2.155929e+00)+(-5.500166e-01*pH)+(-2.631814e-03*PAC)+(3.618751e-02*(pH)^2)+(-5.309768e-04*(PAC)^2)+(1.065628e-03*pH*PAC);
Y9=(-5.714772e-01)+(3.868021e-02*T)+(3.074136e-03*EC)+(-4.232457e-04*(T)^2)+(-3.589102e-06*(EC)^2)+(-9.949798e-05*T*EC);
Y10=(1.611915e+00)+(1.659953e+00*PE)+(-4.326280e-01*pH)+(-1.518830e-01*(PE)^2)+(3.009217e-02*(pH)^2)+(-1.962753e-01*PE*pH);
Y11=(2.842984e-01)+(-1.031613e-02*pH)+(-2.849727e-02*T)+(-7.071133e-03*(pH)^2)+(-1.038367e-03*(T)^2)+(8.291753e-03*pH*T);
Y12=(1.717629e+00)+(-4.421779e-01*pH)+(7.262944e-04*Tu)+(2.970533e-02*(pH)^2)+(-1.097958e-06*(Tu)^2)+(-7.185976e-05*pH*Tu);
Y13=(-1.555620e-01)+(2.452657e-02*PAC)+(1.999944e-03*EC)+(-7.878449e-04*(PAC)^2)+(-4.146337e-06*(EC)^2)+(-8.695351e-05*PAC*EC);
Y14=(5.920667e-02)+(6.427719e-01*PE)+(-7.099443e-04*Tu)+(-2.692234e+00*(PE)^2)+(-5.167198e-06*(Tu)^2)+(7.281772e-03*PE*Tu);
Y15=(1.369970e-01)+(-1.415961e+00*PE)+(1.503667e-02*PAC)+(-6.204435e+00*(PE)^2)+(-1.127340e-02*(PAC)^2)+(6.868821e-01*PE*PAC);
Y16=(2.185037e-01)+(8.541725e-01*PE)+(-5.344332e+00*Y1)+(1.079376e+00*(PE)^2)+(4.609771e+01*(Y1)^2)+(-1.374756e+01*PE*Y1);
Y17=(1.343818e-01)+(2.149488e+00*Y2)+(-5.477240e+00*Y3)+(-9.106743e+00*(Y2)^2)+(3.723016e+01*(Y3)^2)+(3.603377e+00*Y2*Y3);
Y18=(5.666695e-01)+(-3.615028e+00*Y4)+(-9.594257e+00*Y5)+(1.219410e+01*(Y4)^2)+(5.002180e+01*(Y5)^2)+(2.526634e+01*Y4*Y5);
Y19=(-6.701909e-02)+(1.721888e+00*Y6)+(2.004564e-03*Tu)+(3.269030e-01*(Y6)^2)+(-7.603734e-07*(Tu)^2)+(-2.199544e-02*Y6*Tu);
Y20=(3.307525e-01)+(2.667156e+01*Y7)+(-3.378600e+01*Y5)+(2.032380e+02*(Y7)^2)+(5.323017e+02*(Y5)^2)+(-6.868417e+02*Y7*Y5);
Y21=(4.871092e-01)+(-4.133341e+00*Y8)+(-6.847067e+00*Y3)+(8.296314e+00*(Y8)^2)+(2.665166e+01*(Y3)^2)+(3.799179e+01*Y8*Y3);
Y22=(-1.642597e+00)+(9.798463e+00*Y4)+(2.631407e+01*Y2)+(4.235934e+01*(Y4)^2)+(-5.411164e+01*(Y2)^2)+(-1.748619e+02*Y4*Y2);
Y23=(3.384227e-01)+(-4.570937e+00*Y3)+(-2.882316e+00*Y9)+(1.694423e+01*(Y3)^2)+(8.619815e+00*(Y9)^2)+(2.635116e+01*Y3*Y9);
Y24=(1.316068e-01)+(3.989315e+00*Y10)+(-6.101945e+00*Y11)+(-1.634729e+01*(Y10)^2)+(5.268875e+01*(Y11)^2)+(-1.864935e+01*Y10*Y11);
Y25=(2.301653e-01)+(-2.150307e-02*T)+(-3.238182e-01*Y9)+(-2.914314e-04*(T)^2)+(-3.595513e+01*(Y9)^2)+(4.109595e-01*T*Y9);
Y26=(6.278603e-01)+(-3.921739e+00*Y12)+(-1.013051e+01*Y5)+(5.442693e+00*(Y12)^2)+(4.847550e+01*(Y5)^2)+(3.506918e+01*Y12*Y5);
Y27=(-2.694828e-01)+(7.106106e+00*Y2)+(-1.014597e+00*Y13)+(-2.443348e+01*(Y2)^2)+(2.296014e+01*(Y13)^2)+(-2.146924e+01*Y2*Y13);
Y28=(4.427188e-01)+(-4.255171e+00*Y14)+(-6.262085e+00*Y13)+(1.775463e+01*(Y14)^2)+(3.067336e+01*(Y13)^2)+(2.511680e+01*Y14*Y13);
Y29=(1.002296e+00)+(-1.107236e+01*Y15)+(-1.297106e+01*Y3)+(3.722526e+01*(Y15)^2)+(5.051247e+01*(Y3)^2)+(6.705163e+01*Y15*Y3);
Y30=(6.648106e-01)+(-6.346648e+00*Y12)+(-8.811894e+00*Y3)+(1.263621e+01*(Y12)^2)+(3.083724e+01*(Y3)^2)+(5.390646e+01*Y12*Y3);
Y31=(3.377283e-01)+(-5.172374e+00*Y2)+(-1.376425e-02*T)+(1.641678e+00*(Y2)^2)+(-3.919779e-04*(T)^2)+(3.642431e-01*Y2*T);
Y32=(1.403684e-02)+(7.760952e+00*Y10)+(-7.101946e+00*Y12)+(-4.986240e+01*(Y10)^2)+(4.013851e+01*(Y12)^2)+(1.176993e+01*Y10*Y12);
Y33=(6.463184e-02)+(-1.700262e+00*Y12)+(6.290647e-03*T)+(-2.411700e+01*(Y12)^2)+(-1.060072e-03*(T)^2)+(3.774901e-01*Y12*T);

(*Continued*)

Appendix 2 (*Continued*)

Y34=(5.914215e-02)+(-4.658054e+00*Y14)+(4.104756e+00*Y2)+(-1.802256e+01*(Y14)^2)+(-5.766816e+01*(Y2)^2)+(8.582254e+01*Y14*Y2);
Y35=(3.015854e-01)+(-6.287214e+00*Y5)+(-1.094381e-01*Y9)+(2.465462e+01*(Y5)^2)+(-1.014813e+01*(Y9)^2)+(3.000888e+01*Y5*Y9);
Y36=(7.087682e-01)+(-9.953921e+00*Y14)+(-6.763801e+00*Y9)+(3.540559e+01*(Y14)^2)+(1.564900e+01*(Y9)^2)+(5.831482e+01*Y14*Y9);
Y37=(5.621038e-01)+(-8.019685e+00*Y15)+(-4.971425e+00*Y9)+(3.251684e+01*(Y15)^2)+(1.501012e+01*(Y9)^2)+(3.836387e+01*Y15*Y9);
Y38=(-4.955778e-02)+(9.138187e-01*Y15)+(9.090873e-01*Y2)+(-3.445020e+01*(Y15)^2)+(-3.218532e+01*(Y2)^2)+(6.391056e+01*Y15*Y2);
Y39=(4.554947e-01)+(-1.016541e+01*Y13)+(-7.565611e-04*T)+(2.610914e+01*(Y13)^2)+(-9.610654e-04*(T)^2)+(4.141756e-01*Y13*T);
Y40=(3.191118e-01)+(-2.158306e+00*Y6)+(-4.440265e+00*Y9)+(-8.888179e+00*(Y6)^2)+(7.455868e+00*(Y9)^2)+(4.615532e+01*Y6*Y9);
Y41=(2.282294e-01)+(-6.446318e+00*Y11)+(9.636250e-01*Y2)+(3.096603e+01*(Y11)^2)+(-1.127202e+01*(Y2)^2)+(2.408074e+01*Y11*Y2);
Y42=(4.548860e-01)+(-8.760339e+00*Y5)+(-1.675484e+00*Y13)+(4.942921e+01*(Y5)^2)+(4.385355e+00*(Y13)^2)+(1.659361e+01*Y5*Y13);
Y43=(-5.305011e-02)+(8.943502e-01*Y6)+(3.604592e-02*PAC)+(6.974033e+00*(Y6)^2)+(-3.822623e-04*(PAC)^2)+(-3.661066e-01*Y6*PAC);
Y44=(-7.829315e-02)+(1.825786e-03*Tu)+(1.895250e+00*Y1)+(-1.326595e-06*(Tu)^2)+(3.073980e-01*(Y1)^2)+(-2.041109e-02*Tu*Y1);
Y45=(0.3408869)+(-3.046889*Y6)+(-4.256494*Y1)+(12.05424*(Y6)^2)+(21.89969*(Y1)^2)+(15.81724*Y6*Y1);
Y46=(1.420851e-01)+(-3.168047e+00*Y16)+(2.831768e-01*Y17)+(8.857065e+00*(Y16)^2)+(-1.166653e+01*(Y17)^2)+(2.847227e+01*Y16*Y17);
Y47=(6.566266e-01)+(-1.139555e+01*Y18)+(-3.256348e+00*Y19)+(1.814263e+01*(Y18)^2)+(-2.795880e+01*(Y19)^2)+(1.020880e+02*Y18*Y19);
Y48=(2.571339e-01)+(-4.127256e+00*Y20)+(-1.109898e+00*Y21)+(5.607210e+00*(Y20)^2)+(-1.497806e+01*(Y21)^2)+(4.668779e+01*Y20*Y21);
Y49=(1.317087e-01)+(-1.319946e+00*Y22)+(-1.380712e+00*Y23)+(8.753604e+00*(Y22)^2)+(8.890915e+00*(Y23)^2)+(7.017114e+00*Y22*Y23);
Y50=(1.617116e-01)+(-3.272548e+00*Y24)+(5.286486e-01*Y25)+(1.045688e+01*(Y24)^2)+(-7.136798e+00*(Y25)^2)+(1.800095e+01*Y24*Y25);
Y51=(3.288396e-01)+(-4.553038e+00*Y26)+(-2.799668e+00*Y27)+(1.792298e+01*(Y26)^2)+(8.431920e+00*(Y27)^2)+(2.556933e+01*Y26*Y27);
Y52=(-2.971049e-01)+(5.748685e+00*Y28)+(2.115851e+00*Y29)+(-8.087998e+01*(Y28)^2)+(-5.916694e+01*(Y29)^2)+(1.010739e+02*Y28*Y29);
Y53=(5.565498e-02)+(-8.427367e-01*Y30)+(3.007034e-01*Y31)+(1.690756e+01*(Y30)^2)+(1.320418e+01*(Y31)^2)+(-2.017835e+01*Y30*Y31);
Y54=(1.515383e-01)+(-1.972281e+00*Y32)+(-4.184907e-01*Y33)+(3.285382e+01*(Y32)^2)+(3.407253e+01*(Y33)^2)+(-4.855808e+01*Y32*Y33);
Y55=(5.812883e-02)+(8.946222e-01*Y34)+(-2.183154e+00*Y35)+(-1.039693e+01*(Y34)^2)+(7.315430e+00*(Y35)^2)+(2.151615e+01*Y34*Y35);
Y56=(-8.122167e-02)+(-6.356687e+00*Y36)+(9.218034e+00*Y37)+(7.532590e+00*(Y36)^2)+(-7.726045e+01*(Y37)^2)+(5.936385e+01*Y36*Y37);
Y57=(3.154600e-01)+(-3.461287e+00*Y38)+(-3.902124e+00*Y39)+(1.296978e+01*(Y38)^2)+(1.512429e+01*(Y39)^2)+(2.585887e+01*Y38*Y39);
Y58=(-2.966618e-02)+(5.170626e-02*pH)+(-3.114333e+00*Y40)+(-5.857013e-03*(pH)^2)+(9.275808e-02*(Y40)^2)+(4.993455e-01*pH*Y40);
Y59=(3.114790e-01)+(1.211976e+00*Y41)+(-7.709163e+00*Y42)+(-1.586467e+01*(Y41)^2)+(3.192999e+01*(Y42)^2)+(2.820228e+01*Y41*Y42);
Y60=(3.627212e-01)+(-3.693760e-01*Y43)+(-7.894409e+00*Y44)+(-2.790019e+01*(Y43)^2)+(1.769592e+01*(Y44)^2)+(6.816923e+01*Y43*Y44);
Y61=(5.091826e-01)+(-6.287808e+00*Y45)+(-4.859507e+00*Y5)+(-9.167714e+00*(Y45)^2)+(-1.183473e+01*(Y5)^2)+(9.228548e+01*Y45*Y5);
Y62=(1.456590e-01)+(8.676825e-01*Y46)+(-3.079996e+00*Y47)+(1.629688e+01*(Y46)^2)+(3.620852e+01*(Y47)^2)+(-3.574769e+01*Y46*Y47);
Y63=(1.583390e-01)+(-4.127409e+00*Y48)+(1.415799e+00*Y49)+(4.121019e+00*(Y48)^2)+(-2.534568e+01*(Y49)^2)+(4.278700e+01*Y48*Y49);
Y64=(2.259574e-01)+(-4.008972e+00*Y50)+(-2.788540e-01*Y51)+(-3.209335e+01*(Y50)^2)+(-5.416222e+01*(Y51)^2)+(1.171980e+02*Y50*Y51);
Y65=(3.095170e-01)+(-4.763668e+00*Y52)+(-2.143051e+00*Y53)+(5.176012e+00*(Y52)^2)+(-9.751222e+00*(Y53)^2)+(5.418689e+01*Y52*Y53);

(*Continued*)

Appendix 2 (*Continued*)

Y66=(1.833375e-01)+(-3.310632e+00*Y54)+(-1.155082e-01*Y55)+(9.120995e+00*(Y54)^2)+(-6.254389e+00*(Y55)^2)+(2.340187e+01*Y54*Y55);

Y67=(-1.549757e-02)+(-1.197713e+00*Y56)+(2.480742e+00*Y57)+(-2.672636e+01*(Y56)^2)+(-5.016641e+01*(Y57)^2)+(7.606205e+01*Y56*Y57);

Y68=(7.209168e-02)+(5.042132e-01*Y58)+(-1.256884e+00*Y59)+(-1.640672e+01*(Y58)^2)+(-5.291244e+00*(Y59)^2)+(3.222831e+01*Y58*Y59);

Y69=(1.362146e-01)+(-1.132576e+00*Y60)+(-7.613028e-01*Y61)+(4.367624e+01*(Y60)^2)+(3.448826e+01*(Y61)^2)+(-6.382457e+01*Y60*Y61);

Y70=(-1.362799e-01)+(-6.202449e-02*Y62)+(4.127659e+00*Y63)+(-4.540964e+01*(Y62)^2)+(-7.890652e+01*(Y63)^2)+(1.082866e+02*Y62*Y63);

Y71=(-5.806236e-02)+(-6.562674e-01*Y64)+(2.849079e+00*Y65)+(-1.918779e+01*(Y64)^2)+(-3.785218e+01*(Y65)^2)+(5.169410e+01*Y64*Y65);

Y72=(2.197707e-02)+(-9.656269e-01*Y66)+(1.429030e+00*Y67)+(2.132677e+01*(Y66)^2)+(8.728264e+00*(Y67)^2)+(-2.710257e+01*Y66*Y67);

Y73=(3.444996e-02)+(1.319262e+00*Y68)+(-1.079021e+00*Y69)+(-6.577506e+00*(Y68)^2)+(5.984380e+00*(Y69)^2)+(4.608135e+00*Y68*Y69);

Y74=(2.880070e-02)+(-3.854216e-01*Y70)+(8.145178e-01*Y71)+(-2.343069e+01*(Y70)^2)+(-3.134929e+01*(Y71)^2)+(5.787011e+01*Y70*Y71);

Y75=(-1.006981e-02)+(1.450968e+00*Y72)+(-5.741348e-02*Y73)+(7.086416e+01*(Y72)^2)+(7.861047e+01*(Y73)^2)+(-1.534599e+02*Y72*Y73);

Y76=(-1.870998e-02)+(2.710722e+00*Y74)+(-1.340864e+00*Y75)+(1.174450e+01*(Y74)^2)+(3.187583e+01*(Y75)^2)+(-4.573037e+01*Y74*Y75);