Artificial neural networks and genetic algorithms: An efficient modelling and optimization methodology for active chlorine production using the electrolysis process

Majid Gholami Shirkoohi,[1,3] Rajeshwar Tyagi,[1,3] Peter A. Vanrolleghem,[2,3] Patrick Drogui[1,3]

[1]Institut National de la Recherche Scientifique (INRS), Centre-Eau Terre Environnement, Université du Québec, Québec (QC), Canada

[2]modelEAU, Département de génie civil et de génie des eaux, Université Laval, Québec (QC), Canada

[3]CentrEau, Centre de recherche sur l'eau, Université Laval, Québec (QC), Canada

Correspondence

Patrick Drogui, Institut National de la Recherche Scientifique (INRS), Centre-Eau Terre Environnement, Université du Québec, 490, Rue de la Couronne, Québec (QC) G1K 9A9, Canada.

Email: patrick.drogui@ete.inrs.ca

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/cjce.24036

Abstract

This study evaluates the effectiveness of a modelling and optimization methodology based on artificial neural networks and genetic algorithms in the prediction of the behaviour of an electrolysis process of active chlorine production from a synthetic saline effluent. Multilayer perceptrons feedforward neural networks were developed for the active chlorine production and energy consumption based on the following inputs: electrolysis time, current intensity, hydrochloric acid concentration, and chloride ion concentration. In order to diagnose and prevent the over-fitting problem during the learning process, learning curves and the regularization factor were utilized. The trained ANN models were able to successfully predict the active chlorine production and energy consumption of the process (R^2=0.979 and MSE=3.826 for active chlorine production and R^2=0.985 and MSE=6.952 for energy consumption). Multi-objective optimization for maximizing active chlorine production and minimizing energy consumption was carried out by a genetic algorithm using the best derived ANN models. The Pareto front obtained led to multiple non-dominated optimal points, which result in insights regarding the optimal operating conditions for the process.

KEYWORDS

ANN-GA, electrochemical processes, learning curves, multi-objective optimization, response surface methodology

1 INTRODUCTION

In recent years, electrochemical processes have been gaining attention as an alternative method for water and wastewater treatment. These processes are considered as eco-friendly and green technologies since the leading reagent involved, the electron, is considered a clean reagent and takes advantage of coupling chemistry (in situ generation of oxidant) with electronic science (electron transfer). Other attractive advantages include versatility, high energy efficiency, amenability to automation, and cost-effectiveness.[1,2] Several studies focusing on the use of electrolysis with different electrochemical methods such as electrooxidation, electrocoagulation, electroflotation, electro-Fenton reaction, and electrodialysis have been published over the last decade for improving the treatment performance of wastewaters and drinking waters.[3-6]

In literature, phenomenological and empirical modelling approaches are generally used for mathematical modelling of electrochemical water and wastewater treatment processes. Although phenomenological (white-box) modelling provides valuable insights into the behaviour of the process and has the ability of extrapolation, heat and mass transport phenomena along with detailed knowledge of the reaction kinetics are required. First principles related to the underlying science and engineering laws lead to governing equations that ultimately arrange these models.[7] In empirical modelling, the structure of the data-fitting model should be specified a priori which makes it challenging as one needs to choose a suitable model structure among the many available ones, especially for non-linear processes.[8] Electrochemical processes for water and wastewater treatment are generally complicated non-linear systems and dependent on many factors such as the influent concentration of contaminants,[9] the applied current density and electrical potential,[10] the types of electrodes,[11] the electrolyte type and concentration,[12] and chemical interactions between contaminants.[13,14] It is thus difficult to use phenomenological or empirical models to model, simulate, and optimize the processes.

In this context, artificial intelligence methods such as artificial neural networks (ANNs) along with genetic algorithms (GAs) have emerged as attractive alternative approaches for modelling and optimization of these non-linear processes in case phenomenological or conventional regression models are not practical.[15] These black-box (data-driven) models are based on empirical data and relationships among input and output variables of the process. Artificial intelligence methods, such as ANNs, have the role of discovering relationships in which patterns

of input data can be linked to the associated output data. These data-driven tools model the system behaviour solely from mapping the input-output data rather than from process knowledge. As the complexity of engineering problems increases, the development of faster computers along with more advanced computational algorithms and availability of cost-efficient sensors results in a noticeable paradigm shift from white-box to black-box modelling.[7,16,17] Various types of problems in science can be cast in the form of such pattern-matching, and among the methods within machine learning tools, ANNs are one of the most effective methods.[18-20] Some recent publications illustrate successful application of ANN models in various electrochemical processes.[21-25]

GAs belong to the category of evolutionary algorithms that are used for the optimization of objective (fitness) functions by means of parameter space coding. Through the algorithm, a GA can obtain acceptable results by using three stochastic operators: selection, crossover, and mutation.[26] Detailed information about the theory of GAs and the combination of ANNs with GAs can be found in the literature.[27,28]

The present study is focused on investigation and analyses of the effectiveness of AI methods for modelling and optimization of an electrolysis process. The database used in this paper was taken from a series of experiments for active chlorine production from a synthetic saline effluent by electrolysis, from the authors' lab and previously published in the literature.[29] Chlorine is one of the most commonly synthetically produced chemicals worldwide and, due to its oxidizing power, has been used as a disinfectant for potable water, wastewater, and swimming pools.[30] Other uses of chlorine products by electrolysis have been reported for the treatment of dye-containing effluents[31] and as electrolyzed oxidizing water in the food industry.[32] It is worth mentioning that studies are still being conducted about the production of undesirable active chlorine species during electrolysis.[33]

While the production of chlorine is commercially dependent on the electrolysis of highly concentrated solutions of sodium chloride (NaCl),[34,35] alternative approaches are being introduced. These approaches include seawater[36] and deep ocean water electrolysis.[37,38] In addition, desalination plants produce brine effluents, which are highly concentrated in salts. One of the techniques for managing this saline concentrate can be to use it as a saline resource for

chlorine production. This would lead to a reduction in chemical costs for the process of chlorine production.[39,40]

The ANN-GA approach for modelling and optimization of electrochemical processes has been applied before. Picos and Peralta-Hernández[41] utilized ANN models to predict the behaviour of an electro-oxidation pilot press-type reactor, which treats synthetic wastewater prepared with a synthetic Violet 54-B dye. Single-objective GA optimization was linked to their ANN model to find the best operational conditions for discoloration efficiency. Tuning ANN models, falling into the domain of hyper-parameter (e.g., number of hidden neurons) optimization, is a crucial task to obtain neural networks with the best performance possible and a strong ability of generalization. In this regard, usually, a trial-and-error procedure is used to derive the best configured network.[42]

In this work, modelling and optimization of active chlorine production by combining ANNs and GAs will be studied. This method includes feedforward neural networks and considers the impacts of learning curves and the regularization factor to improve the training process. It is followed by a multi-objective GA for the optimization process regarding active chlorine production and energy consumption. Learning curves help to acquire an insight throughout the modelling problem in order to diagnose the problem as high-variance or high-bias, which can then help to optimally select the most suitable configuration of the network. Regularization is also utilized to prevent over-fitting, which can occur with too complex a model. These techniques can give an insight into the ANN modelling process and can be used instead of or along with a trial-and-error procedure during training of neural networks. To the best of our knowledge, learning curves and the impact of the regularization factor in the cost function of ANNs have not been studied before for modelling of electrochemical processes. Further, Pareto optimal solutions obtained by multi-objective optimization using a GA can help to identify optimal operating conditions regarding the production of active chlorine and energy consumption of the process.

2 EXPERIMENTAL PROCEDURE

The database used in our work was acquired from the experiments of a published study of our group entitled “Statistical optimization of active chlorine production from a synthetic saline effluent by electrolysis”.[29] To prepare the synthetic saline effluent (SSE) used in these

experiments, sodium chloride (NaCl, Fisher Scientific, ACS reagent) was added to distilled water to produce solutions at different concentrations from 0.05 mol/L-0.105 mol/L. It was observed that produced chlorine gas could be converted to hypochlorous acid (HClO) and hypochlorite ion (ClO^-) after a value of pH = 2.0. Therefore, the initial pH of the solution was adjusted by hydrochloric acid (from 0.02 mol/L-0.14 mol/L) in the range of 0.9-1.3. A batch electrolytic cell was designed for conducting the assays using a power supply, an air diffuser, a 4 L glass tank, and a peristaltic pump. An expanded metal Ti/IrO_2 anode and a stainless steel cathode in the form of plates were utilized as electrodes. A 400 mL (135 mm ×35 mm ×140 mm) PVC electrolytic reactor was used to carry out the experiments.

The Wessler reaction was used to estimate the hypochlorous acid production which is based on the oxidation of iodide ions (I^-) to iodine (I_2) in the presence of active chlorine. Then tri-iodide (I_3^-) can be formed by the reaction of surplus iodide ions with iodine.[43] A Carry UV 50 spectrophotometer (Varian, Canada) was used to analyze the tri-iodide ion by measuring the absorbance at 353 nm. The electrical intensity and voltage were applied via an Enduro 250 V power supply.

Response surface methodology (RSM) approach was utilized to design the experimental assays using a factorial design (FD) followed by a central composite design (CCD). The experiments consisted of 16 experiments for FD and an extra 14 experiments for CCD, a total of 30 experiments. Table 1 represents the experimental region for gas chlorine production.

3 PROCESS MODELLING AND OPTIMIZATION

3.1 ANNs modelling

As the name implies, ANNs, commonly referred to as neural networks, imitate the essential characteristics of the human brain (which itself is a highly non-linear, complex, and parallel computer), such as self-adaptability, self-organization, and error tolerance.[44,45] ANNs can explore many competing hypotheses simultaneously using a massively parallel network composed of non-linear computational elements (neurons or nodes) that are interconnected by links with variable weights. The mentioned interconnected set of weights contains the knowledge generated by the network.[46] Each neuron at certain times examines its inputs and computes an output called an activation. The new activation is then passed along those connections to other neurons.

One of the most common architectures of ANNs, considering how the different neurons are positioned and connected to each other as well as the composition of layers, is the multilayer percepterons (MLP) feedforward network. These networks are usually applied to diverse problems, including function approximation, pattern classification, system identification, process control, process optimization, and so on.[47,48] The nature and complexity of the problem in addition to the desired accuracy and the available data determine the number of hidden layers and the corresponding number of neurons in each hidden layer. In addition, the configuration of the MLP network including the number of hidden layers and hidden neurons, can be derived by a trial-and-error procedure.[49]

The standard learning algorithm for MLP neural networks for any pattern recognition or function fitting process is known as the back-propagation (BP) algorithm.[48] The BP algorithm can be viewed as a generalization of the least mean square procedure that can be used for the training of multilayer neural networks. In the BP algorithm, data enter the network via the input layer which merely transfers the data value to the hidden layer over weighted connections. The hidden and output neurons process their inputs by multiplying each input by its weight, adding the product to a total amount, and then passing it through a (transfer or activation) function to generate its result. The whole aim of the BP algorithm is to change the values of the network weights to achieve the minimum error between the predicted output and actual targets.

Figure 1 shows the MLP neural network used for the modelling and optimization of the active chlorine production from a synthetic saline effluent by electrolysis. In continuation of the work of Zaviska et al,[29] current intensity, electrolysis time, chloride ion concentration, and hydrochloric acid concentration were selected as the input neurons, whereas the output layer contains the active chlorine production or the energy consumption.

The gradient descent algorithm has been selected as the learning algorithm for training the neural network with the tansig and purelin transfer function in the hidden and output layer, respectively.

3.2 Learning curves

Learning curves of model performance on the training and validation datasets can be used to diagnose an underfit (high bias), overfit (high variance), or well-fit model.

At first, the data were split into two sets: training and validation. One single example from the training set was taken and used to fit a model. The error related to the model on the validation set and that single training example were measured. The error related to this training instance would be 0, since it is not too overwhelming to fit a single data point perfectly. Since the model is built around a single instance, the error related to the validation set will be quite large. This is due to the lack of generalization ability to the data that it has not seen before. Then the number of training samples is gradually increased until the entire training set is used. As the training set changes, the error values will vary more or less. Thus, two error values have to be monitored: one for the validation set, and one for the training set. If the evolution of the two error values is plotted as the training sample sets change, two curves (so-called learning curves) are obtained. In brief, a learning curve demonstrates how the error varies with an increase in the training set size and demonstrates if one needs a more complex model for the predictions or not.

In this work, learning curves will be plotted for the training samples. To avoid the uncertainty related to the selection of the training sample, each training sample selection was replicated 50 times and the overall mean value has been calculated for that training sample batch.

3.3 Regularization factor

The regularization parameter (lambda) is an input to the objective function to reduce overfitting. This reduces the variance of the estimated regression parameters. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting. It is defined as a term added to the cost function of the model:

$$J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda\sum_{j=1}^{n}\theta_j^2\right] \quad (1)$$

where $J(\theta)$ is the cost function (error), m is the number of data points used for training, x are the input neurons, $h_\theta\left(x^{(i)}\right)$ is the predicted value of sample i, $y^{(i)}$ is the actual value of sample i, λ is the regularization parameter, and θ are the network parameters (weights). In fact, to have control of the fitting parameters, a regularization parameter is used. With any increase in the magnitudes of the network parameters, an increasing penalty will be applied on the cost function. As can be seen, this penalty is relevant on the magnitude of lambda and the squares of the weights. Any increase in lambda can be advantageous up to a certain point, since it reduces the variance which

avoids overfitting. But after this point, important properties of the model start to be lost, introducing more bias into the model (underfitting problem). This implies the importance of the selection of the lambda value. In this study, different lambda values have been tested each time to obtain the best training and cross-validation errors and to present these errors versus the lambda value. The optimum lambda value was selected from this graph by considering the cross-validation and training errors.

3.4 Relative importance of input variables

The weights obtained from ANN training are coefficients between artificial neurons that are analogous to synaptic strengths between the axon and dendrites in a biological neuron in the brain. As in real life, the proportion of the incoming signal to be transmitted to the neuron's body is decided by these weights.[50] Despite the black-box nature of ANNs, to estimate the influence of different independent variables on the output, it is possible to conduct a sensitivity analysis on the ANNs. The relative importance of each input independent variable on the desired output can be obtained through the neural connection weight matrix. First Garson[51] and then Goh[52] proposed a procedure for partitioning the connection weights to determine the relative importance of the various inputs. This method basically involves partitioning the hidden-output connection weights of each hidden neuron into components associated with each input neuron.[53]

Garson's equation based on the partitioning of connection weights can be applied:

$$R_j = \frac{\sum_{m=1}^{m=U_h}\left(\left(\left|W_{jm}^{ih}\right| / \sum_{k=1}^{U_i}\left|W_{km}^{ih}\right|\right)\times\left|W_{mn}^{ho}\right|\right)}{\sum_{k=1}^{k=U_i}\left[\sum_{m=1}^{m=U_h}\left(\left|W_{km}^{ih}\right| / \sum_{k=1}^{U_i}\left|W_{km}^{ih}\right|\right)\times\left|W_{mn}^{ho}\right|\right]} \quad (2)$$

where R_j is the relative importance of the j^{th} independent variable on the output variable; U_i and U_h denote the number of input and hidden neurons, respectively; W is the connection weight value; and the superscripts i, h, and o refer to input, hidden, and output layers, respectively. Also, the subscripts k, m, and n refer to input, hidden, and output neurons, respectively.

3.5 Genetic algorithm and multi-objective optimization

In recent years, evolutionary algorithms, and in particular GAs, have received growing attention among optimization techniques. Gas, with their good global searching ability and flexibility, ease of operation, and without the need for gradient information on the objective (fitness) functions, have become powerful techniques for optimization problems.[54,55] A GA starts with a primary population of candidate solutions and a fitness value is calculated for each solution. Through the algorithm, three stochastic operators are applied to each population which are analogous to chromosomes in a biological context. Selection is choosing the solutions with the highest fitness value to create an intermediate population. The next population is the result of crossover or mutation. By crossover, the selected members are mated in pairs and recombined through genetic manipulation of chromosomes to generate two new solutions (offsprings). Mutation acts as an assurance against lost genetic material and consists of replacing some of the chromosome's genes with new genes. The generation of new populations and calculation of the fitness value for each population is repeated over and over in an iterative method. When a specific termination criterion is met (e.g.,when there is no more change in the population from one iteration to the next or when a satisfactory fitness value is achieved), this process ends.[56-58]

When multiple objectives are specified to a problem, selecting a single solution with specific decision variables could not satisfy all the objectives in a single manner. In fact, objective functions could have non-linear and opposite behaviour to each other. Therefore, a trade-off between all these conflicting objective functions should be made to find the decision variables. This trade-off can be illustrated as Pareto front, which is based on the domination concept. Best solutions in the problem space will be represented in this front, which are the solutions for which there would be no other solution having better values regarding the objective functions. Having Pareto front helps obtain a clear insight throughout the trade-off between different objective functions. This would help to find and focus on promising solutions from a possibly large population of solutions and choose the suitable decision variables regarding the objective functions.[59]

In this regard, the well-known non-dominated sorting genetic algorithm (NSGA-II)[60] has been utilized for multi-objective optimization, leading to a set of solutions (Pareto front), that are the experimental conditions, with respect to maximization of active chlorine production and minimization of energy consumption. The flowchart of the adopted ANN-GA approach in this study is shown in Figure 2.

4 RESULTS AND DISCUSSION

4.1 ANN modelling

After initial data collection, data preprocessing was necessary to manipulate the data into a usable format for processing by the ANNs. Feature normalization, Equation (3), has been selected and returns a normalized version of feature (input) X where the mean value of each feature is 0 and the standard deviation is 1:

$$y=\frac{X_i-\mu_i}{S_i} \quad (3)$$

where y is the normalized value of X_i. The μ_i and the S_i are the mean and standard deviation values of X_i, respectively. Normalization helps because it ensures both that the network's learning regards all input features to a similar extent and that there are both positive and negative values used as inputs for the next layer, which makes learning more flexible.

At the first step of the ANN modelling, a three-layer network was configured with the 16 FD experiments. The ANN model was constructed with five neurons in the hidden layer with tansig and purelin transfer functions in the hidden and output layer, respectively, and trained by the gradient descent algorithm. While the coefficient of determination for the FD was reported as R^2=0.982, this value increased to R^2=0.999 with the ANN model.

Finding optimal conditions to produce active chlorine is a multi-objective optimization by taking into account the energy consumption of the process.This has been done by conducting 14 more experiments using a CCD.[29] In total, 30 experimental data points were used for ANN modelling for the purpose of training and validation, including sets of 24 and six samples for each, respectively.

4.1.1 Learning curves and impact of regularization

For an ANN model, it is necessary to have an overview of the state of the model in order to check whether there is a high bias (underfit) problem or high variance (overfit) issue. This helps decide whether a more complex model (with more hidden layers and neurons) is required or not. Figure 3 shows the mean learning curve obtained for different numbers of training examples. As explained in Section 3.2, learning curves show how the error changes as the training set size increases and demonstrate whether or not one needs a more complex model for the predictions.

In this figure, the training and cross-validation error have been plotted versus the number of training examples in the training set. This figure helps to have an overview of the type of problem dealt with. In case of a high bias model (underfit), there would be high errors for both training and cross validation data sets. For a high variance model (overfit), the training error would be low and the cross validation error would be much higher. Also, in case of a high variance problem, having more data along with not having a more complex model (e.g. adding more hidden layers or hidden neurons) would help the modelling process. As Figure 2 shows, there is a gap between the training and cross-validation error, with a very small error for the training set and a much higher error for the cross-validation set. Also, as the number of samples in the training set increases, the cross-validation error decreases, which proves that increasing the number of samples is a good solution for a high variance problems as diagnosed for this case. The ANN model for each point in the learning curve has been trained 50 times with random sampling from the available data and the mean, minimum, and maximum error values have been represented. Decision making based on the mean value of 50 times iterations for each training with random sampling helps to decrease the risk of uncertainty related to stochastic behaviour of ANN modelling.

Figure 3 shows the mean learning curve for the validation samples with the minimum and maximum values obtained for the 50 iterations of training. As said before, for a high variance problem, having a more complex model does not help. This is shown in Table 2 where three different neural network configurations are presented with their correlation coefficients for the training, validation, and all data sets. It can be concluded that for this high variance problem there is no need for a more complex model that inherently would increase the overfitting issue.

In the presence of the high variance problem, using the regularization factor can help. Regularization makes slight modifications to the learning algorithm such that the model generalizes better and the model's performance on unseen data is improved. Therefore, a graph of error versus regularization factor (lambda) helps to optimally select the best lambda value. In our case, the best lambda value is 3 (Figure 4).

Thus, a three-layer feedforward back-propagation network with five hidden neurons and a regularization factor value of 3 is selected for the optimization.

The selected network for active chlorine production has a coefficient of determination R^2=0.979 while this value for the RSM with the CCD was reported as R^2=0.964.

Also, for the multi-objective optimization, the selected neural network for predicting the energy consumption is configured with four hidden layers and a sigmoid transfer function at the hidden layer. This network has a performance of R^2_{train}=0.997, $R^2_{validation}$=0.951, and R^2_{All}=0.985, which compares favourably to the RSM regression performance R^2=0.990. Parity plots for ANN and RSM models regarding active chlorine production and energy consumption are represented in Figure 5.

It should be mentioned that in RSM, all available data was used for the linear regression method for curve fitting. In ANN modelling, however the data is divided into training and validation sets. Table 3 represents the CCD experimental plan, actual and predicted values of the ANN, and RSM models for active chlorine production and energy consumption. Also, performance criteria (R^2) and mean squared error (MSE) for each model and dependent variable are reported in Table 3. As can be seen, the ANN method performs slightly better than RSM for predicting active chlorine prediction and energy consumption of the electrolysis process.

4.1.2 Relative importance of input variables

Using the factorial design method, the influence of four main experimental factors was investigated. Based on the sensitivity analysis results, electrolysis time and current intensity with 82.8% contribution on the active chlorine production were the two most influential factors. In order to assess the relative importance of the input variables for the ANN model, the neural net weight matrix can be used. The relative importance of the various variables, calculated by Equation (2), is shown in Figure 6.

Like the FD method, ANN weight analysis derived by Garson's algorithm, described in Section 3.4, illustrates that electrolysis time and current intensity are the most important factors for predicting the production of active chlorine. Garson's algorithm investigation on neural network weights shows about 81.5% influence on active chlorine production for these two main independent variables (compared to an 82.8% influence in the FD analysis). The H_3O^+ and NaCl concentration represent the remaining 18.5% of the investigated response (active chlorine production). Figure 6 shows the compatible and reliable results of the ANN model, similar to the RSM outcomes.

4.1.3 Response surfaces of the RSM and ANN

The effect of electrolysis time and current intensity on the production of active chlorine is illustrated in Figure 7. Note that the concentrations of acid and chloride are kept constant at the centre of the investigated experimental ranges (0.08 mol/L and 0.55 mol/L, respectively). In these conditions, by increasing the electrolysis time, the active chlorine concentration rises for all current intensities studied. As can be seen, the RSM response surface is a quadratic model that has to fit the predicted values on this surface, whereas the ANN model with its high ability for nonlinearity can fit the data in a much finer way. It can be concluded that active chlorine can be produced up to more than 33 mg/L at electrolysis times longer than 30 minutes and for current intensities at the higher values of 1.4 A.

4.2 Multi-objective optimisation with GA

Simultaneous optimization of hypochlorous acid production and the energy consumption is defined in the category of multi-objective optimization. No unique solution can be derived for a multi-objective optimization problem, except for Pareto front solutions which are inherently non-dominated. A MATLAB script using two ANN models developed for the hypochlorous acid production and energy consumption was written to create a cost (fitness) function. The multi-objective optimization was conducted by aiming for both maximizations of the hypochlorous acid production and minimization of energy consumption. The bounds of the four independent variables were chosen by the ranges of the experiments. The following NSGA-II algorithm options were set:

- Population size: 50
- Maximum number of iterations: 150
- Selection function: Tournament selection
- Crossover strength: 0.7
- Mutation strength: 0.3
- Distance measure function: Distance crowding

The maximum number of iterations was used as stopping criterion. For the purpose of comparison, Pareto fronts have also been generated using RSM models with the same NSGA-II algorithm. After 150 iterations, the Pareto front of Figure 8 is obtained. The decision variables of the electrolysis process corresponding to each of the ANN-GA Pareto front solutions are tabulated in Table 4.

The general method used for RSM optimization is single optimization with multiple responses using desirability functions and weighting factors representing the importance of each response. In this approach, usually, just a single optimal point is reported based on the desirability value. No Pareto front will be provided. Since the RSM models are generated by linear regression method, the Pareto front provided by NSGA-II for this approach is linear. Each point on the Pareto front indicates that there is no other process decision variables that can have the same active chlorine production with lower cost or, in other words, with the same cost there are no other process decision variables that can produce higher active chlorine. As reported in the paper on RSM,[29] the optimal conditions for the electrolytic reactor were obtained by 27 minutes of electrolysis time with the concentrations of hydrochloric acid and chloride sodium of 0.11 and 0.8 mol/L, respectively. Under these conditions, it was mentioned that production of 30.60 mg/L of active chlorine could be possible with 0.54 kWh/m^3 energy consumption. Conversely, in the ANN-GA approach of this paper, one of the solutions implies that with 0.53 kWh/m^3 of energy consumption, 34.92 mg/L of active chlorine can be produced under different operating conditions. Also, production of 30.47 mg/L of active chlorine is achievable with 0.47 kWh/m^3 energy consumption. These imply more economic conditions for higher active chlorine production. Unfortunately, our results could not be verified by the experiments since the experimental set up was no longer available. However, by comparing Tables 3 and 4, at least one Pareto optimal point appears similar to an experimental assay (Solution number 1: Time=35 minutes, current=1.6 A, $[H_3O^+]$=0.11 mol/L, and $[Cl^-]$=0.8 mol/L). Under these conditions, 42.23 mg/L of active chlorine can be produced with 0.75 kWh/m^3 energy consumption compared to experimental values of 46 mg/L and 0.76 kWh/m^3 for active chlorine and energy consumption, respectively. The relative error for this optimal point is 0.082 and 0.013 for active chlorine production and energy consumption, respectively.

The ANN-GA approach introduced in this study provides optimal operational conditions based on active chlorine production and energy cost. The advantage of having a Pareto front for

industrial process designers and operators is that different operational conditions (decision variables) can be selected based on preference for each objective. This gives an insight on the trade-off between the different objective functions involved in this industrial process.

Although some effort has been made in this study to obtain the best possible ANN models for describing the process, there remain some ANN hyper-parameters that can be optimized in further studies. These hyper-parameters, including transfer functions and learning rate, may have direct impact on the ANN modelling performance.

5 CONCLUSIONS

The ANN-GA methodology was successfully applied to an electrolysis process for active chlorine production. MLP feedforward neural networks were developed for actives chlorine production and enegy consumption. To diagnose whether there is a danger for high-variance or high-bias error and to prevent over-fitting of the model, learning curves along with regularization factor analysis were utilized during the training of the neural network models. Figure 7 indicated that the ANN model was able to describe the nonlinearities related to the experimental process better than the previously proposed RSM model with a coefficient of determination of 0.979 and 0.985 for production of active chlorine and energy consumption, respectively. Analysis of the relative importance of the variables indicated that electrolysis time and current intensity are the two most influential parameters with a total effect of 81.5% on active chlorine production.

To obtain a clear insight throughout the trade-off between different objective functions involved in the electrolysis process, the NSGA-II algorithm was used for multi-objective optimization of the process regarding active chlorine production and energy consumption. The Pareto front derived by GA led to the generation of non-dominated optimal points (operating conditions) for maximum active chlorine production at minimum energy consumption. The proposed ANN-GA methodology can give insight in how to efficiently choose the process operation parameters (decision variables) for the desired objectives. This approach can be adapted to other processes if the experimental data already exist.

ACKNOWLEDGEMENTS

Sincere thanks are extended to the National Sciences and Engineering Research Council of Canada and CREATE TEDGIEER program for their financial contribution to this study. Peter A. Vanrolleghem holds the Canada Research Chair on Water Quality Modelling.

CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] K. Rajeshwar, J. G. Ibanez, *Environmental Electrochemistry: Fundamentals and Applications in Pollution Abatement*, Academic Press, San Diego **1997**

[2] Y. Feng, L. Yang, J. Liu, B. E. Logan, *Environ Sci-Wat Res* **2016**, *2*, 800.

[3] M. J. Martín de Vidales, C. Sáez, P. Cañizares, M. A. Rodrigo, *J Chem Technol Biotechnol* **2012**, *87*, 225.

[4] R. Daghrir, P. Drogui, J. François Blais, G. Mercier, *J Environ Eng* **2012**, *138*, 1146.

[5] H. Olvera-Vargas, N. Oturan, M. A. Oturan, E. Brillas, *Sep Purif Technol* **2015**, *146*, 127.

[6] Y. Zhang, K. Ghyselbrecht, B. Meesschaert, L. Pinoy, B. Van der Bruggen, *J Membr Sci* **2011**, *378*, 101.

[7] S. Zendehboudi, N. Rezaei, A. Lohi, *Applied Energy* **2018**, *228*, 2539.

[8] S. Nandi, Y. Badhe, J. Lonari, U. Sridevi, B. S. Rao, S. S. Tambe, B. D. Kulkarni, *Chem Eng J* **2004**, *97*, 115.

[9] K. Jardak, A. Dirany, P. Drogui, M. A. El Khakani, *Process Saf Environ Prot* **2017**, *105*, 12.

[10] N. Tran, P. Drogui, S. K. Brar, *J Chem Technol Biotechnol* **2015**, *90*, 921.

[11] O. Dia, P. Drogui, G. Buelna, R. Dube, B. S. Ihsen, *Chemosphere* **2017**, *168*, 1136.

[12] K. Jardak, A. Dirany, P. Drogui, M. A. El Khakani, *Sep Purif Technol* **2016**, *168*, 215.

[13] H. A. Moreno-Casillas, D. L. Cocke, J. A. G. Gomes, P. Morkovsky, J. R. Parga, E. Peterson, *Sep Purif Technol* **2007**, *56*, 204.

[14] A. Giwa, S. Daer, I. Ahmed, P. R. Marpu, S. W. Hasan, *J Water Process Eng* **2016**, *11*, 88.

[15] S. Curteanu, K. Godini, C. G. Piuleac, G. Azarian, A. R. Rahmani, C. Butnariu, *Ind Eng Chem Res* **2014**, *53*, 4902.

[16] A. Kamari, A. H. Mohammadi, A. Bahadori, S. Zendehboudi, *Pet Sci Technol* **2014**, *32*, 2837.

[17] I. Nejatian, M. Kanani, M. Arabloo, A. Bahadori, S. Zendehboudi, *Journal of Natural Gas Science and Engineering* **2014**, *18*, 155.

[18] C. G. Piuleac, S. Curteanu, M. A. Rodrigo, C. Sáez, F. J. Fernández, *Cent Eur J Chem* **2013**, *11*, 1213.

[19] H. Karimi, M. Ghaedi, *J Ind Eng Chem* **2014**, *20*, 2471.

[20] A. Boucheikhchoukh, J. Thibault, C. Fauteux-Lefebvre, *Can J Chem Eng* **2020**, *1*, 16.

[21] V. Sangal, P. Kaur, J. Kushwaha, *RSC Advances* **2015**, *5*, 34663.

[22] T. da Silva Ribeiro, C. D. Grossi, A. G. Merma, B. F. dos Santos, M. L. Torem, *Miner Eng* **2019**, *131*, 8.

[23] M. Radwan, M. Gar Alalm, H. Eletriby, *J Water Process Eng* **2018**, *22*, 155.

[24] N. Chindapan, S. S. Sablani, N. Chiewchan, S. Devahastin, *Food Bioprocess Tech* **2013**, *6*, 2695.

[25] Z. Wang, S. Wiebe, H. Shang, *Can J Chem Eng* **2016**, *94*, 415.

[26] M. Ghaedi, A. M. Ghaedi, F. Abdi, M. Roosta, R. Sahraei, A. Daneshfar, *J Ind Eng Chem* **2014**, *20*, 787.

[27] D. Whitley, *Statistics and Computing* **1994**, *4*, 65.

[28] J. D. Schaffer, D. Whitley, L. J. Eshelman, presented at [Proceedings] COGANN-92: International Workshop on Combinations of Genetic Algorithms and Neural Networks, Baltimore, June **1992**

[29] F. Zaviska, P. Drogui, G. Pablo, *Desalination* **2012**, *296*, 16.

[30] S. Elmas, F. Ambroz, D. Chugh, T. Nann, *Langmuir* **2016**, *32*, 4952.

[31] F. Zaviska, P. Drogui, J.-F. Blais, G. Mercier, *J Appl Electrochem* **2009**, *39*, 2397.

[32] Y. U. R. Huang, Y.-C. Hung, S.-Y. Hsu, Y.-w. Huang, D.-F. Hwang, *Food Control* **2008**, *19*, 329.

[33] C. d. N. Brito, D. M. de Araújo, C. A. Martínez-Huitle, M. A. Rodrigo, *Electrochem Commun* **2015**, *55*, 34.

[34] K. Y. Khouzam, presented at IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, Pittsburgh, July **2008**

[35] F. H. Oliveira, M. E. Osugi, F. M. M. Paschoal, D. Profeti, P. Olivi, M. V. B. Zanoni, *J Appl Electrochem* **2007**, *37*, 583.

[36] G.-S. W. Hsu, C.-W. Hsia, S.-Y. Hsu, *J Food Drug Anal* **2015**, *23*, 729.

[37] G.-S. W. Hsu, C.-W. Hsia, S.-Y. Hsu, *J Food Drug Anal* **2015**, *23*, 735.

[38] G.-S. W. Hsu, S.-Y. Hsu, *J Food Drug Anal* **2016**, *24*, 575.

[39] K. C. Pillai, T. O. Kwon, B. B. Park, I. S. Moon, *J Hazard Mater* **2009**, *164*, 812.

[40] S. A. Abdul-Wahab, M. A. Al-Weshahi, *Water Resour Manage* **2009**, *23*, 2437.

[41] A. Picos, J. M. Peralta-Hernández, *Water Sci Technol* **2018**, *78*, 925.

[42] M. M. Ghiasi, A. Bahadori, S. Zendehboudi, *Fuel* **2014**, *117*, 33.

[43] M. H. Entezari, P. Kruus, *Ultrason Sonochem* **1994**, *1*, S75.

[44] S. Haykin, *Neural Networks. A Comprehensive Foundation*, Prentice Hall PTR, USA **1998**

[45] K. P. Singh, A. Basant, A. Malik, G. Jain, *Ecol Model* **2009**, *220*, 888.

[46] M. Adya, F. Collopy, *J Forecast* **1998**, *17*, 481.

[47] I. N. Da Silva, D. Hernane Spatti, R. Andrade Flauzino, L. H. B. Liboni, S. F. Dos Reis Alves, *Artificial Neural Networks : A Practical Course*, Springer, Cham, Switzerland **2017**

[48] A. R. Carvalho, F. M. Ramos, A. A. Chaves, *Neural Comput Appl* **2011**, *20*, 1273.

[49] M. M. Ghiasi, A. Bahadori, S. Zendehboudi, I. Chatzis, *Fuel* **2015**, *140*, 421.

[50] A. R. Khataee, M. B. Kasiri, *J Mol Catal A: Chem* **2010**, *331*, 86.

[51] G. D. Garson, *AI Expert* **1991**, *6*, 46.

[52] A. T. C. Goh, *Artificial Intelligence in Engineering* **1995**, *9*, 143.

[53] M. Gevrey, I. Dimopoulos, S. Lek, *Ecol Model* **2003**, *160*, 249.

[54] S. Ding, C. Su, J. Yu, *Artif Intel Rev* **2011**, *36*, 153.

[55] S. Curteanu, M. Cazacu, *J Macromol Sci A* **2007**, *45*, 23.

[56] H. Ridha, M. Njah, M. Chtourou, *Model Ident Control* **2008**, *5*, 305.

[57] A. Ansari, A. A. Bakar, presented at 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology, Kota Kinabalu, December **2014**

[58] S. P. Niculescu, *J Mol Struc-THEOCHEM* **2003**, *622*, 71.

[59] X. Hu, M. Wang, X. Hu, M. S. Leeson, presented at IEEE Congress on Evolutionary Computation (CEC), Beijing, July **2014**

[60] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, *IEEE T Evolut Comput* **2002**, *6*, 182.

Figure Captions

FIGURE 1 Schematic of the multilayer percepterons (MLP) neural network used for the modelling and optimization of the active chlorine production

FIGURE 2 Flowchart of ANN-GA (artificial neural network-genetic algorithm) methodology used for multi-objective optimization

FIGURE 3 Mean learning curves for different numbers of training examples (Error bars are generated with 50 time training for each training set)

FIGURE 4 Impact of regularization factor on model performance

FIGURE 5 Parity plots of predicted versus experimental values of active chlorine production and energy consumption for ANN (artificial neural networks) and RSM (response surface methodology) models

FIGURE 6 Importance (%) of the input variables on the electrochemical active chlorine production

FIGURE 7 Response surface graph of active chlorine production versus electrolysis time and current intensity: (A) ANN, artificial neural networks and (B) RSM, response surface methodology

FIGURE 8 Pareto fronts for multi-objective optimization of active chlorine production and energy consumption

Table Captions

TABLE 1 Experimental operating conditions range

Parameter	**Min. value**	**Max. value**
Electrolysis time (min)	15	35
Current intensity (A)	0.8	1.6
$[H_3O^+]$ (mol/L)	0.05	0.11
[NaCl] (mol/L)	0.3	0.8

TABLE 1 Feedforward backpropagation networks

Configuration	**R^2 training**	**R^2 validation**	**R^2 All**
# Samples	24	6	30
5 hidden neurons	0.9803	0.9701	0.9791
6 hidden neurons	0.9404	0.9592	0.9453
7 hidden neurons	0.9445	0.9562	0.9481

TABLE 2 Actual and predicted values of central composite designed experiments.

Experimental assays				**Chlorine production (mg/l)**			**Energy consumption (kWh/m^3)**		
Time (min)	Current (A)	[H_3O^+] (mol/L)	[Cl^-] (mol/L)	Actual	ANN predicted	RSM predicted	Actual	ANN predicted	RSM predicted
[illegible]	1.6	0.05	0.3	5.04	5.34	4.57	0.250	0.280	0.279
25	1.2	0.08	0.55	10.7	10.28	10.7	0.312	0.322	0.330
15	1.6	0.11	0.3	5.32	6.84	9.09	0.300	0.302	0.297
15	0.8	0.05	0.3	1.6	1.52	2.71	0.120	0.122	0.111
25	1.2	0.08	0.55	9	10.28	10.7	0.310	0.322	0.330
[illegible]	1.6	0.11	0.8	46	42.23	44.33	0.760	0.747	0.783
25	2	0.08	0.55	46.6	42.39	41.24	0.708	0.706	0.726
15	1.6	0.11	0.8	6.94	7.96	13.35	0.280	0.281	0.311
15	0.8	0.11	0.8	0.56	1.57	0.61	0.120	0.114	0.143
[illegible]	1.2	0.08	0.55	13.8	10.28	10.7	0.330	0.322	0.330
25	1.2	0.08	0.55	13.76	10.28	10.7	0.323	0.322	0.330
25	1.2	0.08	0.55	10.2	10.28	10.7	0.314	0.322	0.330
15	0.8	0.05	0.8	1.54	1.52	0.61	0.120	0.115	0.125
35	0.8	0.11	0.8	7.26	10.67	8.63	0.300	0.295	0.311
25	1.2	0.08	0.05	13	11.88	9.62	0.375	0.342	0.316
35	1.6	0.11	0.3	36.5	36.96	40.07	0.770	0.770	0.769
25	1.2	0.08	0.55	9.4	10.28	10.7	0.318	0.322	0.330
15	1.6	0.05	0.8	8.96	8.75	8.83	0.300	0.288	0.293
25	1.2	0.08	1.05	13.4	13.28	11.78	0.308	0.315	0.344
25	0.4	0.08	0.55	1.68	3.06	3.68	0.083	0.085	0.086
25	1.2	0.02	0.55	4	5.77	8.44	0.375	0.363	0.312
35	0.8	0.11	0.3	9.52	11.09	10.73	0.310	0.326	0.297
5	1.2	0.08	0.55	0.1	1.52	-3.12	0.125	0.120	0.098
35	1.6	0.05	0.8	37.2	38.01	39.81	0.780	0.785	0.765
45	1.2	0.08	0.55	36	34.34	35.88	0.688	0.795	0.738

25	1.2	0.14	0.55	16.4	11.90	12.96	0.300	0.313	0.348
35	1.6	0.05	0.3	34.4	34.39	35.55	0.820	0.794	0.751
35	0.8	0.05	0.8	10.3	11.17	8.63	0.310	0.364	0.293
15	0.8	0.11	0.3	1.74	1.62	2.71	0.120	0.145	0.129
35	0.8	0.05	0.3	13.5	11.46	10.73	0.330	0.368	0.279
R-squared (R^2)				-	**0.979**	**0.964**	-	**0.985**	**0.990**
Mean squared error (MSE)				-	**3.826**	**6.952**	-	**6.903e-04**	**9.043e-04**

TABLE 4 Decision variables of the electrolysis process corresponding to each of the Pareto front solutions presented in Figure 8

Solution No.	Time (min)	Current (A)	$[H_3O^+]$ (mol/L)	$[Cl^-]$ (mol/L)	HClO production (mg/L)	Energy consumption (kWh/m^3)
1	35.00	1.60	0.11	0.80	42.23	0.75
2	15.00	0.80	0.08	0.80	1.53	0.11
3	22.75	1.57	0.11	0.80	25.45	0.42
4	20.70	1.45	0.11	0.80	13.69	0.33
5	17.92	0.80	0.11	0.80	1.87	0.12
6	18.81	0.80	0.11	0.80	2.16	0.13
7	21.43	1.36	0.11	0.80	11.40	0.30
8	33.27	1.60	0.11	0.80	41.49	0.70
9	24.13	0.80	0.11	0.80	7.33	0.16
10	20.33	1.53	0.11	0.80	16.89	0.35
11	33.48	1.53	0.11	0.80	40.13	0.67
12	27.36	1.05	0.11	0.80	10.62	0.28
13	20.38	1.59	0.11	0.80	21.11	0.37
14	25.29	0.80	0.11	0.80	8.34	0.17
15	23.04	0.80	0.11	0.80	6.03	0.15
16	23.81	1.59	0.11	0.80	28.84	0.45
17	21.47	0.80	0.11	0.80	4.08	0.14
18	26.53	0.96	0.11	0.80	9.72	0.24
19	28.25	0.81	0.11	0.80	9.60	0.20
20	28.75	1.59	0.11	0.80	37.21	0.57
21	29.71	1.59	0.11	0.80	38.50	0.60
22	28.56	0.93	0.11	0.80	10.12	0.25
23	27.03	0.81	0.11	0.80	9.26	0.19
24	19.61	0.80	0.11	0.80	2.48	0.13
25	20.12	0.80	0.11	0.80	2.84	0.13

26	20.64	0.80	0.11	0.80	3.13	0.14
27	19.92	0.80	0.11	0.80	2.71	0.13
28	26.19	1.59	0.11	0.80	33.72	0.51
29	21.90	0.80	0.11	0.80	4.52	0.14
30	25.41	1.59	0.11	0.80	32.03	0.49
31	21.12	0.81	0.11	0.80	3.71	0.14
32	28.41	1.56	0.11	0.80	35.47	0.55
33	30.77	1.60	0.11	0.80	39.76	0.63
34	24.65	1.59	0.11	0.80	30.47	0.47
35	20.97	1.55	0.11	0.80	19.53	0.37
36	27.16	1.59	0.11	0.80	34.92	0.53
37	22.15	0.80	0.11	0.80	4.82	0.15
38	20.95	0.80	0.11	0.80	3.51	0.14
39	17.36	0.80	0.11	0.80	1.77	0.12
40	15.78	0.80	0.11	0.80	1.60	0.12
41	20.80	0.80	0.11	0.80	3.27	0.14
42	22.46	0.80	0.11	0.80	5.28	0.15
43	19.31	0.80	0.11	0.80	2.29	0.13
44	30.55	1.59	0.11	0.80	39.47	0.63
45	22.57	0.81	0.11	0.80	5.50	0.15
46	16.98	0.80	0.11	0.80	1.71	0.12
47	16.45	0.80	0.11	0.80	1.65	0.12
48	19.31	0.80	0.11	0.80	2.35	0.13

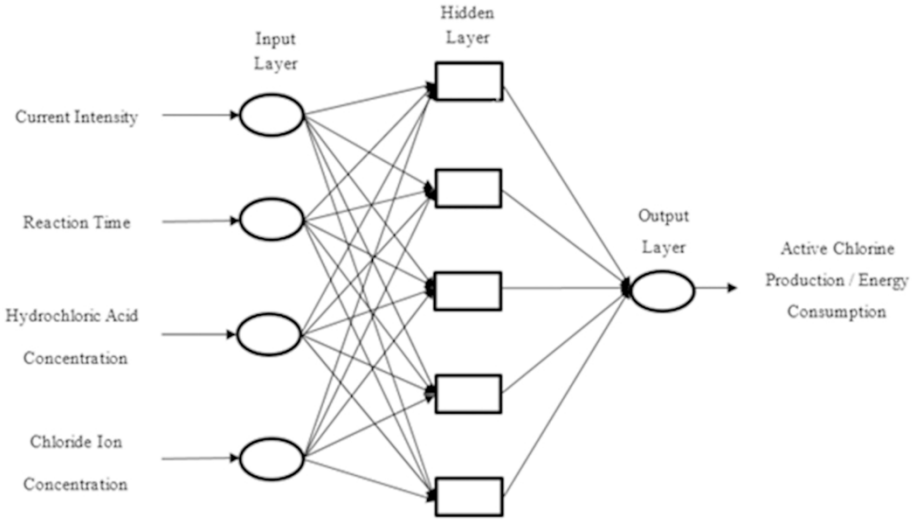

FIGURE 1 Schematic of the MLP neural network used for the modelling and optimization of the active chlorine production

177x101mm (300 x 300 DPI)

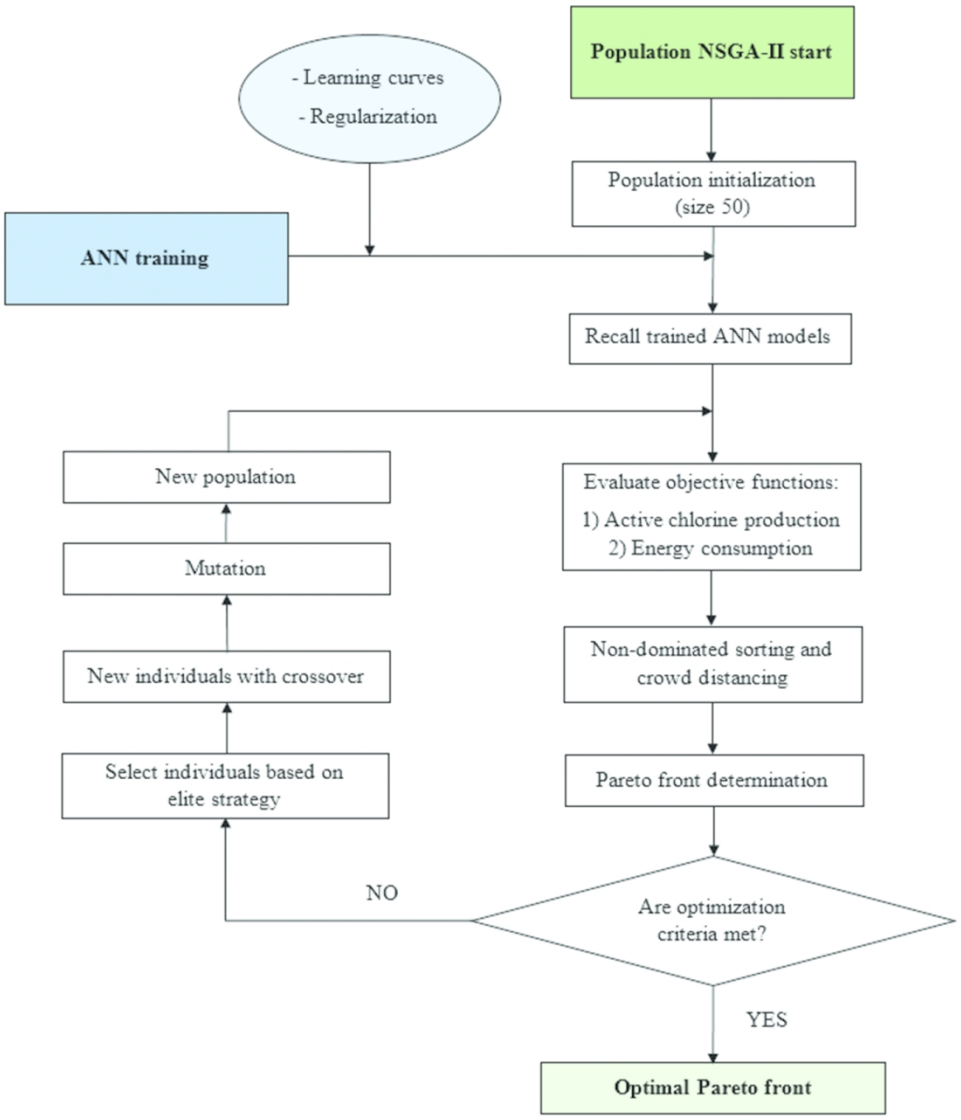

FIGURE 2 Flowchart of ANN-GA methodology used for multi-objective optimization

127x144mm (300 x 300 DPI)

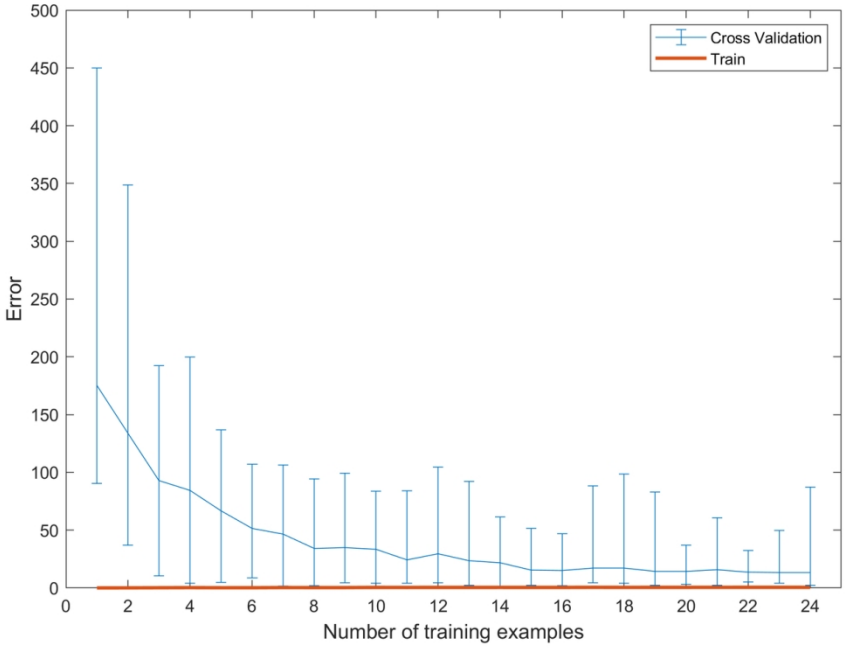

FIGURE 3 Mean learning curves for different numbers of training examples (Error bars are generated with 50 time training for each training set)

127x90mm (300 x 300 DPI)

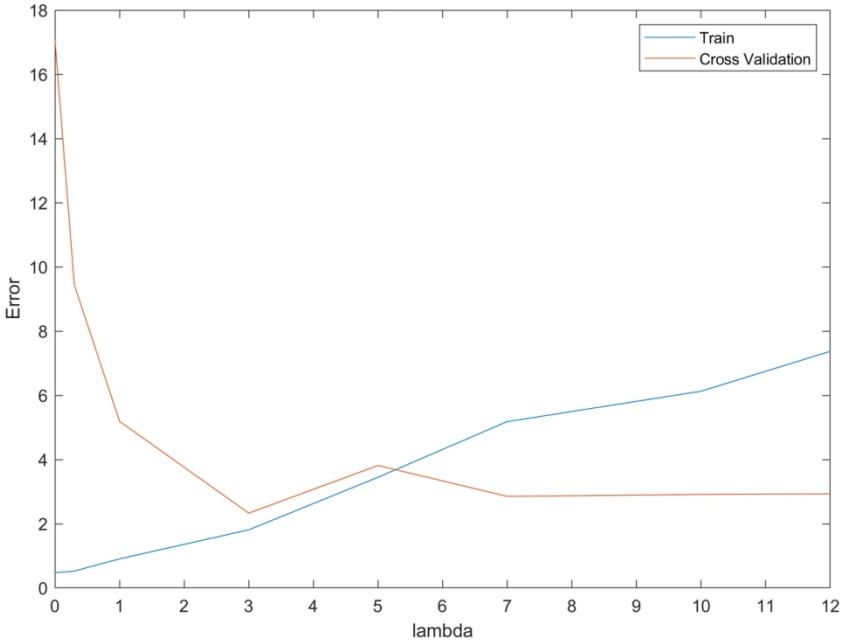

FIGURE 4 Impact of regularization factor on model performance

127x90mm (300 x 300 DPI)

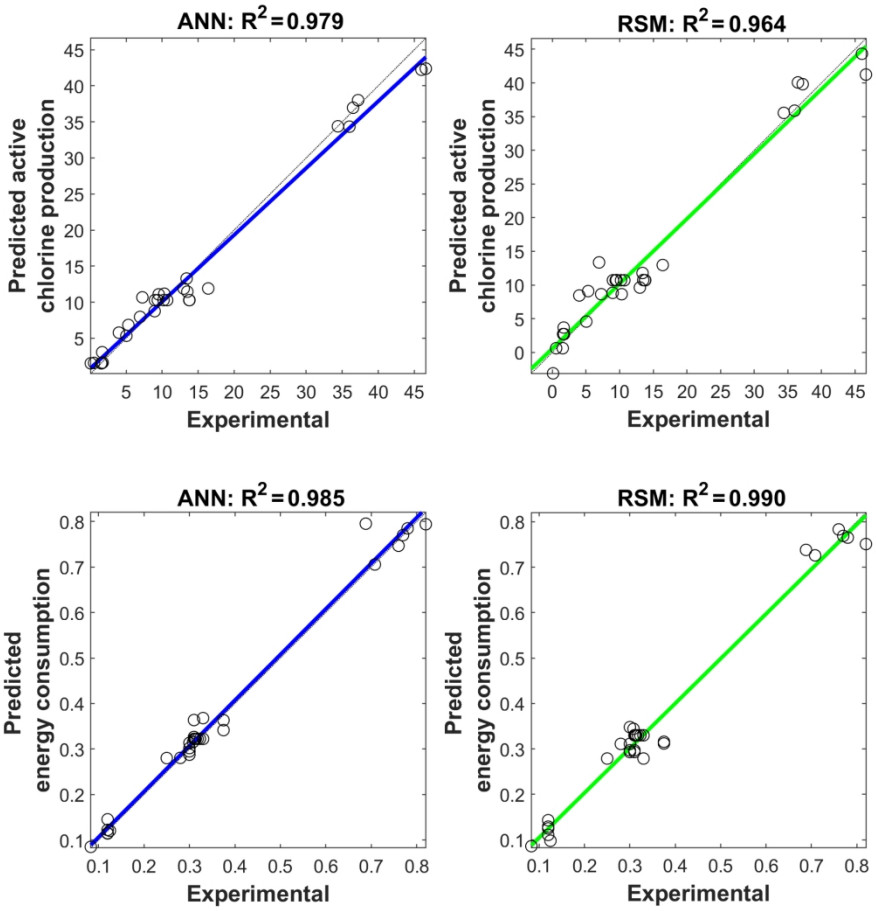

FIGURE 5 Parity plots of predicted versus experimental values of active chlorine production and energy consumption for ANN and RSM models

177x177mm (300 x 300 DPI)

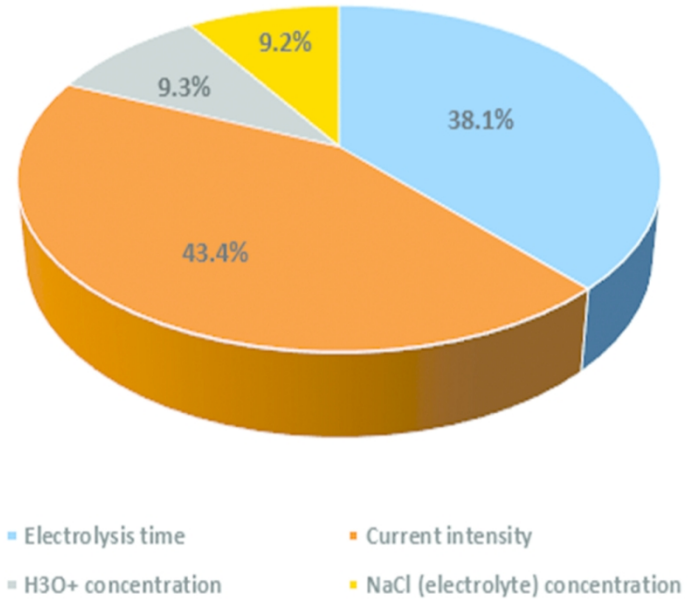

FIGURE 6 Importance (%) of the input variables on the electrochemical active chlorine production

127x88mm (300 x 300 DPI)

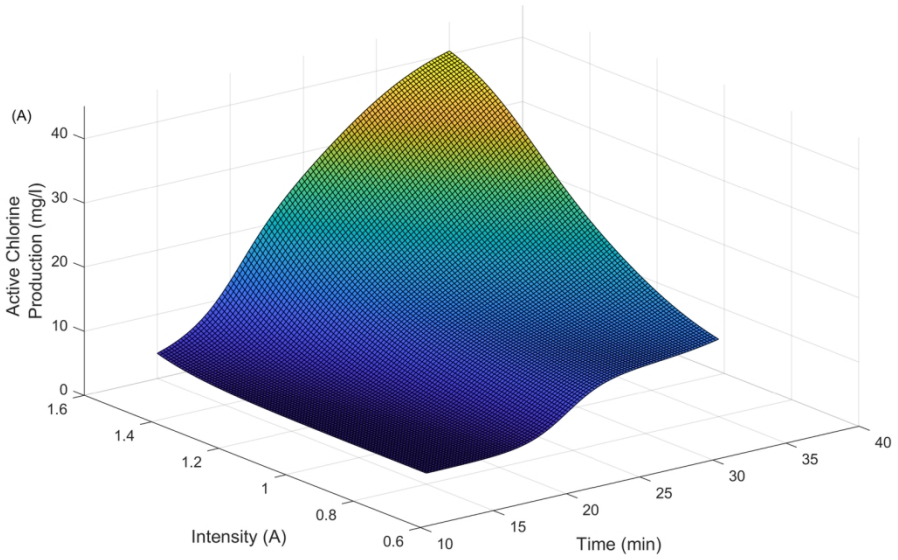

FIGURE 7 Response surface graph of active chlorine production versus electrolysis time and current intensity, (A) ANN, (B) RSM

152x97mm (300 x 300 DPI)

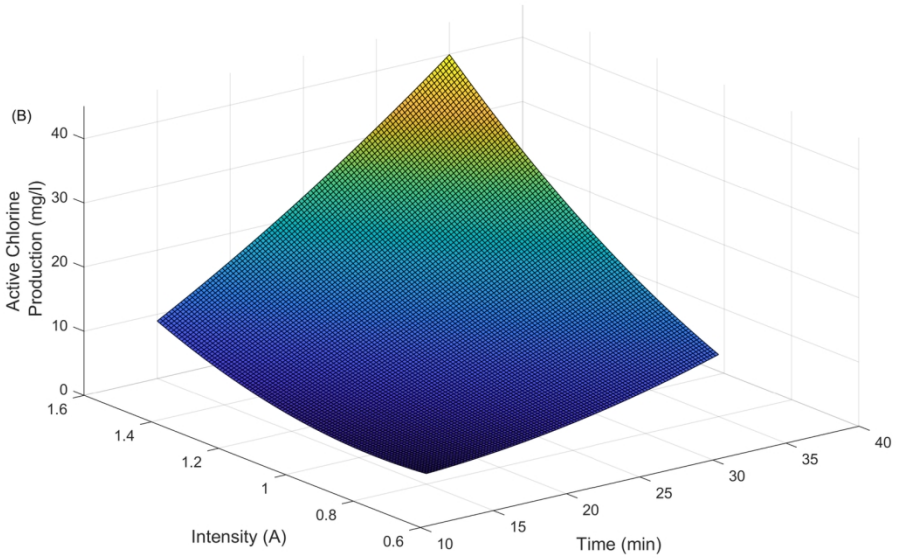

FIGURE 7 Response surface graph of active chlorine production versus electrolysis time and current intensity, (A) ANN, (B) RSM

152x97mm (300 x 300 DPI)

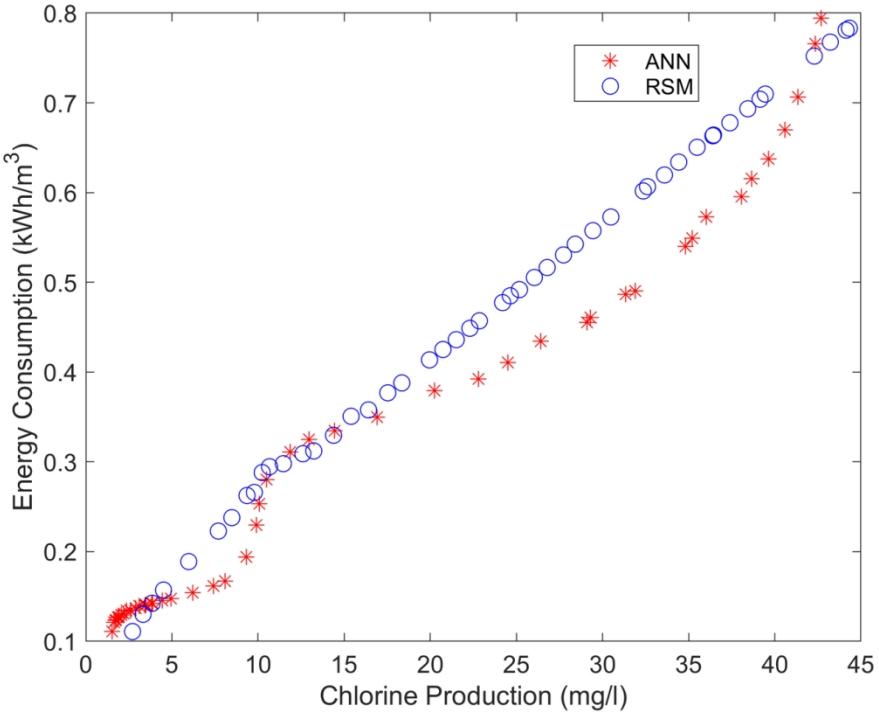

FIGURE 8 Pareto fronts for multi-objective optimization of active chlorine production and energy consumption

127x97mm (300 x 300 DPI)