

<https://doi.org/10.15255/KUI.2021.001>

KUI-47/2021

Original scientific paper

Received January 8, 2021

Accepted March 18, 2021

Optimisation and Prediction of the Coagulant Dose for the Elimination of Organic Micropollutants Based on Turbidity

H. Tahraoui,^{a*} A.-E. Belhadj,^a N. Moula,^bS. Bouranene,^c and A. Amrane^dThis work is licensed under a
Creative Commons Attribution 4.0
International License^a Laboratory of Biomaterials and Transport Phenomena (LBMPT), Nouveau Pôle Urbain, University of Médéa, 26 000 Médéa, Algeria^b Fundamental and Applied Research in Animal and Health (FARAH), Department of Veterinary Management of Animal Resources, Faculty of Veterinary Medicine, University of Liege, Liege 4000, Belgium^c University of Souk Ahras, Dept. Process Engineering, STEE Lab., Rue d'annaba, BP 1553, 41 000 Souk-Ahras, Algeria^d University of Rennes, Ecole Nationale Supérieure de Chimie de Rennes, CNRS, ISCR – UMR6226, F-35 000 Rennes, France

Abstract

In this study, four different mathematical models were considered to predict the coagulant dose in view of turbidity removal: response surface methodology (RSM), artificial neural networks (ANN), support vector machine (SVM), and adaptive neuro-fuzzy inference system (ANFIS). The results showed that all models accurately fitted the experimental data, even if the ANN model was slightly above the other models. The SVM model led to almost similar results as the ANN model; the only difference was in the validation phase, since the correlation coefficient was very high and the statistical indicators were very low for the ANN model compared to the SVM model. However, from an economic point of view, the SVM model was more appropriate than the ANN model, since its number of parameters was 22, *i.e.*, almost half the number of parameters of the ANN model (43 parameters), while the results were almost similar in all the data phase. To reduce the economic costs further, the RSM model can also be used, which remained very useful due to its high coefficients related to the number of parameters – only 13. In addition, the statistical indicators of the RSM model remained acceptable.

Keywords

Coagulation, physicochemical analysis, response surface methodology, artificial neural networks, support vector machine, adaptive neuro-fuzzy inference system

1 Introduction

Water treatment is essential owing to the unavailability of drinking water and its lack in developing countries in Africa, which is mainly due to human and industrial pollution. Water treatment requires specific attention to be in agreement with the standards of consumption.¹ It includes several methods, such as coagulation-flocculation, sedimentation or flotation, filtration and disinfection, *etc.*² Among these processes, coagulation-flocculation has been classified as the best available technology for large-scale drinking water production.³ It has played and will continue to play an important role, directly or indirectly, in the control of particles, microorganisms, natural organic matter (NOM), synthetic organic carbon, disinfection by-product precursors (DBP), and certain inorganic ions and metals, and finally, in the control of drinking water quality.⁴ Coagulation is a process of destabilisation of particles in water by neutralising the negative charge on the particle surface, followed by the agglomeration of the neutralised micro-flakes to form settleable and filterable macro-flocculants.^{5,6} The coagulation process is primarily used to remove turbidity

in a water treatment plant.⁷ It is used in more than half of the water purification plants in Algeria. Coagulation efficiency depends on the water characteristics, such as colour, turbidity, temperature, pH, alkalinity, dissolved salts not least the influence of mineralisation.⁸ It also depends on the treatment itself, such as the nature and dosage of coagulant, the conditions of agitation, as well as on the treatment conditions.⁹ Jar tests are generally conducted to determine optimal coagulant doses,^{10,11} with the goal of reducing costly chemical wastage and operating costs, and achieving drinking water quality objectives.¹² However, to date, this process poses a difficult control problem because it is non-linear and complex. Several studies have been conducted to apply different control strategies to the coagulation process.^{13–15} In recent years, mathematical models have emerged as a viable method to model complex processes of water treatment.¹⁶ These models have been very successful in modelling and predicting environmental parameters, and in predicting coagulant quantity as a function of many physicochemical parameters.¹⁷ For example, a study was focused on the use of *Moringa oleifera* seed as a coagulant for the treatment of surface water using response surface methodology (RSM).¹⁸ Four parameters were varied *viz.* stabilisation time, stirring time, stirring speed, and concentration of *Moringa oleifera* seed extract (MOSE).

* Corresponding author: Hichem Tahraoui, PhD
Email: tahraoui.hichem@univ-medea.dz

The model predicts the lowest turbidity of 5.49 NTU with optimal conditions of 120 min settling time, stirring speed of 100 rpm, 10 min stirring time, and 3 g l^{-1} concentration of MOSE. The condition was verified in aftershocks and a turbidity of 5.51 NTU was obtained.¹⁸

Other works reported the effectiveness of three coagulants, CuSO_4 , FeCl_3 , and $\text{CuSO}_4 + \text{FeCl}_3$ (the two salts mixed in 1 : 1 ratio (v/v)) for the treatment of wastewater from the effluent treatment plant of an oil refinery by coagulation and flocculation process.¹⁹ Independent parameters, such as pH and coagulant dosage, were optimised using the response surface methodology with central composite design technique considering final pH, COD reduction, turbidity, TDS, and colour as dependent variables.¹⁹ Results showed that the mixed coagulant $\text{CuSO}_4 + \text{FeCl}_3$ gave better results than CuSO_4 and FeCl_3 individually. $\text{CuSO}_4 + \text{FeCl}_3$ along with adsorption and sweep flocculation, forms tribasic copper chloride (TBCC) as an intermediate in the pH range 5 to 7. TBCC formation improved flocculation by destabilising colloidal and suspended particles due to its structural catalytic, bleaching, and octahedral properties. A maximum reduction in COD (76.77 %), turbidity (89.47 %), TDS (94.16 %), and colour (95.29 %) were observed at pH 7.12 and dosage 0.20 g l^{-1} of $\text{CuSO}_4 + \text{FeCl}_3$ coagulant. The final pH of the solution under optimal conditions for all three coagulants was below the removal limits.¹⁹

Another application by the response surface method (RSM) was to find the optimal combination of coagulant dose and pH relative to the highest removal efficiency of turbidity and dissolved organic carbon (COD).²⁰ The results obtained with polyaluminum chloride (PACl) were compared with those obtained using a conventional coagulant such as alum. Quadratic models developed for both responses (turbidity removal and DOC removal) indicated the optimal conditions for a PAC1 concentration of 0.11 mM at pH 7.4 and alum concentration of 0.15 mM at pH 6.6. The compromise to optimise both responses simultaneously resulted in 91.4 % turbidity removal and 31.2 % DOC removal using PACl, while 86.3 % turbidity and 34.3 % DOC. CODs were removed using alum.²⁰

Another study was conducted to predict both turbidity and removal of dissolved organic matter (DOM) during the coagulation process at Akron Water Treatment Plant (Akron, Ohio, USA) with four different neural network models.²¹ DOM was monitored and characterised using fluorescence spectroscopy and parallel factor analysis (PARAFAC), building on previous research that identified three fluorescence components (C1, C2, and C3). Neural network models were constructed using operational data to predict each of the components of fluorescence and turbidity after coagulation as a function of variable raw water quality and chemical doses. Correlation coefficients between measured and model-predicted values for final turbidity models, C1, C2, and C3, on an invisible test data set were 0.91, 0.95, 0.97, and 0.51, respectively.²¹

Other researchers used extreme machine learning (ELM) coupled with radial base function (RBF) neural networks to predict coagulation doses.²² The coagulation data was divided into two categories based on low and high turbidity. The optimal number of input parameters for low turbidity water

coagulation modelling was found to be 3, while the optimal number of input parameters for turbidity water coagulation modelling were found to be 4. Re-selection of the number of input parameters was necessary, since the alkalinity of the raw water was an important factor in improving the performance of the high turbidity model. The low turbidity model was able to predict coagulant dosage with a correlation coefficient greater than 0.97.²² The high turbidity model was able to predict coagulation dosage with a reasonably acceptable correlation coefficient of at least 0.80.²²

Another study was performed by a hybrid of k -signifies an adaptive neuro-fuzzy inference system (k -means-ANFIS) for the turbidity of settled water prediction and optimal determination of the coagulant dose using historical data at large scale.²³ To construct a well adaptive model to different states of inflow water process, raw water quality data was classified into four groups according to its properties by a k -means clustering technique. The sub-models were developed individually based on each clustered data set. The results revealed that the sub-models constructed by a k -means-ANFIS hybrid perform better than only a single ANFIS model, but also seasonal artificial neural network (ANN) models. The finally completed model composed of sub-models shows more precise and consistent prediction capacity than a single ANFIS model and a single ANN model based on all five evaluation indices.²³

In another report, they used artificial neural network (ANN) and adaptive network-based fuzzy inference system (ANFIS) models to model the dosage of polyaluminum chloride (PAC) of surface water in northern Taiwan.²⁴ Each of them was built based on 819 data sets controlled by the processes. Input parameters included yesterday's PAC dosage, the day before yesterday's PAC dosage, and temperature, turbidity, colour, pH in each raw water, flocculation, sedimentation, and treated water.²⁴ The ANN model was better than the ANFIS model in achieving the optimal prediction model for the optimal dosage of PAC in real time when storm water brought high turbidity to the source water.²⁴

In this work, response surface methodology (RSM), artificial neural networks (ANN), machine vector support (SVM), and adaptive neuro-fuzzy inference system (ANFIS) were selected to predict coagulant doses as a function of physical and chemical parameters, and to compare them to select the most efficient one. Indeed, to our knowledge, the ANN, SVM, and ANFIS models have never been used in this kind of data-based study built from RSM. In addition, such a comparison has not been made so far.

2 Materials and methods

2.1 Data preparation

2.1.1 Jar test

Aluminum salt was used in this study as coagulant. The jar tests were carried out using beakers that were filled with 1 l of water, namely raw waters collected in different localities of the Médéa region, with various turbidity levels, which were adjusted with humic acid. Considering that acidic

and oligotrophic waters often have a brown to yellow “tea” tint,²⁵ in particular due to humic acids (humic acids are one of the most important fractions of humus, thanks to their hydrophilic functions of the carboxylic acid type), they can retain about fifteen times their weight in water, and thus play a fundamental role in water retention and the useful water reserve of a soil.²⁶

After having created turbidity with humic acid, the water takes on a “tea” hue, not to mention that the turbidity contains various mineral and organic matter in suspension or in solution. Dissolved and colloidal matter alone constitutes 60 to 80 % of the organic load (humic acids). If we manage to remove turbidity, we can say that we have removed organic and mineral matter (complexation),^{27–29} because the removal of the colour created by humic acid means the removal of organic and dissolved colloidal organics and micropollutants.³⁰

The samples were mixed at 180 rpm for 2 min after addition of coagulant to provide mixing; the speed was then reduced to 40 rpm for 30 min to ensure flocculation.³¹ The time and speed of mixing were set with an automatic controller. After 45 min of settling, the supernatant from each pot was removed from the port sample and analysed for pH, temperature, turbidity, conductivity, and total alkali metric titre (TAC). The pH, conductivity, temperature, and turbidity were measured. TAC was analysed immediately following the assay method described by *Rodier*.³²

2.2 Database

The data used in this study were those obtained from the analysis of samples purified with different doses of coagulant that were collected during several campaigns conducted in the Médéa region. The database that was built by RSM was used for the other models (ANN, SVM, and ANFIS), in order to be able to compare them.

The dependent variable was the coagulant dose. The independent variables were the physicochemical parameters: pH, temperature, conductivity, turbidity, and total alkalinity titre (TAC).

2.3 Prediction methods

Several methods have been applied to solve problems related to the prediction and modelling of complex non-linear systems. These methods are particularly useful when such systems are difficult to model using conventional methods.³³ In this study, four methods were tested for predicting coagulant doses from physicochemical water parameters, namely RSM, ANN, SVM, and ANFIS.

After constructing the database according to RSM, the database was normalised in the interval $[-1, +1]$ by using the *mapminmax* function (Eq. 1) in MATLAB software.³⁴ It was also divided in two parts for the three other models (ANN, SVM, and ANFIS): 70 % of the dataset for training and the remaining 30 % of the samples, which did not participate

in model learning, were used for validation and prediction performances of the models:^{35,36}

$$x_N = (y_{\max} - y_{\min}) \left(\frac{x - x_{\min}}{x_{\max} - x_{\min}} \right) + y_{\min} \quad (1)$$

where x_N is the data value after normalisation, x_{\max} and x_{\min} are the maximum and minimum of data, respectively; y_{\max} and y_{\min} are taken as -1 and 1 ; x represents the original value.

The correlation coefficient (R), determination coefficient (R^2), adjusted coefficient of determination (R^2_{adj}), root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), error standard prediction (ESP), and error prediction model (EPM) were used to estimate the performances of each model.^{36–41}

2.3.1 RSM

The principle of RSM has been described by *Khuri* and *Cornell*²⁰ as a set of mathematical and statistical methods for evaluating the relationships between a group of independent variables and one or more responses.⁴² To obtain adequate and reliable measures of the responses of interest, the design of the experiment is necessary. Normally, the relationship between the response and the independent variables cannot be well modelled by a linear function.

A model that incorporates curvature is generally necessary to approximate the response in the region close to optimal, and in most cases a second-order model is adequate.⁴³ A central composite design (CCD) is considered, which is a very effective design tool for fitting second-order models.⁴³ This technique is generally used when there are several input parameters affecting the output (response). The CCD was selected for use in this study.⁴³ Two levels, a value of $a = 1$ and 5 repetitions at the centre point were considered. Finally, the input variables were coded in intervals $[-1, +1]$ (Table 1).

The number of experiments is determined according to Eq. (2)⁴²:

$$N = 2^k + 2k + c \quad (2)$$

where k is the number of independent variables, 2^k is number of the factorial experiments, $2k$ is the number of axial experiments, and c is the number of experiments in the centre point.⁴⁴ The model used to predict the responses and describe the relationship between the independent variables was the second-order polynomial (Eq. (3))⁴².

$$Y = B_0 + \sum_i^k B_i X_i + \sum_i^k B_{ii} X_i^2 + \sum_{ij}^k B_{ij} X_i X_j + E \quad (3)$$

where, Y represents the response functions (in our case, it is force compression): B_0 is a constant coefficient; B_i , B_{ii} , and B_{ij} are the coefficients of the linear, quadratic, and interactive terms, respectively.⁴⁵

Table 1 – Coding and real levels for CCD model

Variables	Symbol	Unit	Code Factor Levels		
			-1	0	+1
pH	X_1	–	7.25	8.02	8.78
temperature	X_2	°C	10.65	19.48	28.30
conductivity	X_3	$\mu\text{S cm}^{-1}$	106	988	1871
turbidity	X_4	NTU	0.86	519.93	1039
TAC	X_5	°F	12.00	18.20	24.4

2.3.2 Artificial neural network

In this work, the optimisation of the ANN architecture using MATLAB R2013a software is described by a flowchart shown in Fig. 1.^{36,46,47}

2.3.3 Support vector machine

Support vector machine was developed by Vapnik in the 1990s, and is based on the statistical learning theory (SLT) tools and structural risk minimisation (SRM) theory.⁴⁸ It was mainly used for non-linear classification and regression analysis.⁴⁹ In this approach, a training data set of N points (X_i, Y_i) is considered, with $i = 1, \dots, N$. X represents the inputs of the model, and Y its output. A SVM model takes the form:

$$y(x) = \omega^T \phi(x) + b \quad (4)$$

where $\phi(\cdot): R^n \rightarrow R^m$ is a non-linear function that maps the finite dimensional space entry into a higher dimensional

space that is implicitly created, ω is a weight vector, and b is the bias.^{50,51}

SVM problems are solved using quadratic programming techniques, which however show some drawbacks:⁵²

1. They are practically difficult to use.
2. They are time-consuming.
3. They require large memory and CPU time.

In this article, SVM was used for non-linear modelling using the different functions of the kernel, since there is no guarantee that a given kernel appears more efficient than another one regarding the specific data. From this, it is necessary to optimise different kernel functions and test their performances. These kernel functions can be found in the MATLAB Toolbox R2018a.⁵³

The kernel functions selected in this article were as follows:⁵³

Poly (polynomial)

$$k(X_i, X_j) = (ax^T + c)^d \quad (5)$$

Linear

$$k(X_i, X_j) = x^T y + c \quad (6)$$

Gaussian

$$k(X_i, X_j) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(X_i - X_j)^2}{2\sigma^2}\right\} \quad (7)$$

with $C = \text{BOxConstraint}$, $\varepsilon = \text{epsilon}$, $d = \text{Polynomial order}$, and $\sigma^2 = \text{sigma}$, where d , c , ε , and σ^2 are user-defined kernel parameters.

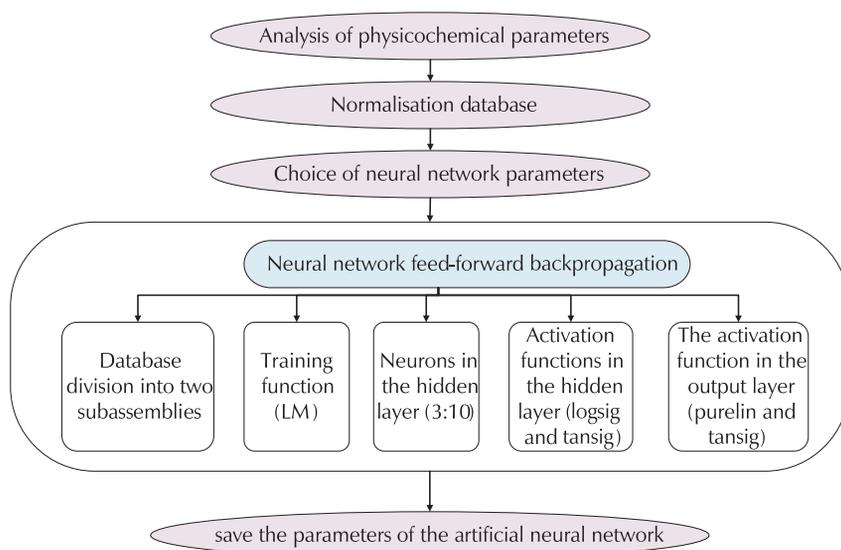


Fig. 1 – Organisation chart for the development and optimisation of the ANN architecture

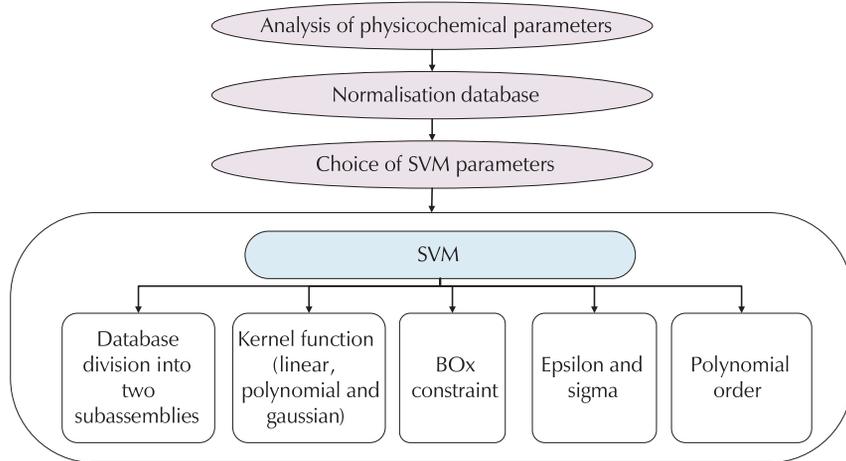


Fig. 2 – Organisation chart for the development and optimisation of the SVM model

2.3.4 Adaptive Neuro-Fuzzy Inference System (ANFIS)

The ANFIS was developed by Jang.⁵⁴ The architecture of ANFIS is shown in Fig. 3. It consists of six key components: input and output data, data division unit, algorithm unit, fuzzy generating system, fuzzy inference system, and adaptive network neuron representing the fuzzy system. After dividing the processing data, each set of training and validation data is assigned to the clusters formed by the algorithm.^{23,24} Each algorithm is formed independently with training and validation together by ANFIS to obtain the optimal fuzzy inference system.²³ In addition, ANFIS is composed of 5 layers – each layer can include different nodes.⁵⁵

Layer 1: Each neuron calculates the degree to which *X* and *Y* inputs belong to the different fuzzy sets (Eqs. 8 and 9).⁵⁶

The parameters inherent to these sets are called premise network parameters. For this system, we have:

$$X_{1,k} = \mu_{A(k)}(X) \quad \text{for } k = 1 \text{ and } 2 \quad (8)$$

$$X_{1,k} = \mu_{B(k-2)}(Y) \quad \text{for } k = 3 \text{ and } 4 \quad (9)$$

Layer 2: is used to calculate the degree of activation of the premises. Each neuron in this layer receives the outputs of the previous fuzzification neurons and calculates its activation. The conjunction of the antecedents is performed with the product operator, which uses the derivability constraint to deploy the learning algorithms. Each node performs a fuzzy T-norm.

$$X_{2,k} = \Pi_{(1,j)} \text{Ant}(2,k)(X) \quad (10)$$

or, *Ant*(2,*k*) indicating the antecedent nodes of node (2,*k*).

Layer 3: Each neuron calculates the normalised degree of truth of a given fuzzy rule. The value obtained represents the contribution of the fuzzy rule to the final result.

$$X_{3,k} = \frac{X_{2,k}}{\sum (2,j) \text{Ant}(2,k)^{X_{2,i}}} \quad (11)$$

Layer 4: Each *i* neuron of this layer is connected to a normalisation neuron corresponding to the initial inputs of the network. The link in this layer fulfils the role of the consequence part of the rules; each node calculates the standard degree of truth of a given fuzzy rule. The value obtained represents the contribution of the fuzzy rule to the final result.

$$X_{4,k} = X_{3,k} [m_{k0} + m_{k1} \cdot X + m_{k2} \cdot Y] \quad (12)$$

where the parameters *m*_{*k0*}, *m*_{*k1*} and *m*_{*k2*}, are called neural fuzzy-system consequential parameters.

Layer 5: This is the output layer comprising a single neuron that provides the output of ANFIS by calculating the sum of the outputs of all output neurons.

$$X_{4,k} = X_{3,k} [m_{k0} + m_{k1} \cdot X + m_{k2} \cdot Y] \quad (13)$$

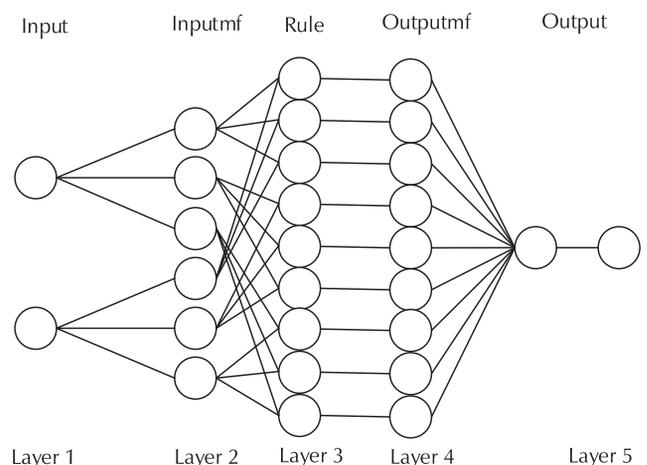


Fig. 3 – Structure of ANFIS

3 Results and discussion

RSM, ANN, SVM, and ANFIS are the four approaches that were used for the prediction of coagulant dose from physicochemical water parameters. All four methods were performed, evaluated, and compared.

3.1 RSM

A statistical analysis by RSM was performed using the “JMP 13” software on the entire database software. This method allowed finding the mathematical relationship (Eq. (14)) between the coagulant doses and the independent variables that corresponded to the physicochemical parameters. Firstly, the equation took into account all 5 variables (Table 2), even those that did not seem to have a significant impact on the dependent variable. Therefore, the relationship obtained was evaluated in order to keep only the independent variables characterised by a high probability value (Table 2).

Table 2 – Values of β_i and characterisation of independent variables of the RSM

i	Term	β_i	Standard error	t ratio	P_r
0	Intercept	29.44685	2.01046	14.64676	$4.40 \cdot 10^{-8}$
1	X_1	-3.88888	1.58898	-2.44740	0.03441
2	X_2	-2.55555	1.58898	-1.60829	0.13884
3	X_3	-11.94440	1.58898	-7.51703	$2.02 \cdot 10^{-5}$
4	X_4	16.38888	1.58898	10.31406	$1.20 \cdot 10^{-6}$
5	X_5	4.16666	1.58898	2.62222	0.02550
6	X_1X_2	5.81250	1.68537	3.44879	0.00623
7	X_1X_3	6.68750	1.68537	3.96796	0.00265
8	X_2X_3	-7.56250	1.68537	-4.48713	0.00116
9	X_1X_4	2.06250	1.68537	1.22376	0.24909
10	X_2X_4	-14.18750	1.68537	-8.41802	$7.52 \cdot 10^{-6}$
11	X_3X_4	-1.31250	1.68537	-0.77875	0.45415
12	X_1X_5	4.68750	1.68537	2.78128	0.01940
13	X_2X_5	-2.56250	1.68537	-1.52043	0.15936
14	X_3X_5	1.56250	1.68537	0.92709	0.37571
15	X_4X_5	6.18750	1.68537	3.67129	0.00430
16	X_1X_1	3.02386	4.29938	0.70332	0.49790
17	X_2X_2	-5.97613	4.29938	-1.38999	0.19469
18	X_3X_3	20.52386	4.29938	4.77367	0.00075
19	X_4X_4	8.52386	4.29938	1.98257	0.07554
20	X_5X_5	-13.47613	4.29938	-3.13443	0.01061

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_1X_2 + \beta_7X_1X_3 + \beta_8X_2X_3 + \beta_9X_1X_4 + \beta_{10}X_2X_4 + \beta_{11}X_3X_4 + \beta_{12}X_1X_5 + \beta_{13}X_2X_5 + \beta_{14}X_3X_5 + \beta_{15}X_4X_5 + \beta_{16}X_1^2 + \beta_{17}X_2^2 + \beta_{18}X_3^2 + \beta_{19}X_4^2 + \beta_{20}X_5^2 \quad (14)$$

Next, only 13 independent variables having high explanatory power for the dependent variable were taken ($P_r < 0.05$); so the relationship could be reduced to Eq. (15):

$$Y = \beta_0 + \beta_1X_1 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_1X_2 + \beta_7X_1X_3 + \beta_8X_2X_3 + \beta_{10}X_2X_4 + \beta_{12}X_1X_5 + \beta_{15}X_4X_5 + \beta_{19}X_4^2 + \beta_{20}X_5^2 \quad (15)$$

The value of the coefficient of determination decreased slightly, but the equation became simpler after removal of the variables with little explanatory power from the dependent variable. This coefficient $R = 0.98691$ means that the model's correlation was moderately positive (Fig. 5). The probability ($P_r < 0.001$) was strictly less than 0.5 %, confirming that the model was significant.

The significance level value p and the F ratio value, which provide a measure of the statistical significance of the regression model, were also determined. A high value of F with a minimum value of p means that the equation is significant.⁵⁷

The proposed model can infer the effect of the factors (X_i), their interactions, and their quadratic effects simultaneously. Table 2 shows the effects of each independent factor, its interaction with the other factors, and its quadratic effect on the coagulant dose (Y).

Indeed, the coefficients of each factor in the model allow to assess the impact of each factor on the response.⁵⁸ Thus, it is evident that changes in turbidity and TAC in the treated effluent increased the dose of the coagulant (Fig. 4). On the other hand, changes in pH and conductivity had a negative effect on the dose of coagulant (Fig. 4).

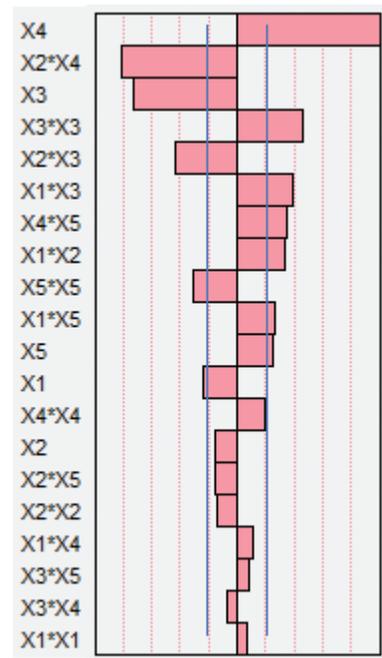


Fig. 4 – Effects of the operating factors and their interactions

Table 3 – RSM performances

	RMSE/mg l ⁻¹	MSE/mg l ⁻¹	ESP/%	EPM/%	MAE/mg l ⁻¹	R	α	β
all	3.82890	14.66053	10.41194	1.47823	3.82890	0.98691	0.97399	0.95615

The interactions and quadratic effects of the factors show that the relationship between the response and the factors is not always linear.⁵⁸ Used at different levels in the coagulation process and when several factors are modified simultaneously, one factor can produce different degrees of response. Thus, interactions between the factors (pH – temperature, pH – conductivity, pH – TAC and turbidity – TAC) showed positive effects on coagulant doses (Fig. 4), whereas the interactions between coagulants (temperature – conductivity and temperature – turbidity) had negative effects on coagulant doses (Fig. 4). In addition, the results showed two factors with interesting quadratic effects, namely, conductivity and TAC, indicating positive and negative quadratic effects, respectively (Fig. 4).

In light of the obtained results, it appears that with each increase in pH (7.25 to 8.78) and temperature (10.65 to 28.3 °C), the dose of coagulant used for treatment decreased from 109.51 to 83.15 mg l⁻¹ for pH, and from 109.15 to 78.28 mg l⁻¹ for temperature. The dose of coagulant also decreased (from 109.51 to 75.24 mg l⁻¹) with increasing conductivity (from 106 μ S to 1243.54 μ S cm⁻¹), and then increased to 85.51 mg l⁻¹ at 1871 μ S cm⁻¹. On the other hand, the results obtained show an increase in the dose of coagulant (from 47.16 to 109.51 mg l⁻¹) when the turbidity increased (from 0.86 NTU to 1039 NTU). Finally, it can be noted that the effect of TAC on the coagulant

dose was not linear. Thus, the dose of coagulant used for effluent treatment increased (from 88.55 to 109.51 mg l⁻¹) for TAC values ranging from 88.55 to 19.71 °F; then decreased (101.88 g l⁻¹) at 24.4 °F. These results confirm the effect of the factors in this study, namely, pH, temperature, conductivity, turbidity, and TAC during the coagulation process.^{8,59–61}

For an optimal chosen coagulant dose equal to 109.50 mg l⁻¹, water with the following properties: pH 7.25, temperature 10.65 °C, conductivity 106 μ S cm⁻¹, turbidity 1039 NTU, and TAC = 19.71 °F, provided 99.96 % effective treatment.

The results of the RSM performance in terms of all errors and in terms of agreement vector values (R, slope: α , and intercept: β) are given in Table 3.

3.2 ANN

In this part, we optimised 3 activation functions with the number of hidden neurons (from 3 to 10). The results of these tests are presented in Table 4.

Table 4 presents the best architectures found. It shows the correlation coefficients and the error for each learning and validation according to the number of neurons in the hidden layer and the network topology. It also shows the activation functions of the hidden layer and the output layer. Architecture 1 (Table 4) seems to be the most relevant ANN model for predicting coagulant doses. Since the correlation coefficient and RMSE were almost equal between architectures 1 and 3, the lowest number of neurons of the hidden layer was chosen in order to have the lowest possible number of parameters.

Table 4 – Performances of the different tested ANN architectures

Nbr	ANN		Activation function		Coefficient of correlation	RMSE
	Neurons per layer	Hidden layer	Output layer	Total	Total	
1	[5-6-1]	logsig	purelin	0.99750	1.7954	
2	[5-9-1]	tansig	purelin	0.99744	1.8132	
3	[5-10-1]	tansig	tansig	0.99751	1.7908	

Once the architecture of the ANN was synthesized, it was validated. The predicted values (estimated values) were compared to the experimental ones in the two stages of

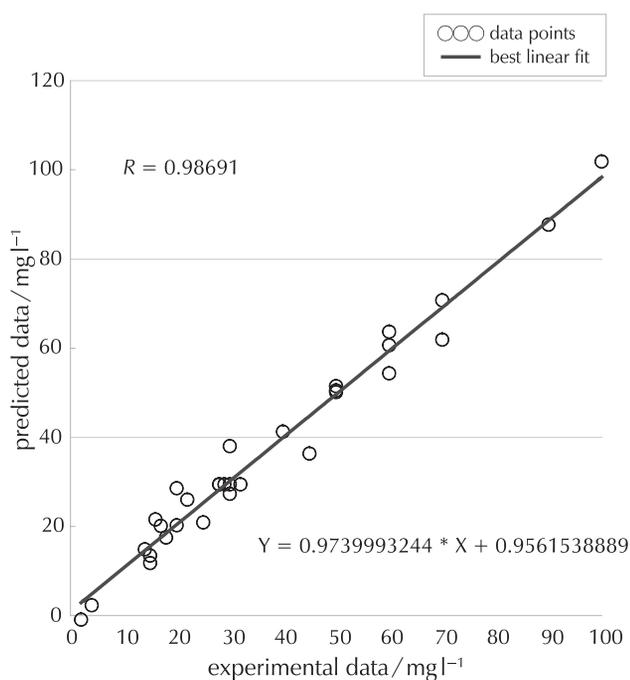


Fig. 5 – Relation between the observed dose of coagulant and those estimated by the RSM model

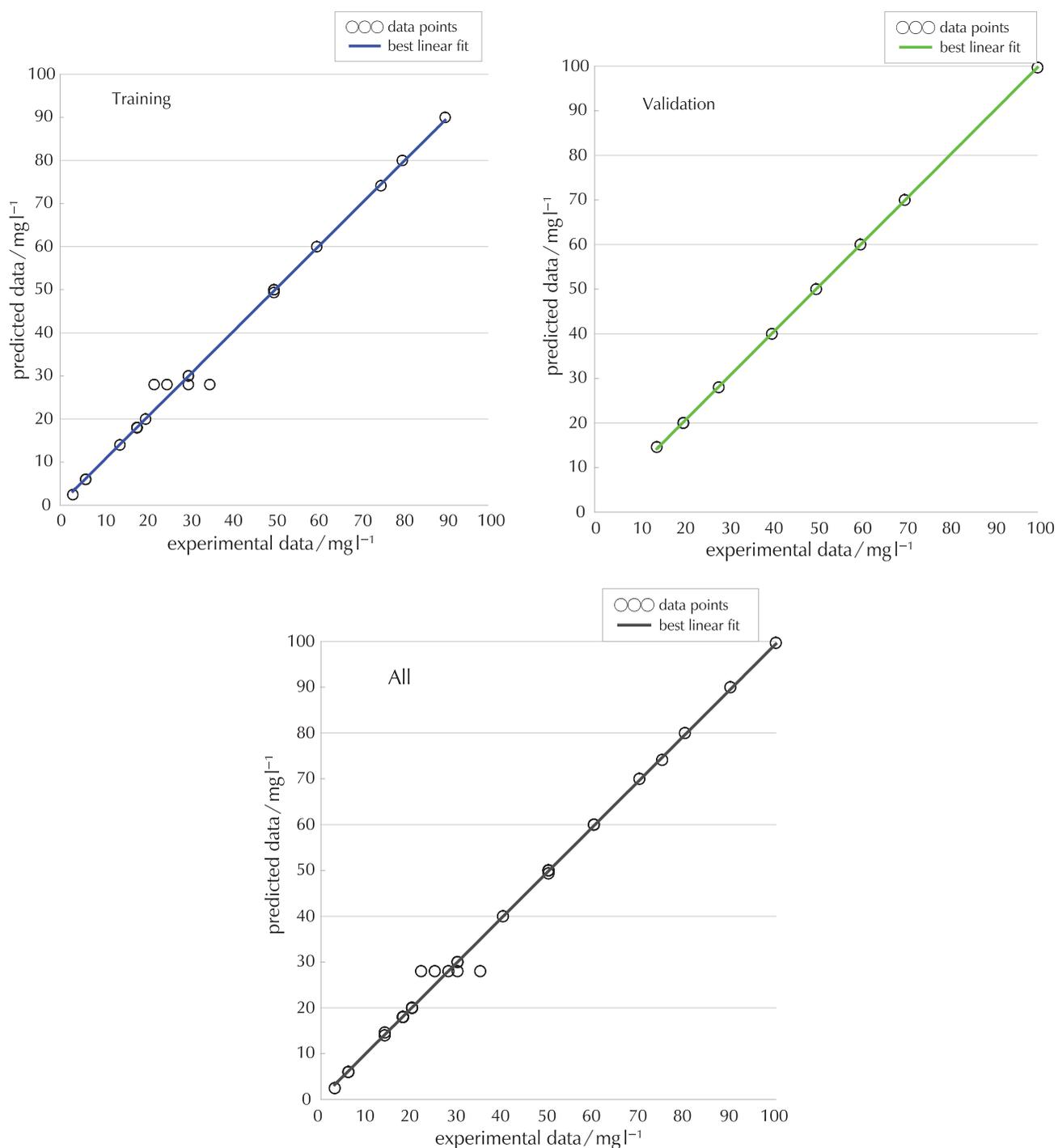


Fig. 6 – Relation between the observed dose of coagulant and those estimated by the ANN model

Table 5 – Performances of prediction using ANN

	RMSE/ mg l^{-1}	MSE/ mg l^{-1}	ESP/%	EPM/%	MAE/ mg l^{-1}	R	α	β
all	1.79535	3.22331	4.83966	0.28386	0.67928	0.99750	0.99403	0.20004
training	2.12626	4.52101	6.25372	0.19098	0.91414	0.99612	0.99011	0.24250
validation	0.22616	0.05115	0.50634	0.00428	0.10519	0.99978	0.99317	0.18521

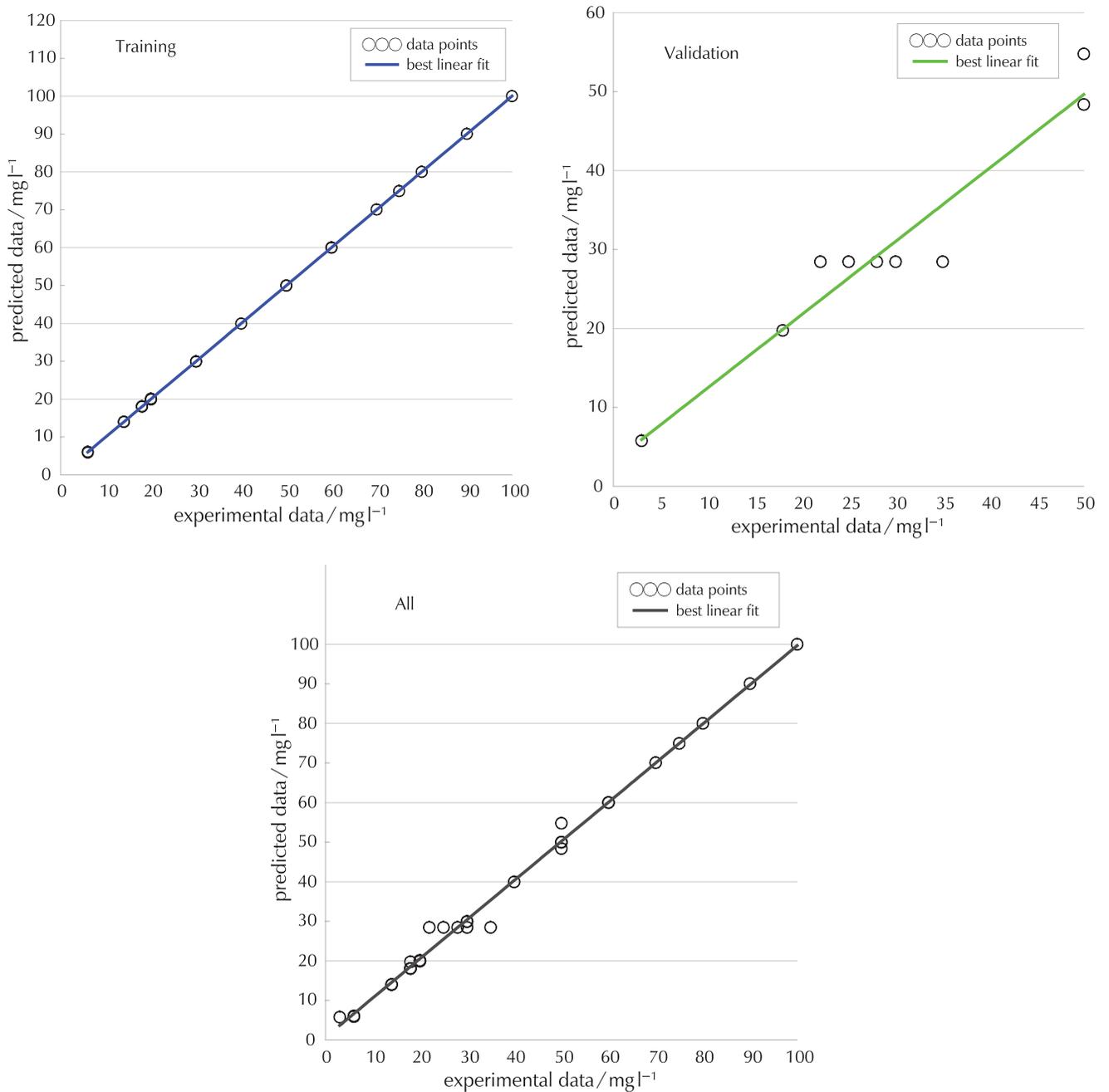


Fig. 7 – Relation between the observed dose of coagulant and those estimated by the SVM model

the learning and validation phases, and the corresponding results are given in Fig. 6. The results were denormalised to actual values in view of comparison with the other models.

The results of the ANN performance in terms of all errors and in terms of agreement vector values (R , slope: α , and y -intercept: β) are given in Table 5.

3.3 SVM

In this part, we optimised 3 kernel functions, each kernel synthesised by parameters that must also be optimised:

The linear function synthesised by BOxConstraint, epsilon and sigma; the Gaussian function synthesised by BOxConstraint, epsilon and sigma; the polynomial function synthesised by BOxConstraint, epsilon and PolynomialOrder, knowing that the PolynomialOrder was optimised from 2 to 5.

Once the result of the learning phase was obtained, it was validated by the validation database. The obtained results (predicted values) were compared with the experimental values in the two stages (learning phase and validation phase) in order to obtain the correlation coefficient and RMSE. The results of these tests are shown in

Table 6 – Performances of the different tested SVM Kernel functions

Kernel function	C	ϵ	σ^2	d	Quantity of support vectors	Coefficients of correlation		RMSE/mg l ⁻¹	
						Total	Total	Total	Total
Linear	846	0.0032	/	/	22	0.66564		19.04920	
Gaussian	479.0010	2.5551 · 10⁻⁴	2.3000	/	22	0.99668		2.08927	
Polynomial	816	0.0052	/	2	22	0.92786		9.54540	
Polynomial	23	0.0035	/	2.5	22	0.93081		9.3885	
Polynomial	102	0.0012	/	3	22	0.98170		4.87540	
Polynomial	614	0.0013	/	3.5	22	0.98311		4.67220	
Polynomial	623	0.001	/	4	22	0.98672		4.17170	
Polynomial	731	5.8782 · 10 ⁻⁴	/	4.5	22	0.99272		3.16830	
Polynomial	855	4.3554 · 10 ⁻⁴	/	5	22	0.99474		0.9947	

Table 7 – Performances of prediction using SVM

	RMSE/mg l ⁻¹	MSE/mg l ⁻¹	ESP/%	EPM/%	MAE/mg l ⁻¹	R	α	β
all	2.08927	4.36505	5.63194	0.58983	0.98625	0.99668	0.99017	0.68643
training	0.06292	0.00395	0.15571	0.01155	0.05477	0.99999	1.00050	-0.01719
validation	3.87627	15.02550	13.36646	0.16488	3.26318	0.96448	0.92988	3.13381

Table 6. It should be noted that the results were denormalised to the actual values to compare to the other models.

It is obvious that the Gaussian function gives better results in terms of correlation coefficient and RMSE; this result has been graphically schematised in Fig. 7.

The results of the SVM performance in terms of all errors and in terms of agreement vector values (R, slope: α and intercept: β) are given in Table 7.

3.4 ANFIS

In this work, we optimised the weighting algorithms, the functions (trimf, trapmf, gbellmf, gaussmf, gauss2mf, pimf, dsigmf and psigmf) for the input and (constant and linear) for the output and many of the nodes for each input to obtain the most accurate result. It should be noted that these algorithms and functions could be found in Matlab R2013 toolbox .

As described previously, two steps were followed for the ANFIS modelling, where two sets of data were used, the training and verification data, respectively. Firstly, the train-

Table 8 – Performances of the different tested ANFIS architectures

ANFIS		Membership function type		Train FIS	Coefficient of correlation	RMSE/mg l ⁻¹
No.	No. of membership functions	Input	Output	Optimisation method	Total	Total
1	[2 2 2 2 2]	trimf	linear	hybrid	0.93691	9.24210
2	[2 2 2 2 2]	trapmf	linear	hybrid	0.95284	7.83380
3	[2 2 2 2 2]	gbellmf	linear	hybrid	0.97137	6.14477
4	[2 2 2 2 2]	gaussmf	linear	hybrid	0.96934	6.35520
5	[2 2 2 2 2]	gauss2mf	linear	hybrid	0.96842	6.45090
6	[2 2 2 2 2]	pimf	linear	hybrid	0.96839	6.45360
7	[2 2 2 2 2]	dsigmf	linear	hybrid	0.96875	6.42420
8	[2 2 2 2 2]	psigmf	linear	hybrid	0.96875	6.42420

ing was performed using the training dataset. Secondly, verification of the dataset was considered to check the accuracy and effectiveness of the ANFIS model.

Two criteria were adopted for optimising the ANFIS model: the number of membership functions assigned to the inputs and output and the RMSE value. In addition, the following four aspects were considered for the ANFIS learn-

ing phase, which were the number of used membership functions, the membership functions type, the overfitting and the training options.

The results of these tests are presented in Table 8. The best results obtained are illustrated in Fig. 8 by denormalising the values to the actual values.

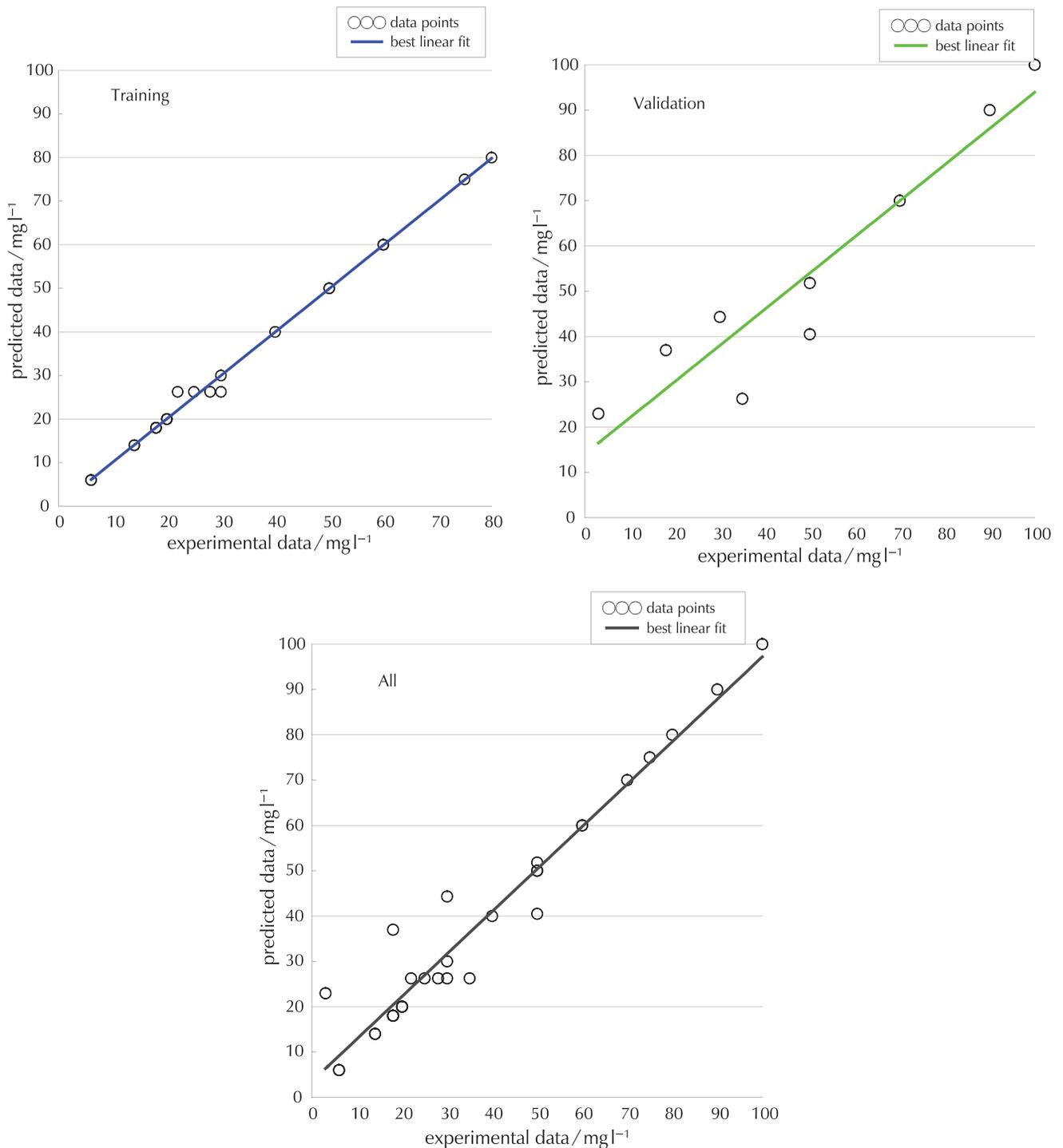


Fig. 8 – Relation between the observed dose of coagulant and those estimated by the ANFIS model

Table 9 – Performances of prediction using ANFIS

	RMSE/mg l ⁻¹	MSE/mg l ⁻¹	ESP/%	EPM/%	MAE/mg l ⁻¹	R	α	β
all	6.14477	37.75828	16.56418	2.81801	2.71994	0.97137	0.93672	3.53417
training	1.29246	1.67045	4.03893	0.09475	0.05000	0.99809	0.99618	0.12218
validation	11.22376	125.97298	22.64886	0.77937	8.14644	0.94454	0.79847	14.07516

Table 10 – Comparison of the predictive performance of the RSM, ANN, SVM and ANFIS models for all data

Model	RMSE/mg l ⁻¹	MSE/mg l ⁻¹	ESP/%	EPM/%	MAE/mg l ⁻¹	R	α	β	R ²	R ² _{adj}	No. of parameters
RSM	3.82890	14.66053	10.41194	1.47823	3.82890	0.98691	0.97399	0.95615	0.97399	0.96878	13
ANN	1.79535	3.22331	4.83966	0.28386	0.67928	0.99750	0.99403	0.20004	0.99500	0.99400	43
SVM	2.08927	4.36505	5.63194	0.58983	0.98625	0.99668	0.99017	0.68643	0.99337	0.99204	22
ANFIS	6.14477	37.75828	16.56418	2.81801	2.71994	0.97137	0.93672	3.53417	0.94355	0.93227	309

The results of the ANFIS performances in terms of all errors and in terms of agreement vector values (R , slope: α , and intercept: β) are given in Table 9.

3.5 Comparison of the models

The comparison of the coefficients of correlation, determination, adjustment and statistical indicators (RMSE, MSE, ESP, EPM, and MAE) obtained by the four models, RSM, ANN, SVM and ANFIS, for all data is presented in Table 10. The coefficients calculated by the model ANN was slightly higher than those obtained by the SVM model, and significantly higher than those obtained by the RSM and the ANFIS models. These results were also confirmed by the statistical indicators, since those given by the ANN model were slightly lower than those obtained by the SVM model, followed by the RSM model and finally the ANFIS model.

The comparison of the different models is illustrated in Fig. 9, confirming the superiority of the ANN and SVM models over the other models (RSM, ANFIS) with a very good superposition of the experimental values and those given by the models. Despite the fact that the ANN and SVM models were almost equal, the ANN model was slightly superior. The correlation coefficient and the statistical indicators of the apprenticeship phase of the ANN model were almost equal to those given by the SVM model with slight superiority of the SVM model (Tables 5 and 7). However, in the validation phase, the correlation coefficient was very high and the statistical indicators were very small for the ANN model ($R = 0.99978$, $RMSE = 0.22616$, $MSE = 0.05115$, $ESP = 0.50634$, $EPM = 0.00428$ and $MAE = 0.10519$) compared to the SVM model ($R = 0.96448$, $RMSE = 3.87627$, $MSE = 15.02550$, $ESP = 13.36646$, $EPM = 0.16488$, and $MAE = 3.26318$) (Tables 5 and 7). However, in terms of reducing economic

costs, the SVM model was superior, since the number of parameters of this model was 22, which is almost half the number of parameters of the ANN model (43 parameters). To reduce the economic costs further, we can also use the RSM model, which remains very useful because of its high coefficients ($R = 0.98691$, $R^2 = 0.97399$, $R^2_{adj} = 0.96878$) compared to the number of parameters, namely 13. In addition, the statistic indicators remained acceptable for the RSM model ($RMSE = 3.82890$, $MSE = 14.66053$, $ESP = 10.441194$, $EPM = 1.47823$, and $MAE = 3.82890$) if related to the number of model parameters.

Finally, the ANFIS model, which was acceptable from the point of view of coefficients and statistical indicators ($R = 0.97137$, $R^2 = 0.94355$, $R^2_{adj} = 0.93227$, $RMSE = 6.14477$, $MSE = 37.75828$, $ESP = 16.56418$, $EPM = 2.81801$, and $MAE = 2.71994$), appeared very expensive because the number of parameters was very high (309 parameters) and results less accurate compared to other models.

3.6 Residues study

Residue analysis is an efficient way to reveal the performances of the optimised model. It consists of the measurement of the absolute or relative error between experimental and predicted values.⁵⁵ Residue analysis methods are mainly graphical analysis methods. Fig. 10 shows the residuals related to each model (RSM, ANN, SVM, and ANFIS). This figure shows that the residuals obtained by the neural network method were less scattered (close to zero) compared to those obtained by the other models. Again, this result shows the accuracy and robustness of the ANN model compared to the other models, and justifies the use of the ANN approach in the prediction of coagulant doses.

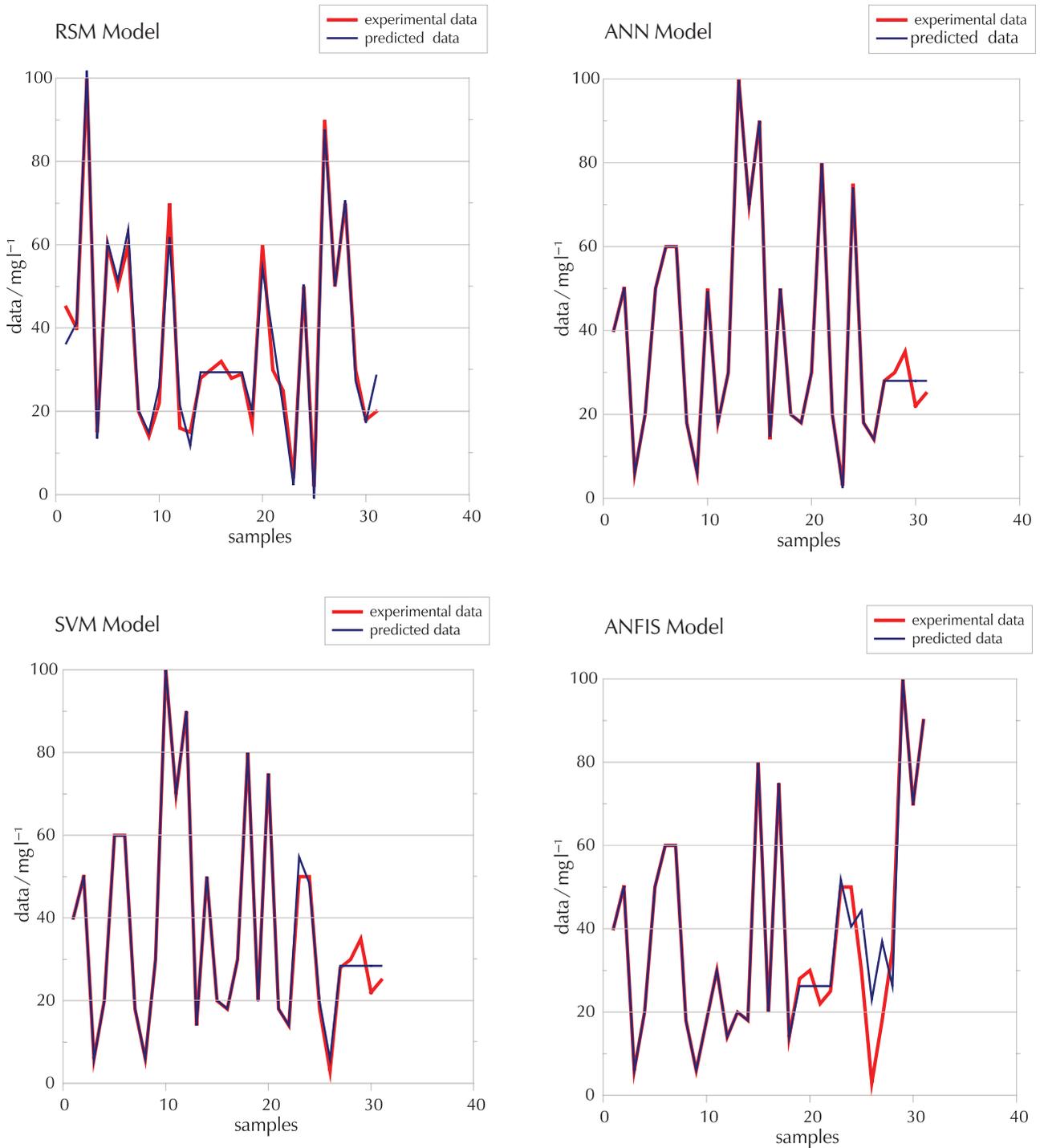


Fig. 9 – Relationship between experimental data and the predicted data of samples using RSM, ANN, SVM, and ANFIS modelling

4 Conclusion

In this study, four mathematical models (RSM, ANN, SVM, and ANFIS) were used to build models capable of predicting the dose of coagulant for removing organic components and turbidity, as well as to reduce the waste of expensive chemicals and operating costs, and achieve drinking water quality objectives. The results obtained from the RSM, ANN, SVM, and ANFIS models proved to be useful and ef-

fective models for predicting the coagulant doses from the point of view of coefficients and statistical indicators. However, the ANN model appeared to be the most effective model, as it led to the highest coefficient ($R^2_{adj} = 0.99400$) and the lowest error indicator (RMSE = 1.79535) compared to the other models. Followed by the SVM model ($R^2_{adj} = 0.99204$, RMSE = 2.08927), the RSM model ($R^2_{adj} = 0.96878$, RMSE = 3.82890), and finally the ANFIS model ($R^2_{adj} = 0.93227$, RMSE = 6.14477). Furthermore,

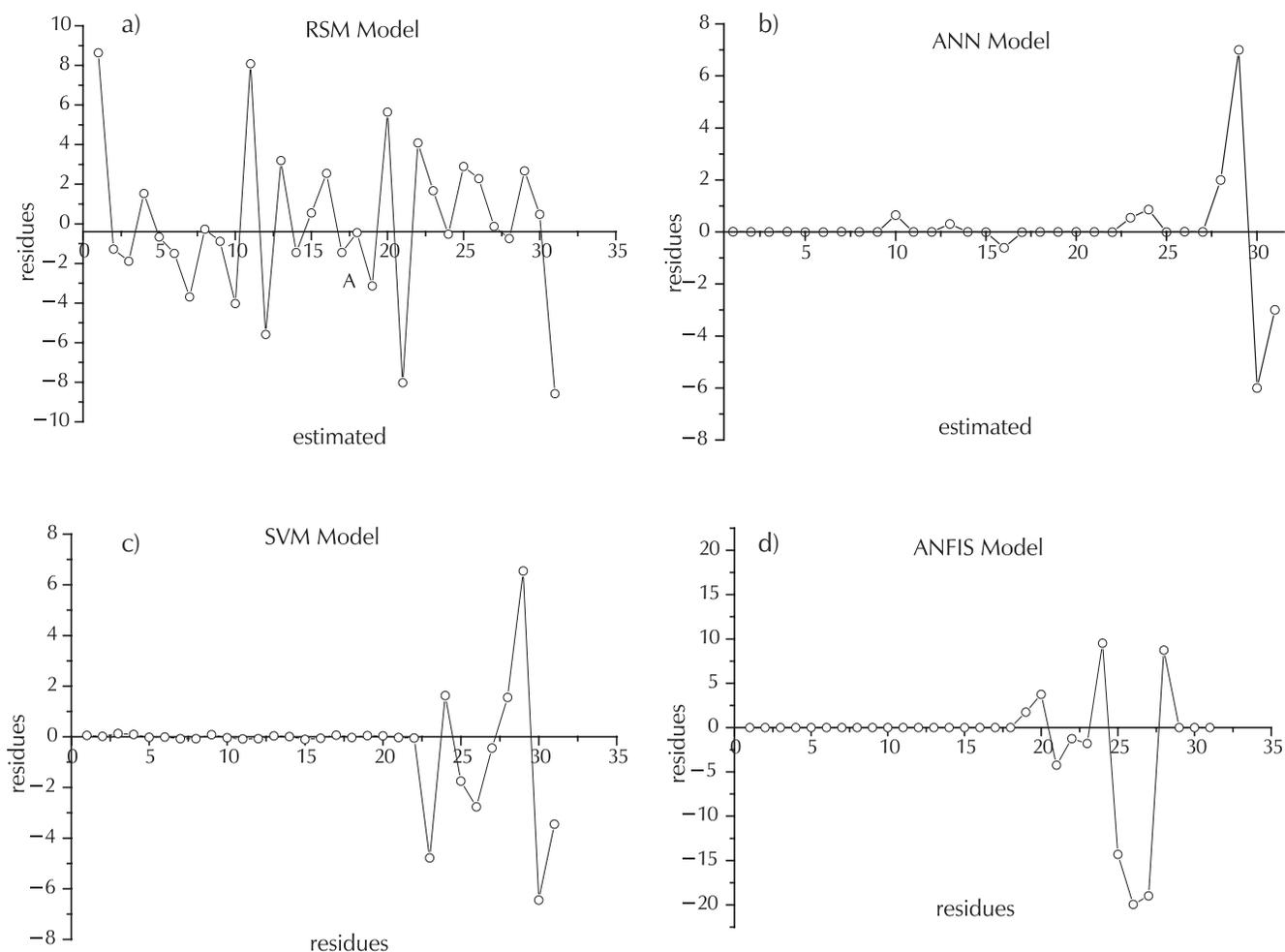


Fig. 10 – Distribution of residuals for the RSM (a), ANN (b), SVM (c), and ANFIS (d) models for all the data

from an economic point of view, only three models appeared relevant: the ANN model which involved 43, the SVM model which was almost similar to the ANN model in terms of coefficients and statistical indicators but involved only 22 parameters, and finally the RSM model, which is very useful in terms of reducing economic costs due to its high coefficient ($R^2_{adj} = 0.96878$) if related to the number of parameters, only 13. In addition, it led to acceptable statistical indicators. Contrarily, the ANFIS model was considered unacceptable, owing to the very high number of parameters, 309.

Literatura References

1. A. T. A. Baptista, P. F. Coldebella, P. H. F. Cardines, R. G. Gomes, M. F. Vieira, R. Bergamasco, A. M. S. Vieira, Coagulation–flocculation process with ultrafiltered saline extract of *Moringa oleifera* for the treatment of surface water, *Chem. Eng. Technol.* **276** (2015) 166–173, doi: <https://doi.org/10.1016/j.cej.2015.04.045>.
2. J. Bratby, Coagulation and flocculation in water and wastewater treatment, IWA publishing, 2016, doi: <https://doi.org/10.2166/9781780407500>.
3. J. K. Edzwald, J. E. Tobiason, Enhanced coagulation: US requirements and a broader view, *Water Sci. Technol.* **40** (1999) 63–70, doi: <https://doi.org/10.2166/wst.1999.0444>.
4. J.-Q. Jiang, Development of coagulation theory and pre-polymerized coagulants for water treatment, *Sep. Purif. Technol.* **30** (2001) 127–141, doi: <https://doi.org/10.1081/SPM-100102986>.
5. B. Bolto, J. Gregory, Organic polyelectrolytes in water treatment, *Water Res.* **41** (2007) 2301–2324, doi: <https://doi.org/10.1016/j.watres.2007.03.012>.
6. R. Jiao, R. Fabris, C.W. Chow, M. Drikas, J. van Leeuwen, D. Wang, Roles of coagulant species and mechanisms on floc characteristics and filterability, *Chemosphere* **150** (2016) 211–218, doi: <https://doi.org/10.1016/j.chemosphere.2016.02.030>.
7. Z. Yang, B. Gao, Q. Yue, Coagulation performance and residual aluminum speciation of $Al_2(SO_4)_3$ and polyaluminum chloride (PAC) in Yellow River water treatment, *Chem. Eng. Technol.* **165** (2010) 122–132, doi: <https://doi.org/10.1016/j.cej.2010.08.076>.
8. B. Lakhdari, Effet de la coagulation-flocculation sur la qualité des eaux épurées de la STEP de Ain El Houtz, 2011, url: <http://dSPACE.univ-tlemcen.dz/handle/112/132>.
9. M. Sillanpää, A. Matilainen, NOM removal by coagulation, Natural Organic Matter in Water: Characterization and Treatment Methods. **55** (2015) 55–80, doi: <https://doi.org/10.1016/B978-0-12-801503-2.00003-3>.
10. D.-S. Joo, D.-J. Choi, H. Park, The effects of data preprocess-

- ing in the determination of coagulant dosing rate, *Water Res.* **34** (2000) 3295–3302, doi: [https://doi.org/10.1016/S0043-1354\(00\)00067-1](https://doi.org/10.1016/S0043-1354(00)00067-1).
11. A. Robenson, S. A. Shukor, N. Aziz, Development of process inverse neural network model to determine the required alum dosage at Segama water treatment plant Sabah, Malaysia, *Comput. Aided Chem. Eng.* **27** (2009) 525–530, doi: [https://doi.org/10.1016/S1570-7946\(09\)70308-6](https://doi.org/10.1016/S1570-7946(09)70308-6).
 12. O. Bello, Y. Hamam, K. Djouani, Control of a coagulation chemical dosing unit for water treatment plants using MMPC based on fuzzy weighting, *J. Water Process. Eng.* **4** (2014) 34–46, doi: <https://doi.org/10.1016/j.jwpe.2014.08.007>.
 13. A. Adgar, C. S. Cox, C. A. Jones, Enhancement of coagulation control using the streaming current detector, *Bioproc. Biosyst. Eng.* **27** (2005) 349–357, doi: <https://doi.org/10.1007/s00449-005-0413-5>.
 14. B. Hua, A fuzzy logic based coagulant real time control scheme for water purification system, in: 2009 International Conference on Mechatronics and Automation, IEEE, 2009; pp. 2123–2127, doi: <https://doi.org/10.1109/ICMA.2009.5246439>.
 15. P. S. Paz, J. P. V. Ocampo, Automation and control of a water treatment plant process, in: Proceedings of 4th Mercosur Congress on Process System Engineering, Mecosur, 2009, p. 2009.
 16. J. S. Torrecilla, M. L. Mena, P. Yáñez-Sedeño, J. García, Application of artificial neural network to the determination of phenolic compounds in olive oil mill wastewater, *J. Food Eng.* **81** (2007) 544–552, doi: <https://doi.org/10.1016/j.jfoodeng.2006.12.003>.
 17. P. Perez, A. Trier, Prediction of NO and NO₂ concentrations near a street with heavy traffic in Santiago, Chile, *Atmos. Environ.* **35** (2001) 1783–1789, doi: [https://doi.org/10.1016/S1352-2310\(00\)00288-0](https://doi.org/10.1016/S1352-2310(00)00288-0).
 18. O. A. Adesina, F. Abdulkareem, A. S. Yusuff, M. Lala, A. Okewale, Response surface methodology approach to optimization of process parameter for coagulation process of surface water using Moringa oleifera seed, *S. Afr. J. Chem. Eng.* **28** (2019) 46–51, doi: <https://doi.org/10.1016/j.sajce.2019.02.002>.
 19. B. Singh, P. Kumar, Pre-treatment of petroleum refinery wastewater by coagulation and flocculation using mixed coagulant: Optimization of process parameters using response surface methodology (RSM), *J. Water Process. Eng.* **36** (2020) 101317, doi: <https://doi.org/10.1016/j.jwpe.2020.101317>.
 20. T. K. Trinh, L. S. Kang, Response surface methodological approach to optimize the coagulation–flocculation process in drinking water treatment, *Chem. Eng. Res. Des.* **89** (2011) 1126–1135, doi: <https://doi.org/10.1016/j.cherd.2010.12.004>.
 21. M. J. Kennedy, A. H. Gandomi, C. M. Miller, Coagulation modeling using artificial neural networks to predict both turbidity and DOM-PARAFAC component removal, *J. Environ. Chem. Eng.* **3** (2015) 2829–2838, doi: <https://doi.org/10.1016/j.jece.2015.10.010>.
 22. C. Jayaweera, M. Othman, N. Aziz, Improved predictive capability of coagulation process by extreme learning machine with radial basis function, *J. Water Process Eng.* **32** (2019) 100977, doi: <https://doi.org/10.1016/j.jwpe.2019.100977>.
 23. C. M. Kim, M. Parnichkun, Prediction of settled water turbidity and optimal coagulant dosage in drinking water treatment plant using a hybrid model of k-means clustering and adaptive neuro-fuzzy inference system, *Appl. Water Sci.* **7** (2017) 3885–3902, doi: <https://doi.org/10.1007/s13201-017-0541-5>.
 24. G.-D. Wu, S.-L. Lo, Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system, *Eng. Appl. Artif. Intell.* **21** (2008) 1189–1195, doi: <https://doi.org/10.1016/j.engappai.2008.03.015>.
 25. G. Rädlinger, K. G. Heumann, Determination of halogen species of humic substances using HPLC/ICP-MS coupling, *Fresenius' J. Anal. Chem.* **359** (4) (1997) 430–433, doi: <https://doi.org/10.1007/s002160050604>.
 26. H. N. P. Dayarathne, M. J. Angove, R. Aryal, H. Abuel-Naga, B. Mainali, Removal of natural organic matter from source water: Review on coagulants, dual coagulation, alternative coagulants, and mechanisms, *J. Water Process. Eng.* (2020) 101820 (in press), doi: <https://doi.org/10.1016/j.jwpe.2020.101820>.
 27. J. J. Milledge, S. Heaven, A review of the harvesting of micro-algae for biofuel production, *Reviews in Rev. Environ. Sci. Biotechnol.* **12** (2) (2013) 165–178, doi: <https://doi.org/10.1007/s11157-012-9301-z>.
 28. E. Lefebvre, B. Legube, Coagulation par Fe (III) de substances humiques extraites d'eaux de surface: Effet du pH et de la concentration en substances humiques, *Water Res.* **24** (5) (1990) 591–606, doi: [https://doi.org/10.1016/0043-1354\(90\)90192-9](https://doi.org/10.1016/0043-1354(90)90192-9).
 29. E. M. Thurman, *Organic geochemistry of natural waters*. Springer Science & Business Media, Vol. 2, 2012.
 30. F. Zidane, K. Cheggari, J.-F. Blais, P. Drogui, J. Bensaid, S. I. Ahmed, Contribution à l'étude de l'effet de la coagulation avant chloration sur la formation des trihalométhanes (THM) et composés organohalogénés (COX) dans les eaux alimentant la ville de Casablanca au Maroc, *Can. J. Civ. Eng.* **37** (2010) 1149–1156, doi: <https://doi.org/10.1139/L10-065>.
 31. H. Yamamura, E.U. Putri, T. Kawakami, A. Suzuki, H.D. Arjesyady, T. Ishii, Dosage optimization of polyaluminum chloride by the application of convolutional neural network to the floc images captured in jar tests, *Sep. Purif. Technol.* **237** (2020) 116467, doi: <https://doi.org/10.1016/j.seppur.2019.116467>.
 32. J. Rodier, B. Legube, N. Merlet, R. Brunet, L'analyse de l'eau - 9e éd.: Eaux naturelles, eaux résiduaires, eau de mer, Dunod, 2009, url: <https://books.google.dz/books?id=qUEGsUBZkLOC>.
 33. D. Jamin, Recherche du boson de Higgs du Modèle Standard dans le canal de désintégration ZH-> nu nu bb sur le collisionneur Tevatron dans l'expérience D0. Développement d'une méthode d'étiquetage des jets de quark b avec des muons de basses impulsions transverses, 2010.
 34. R. Yacef, A. Mellit, S. Belaid, Z. Şen, New combined models for estimating daily global solar radiation from measured air temperature in semi-arid climates: application in Ghardaïa, Algeria, *Energ. Conver. Manag.* **79** (2014) 606–615, doi: <https://doi.org/10.1016/j.enconman.2013.12.057>.
 35. H. El Badaoui, A. Abdallaoui, I. Manssouri, L. Lancelot, Elaboration de modèles mathématiques stochastiques pour la prédiction des teneurs en métaux lourds des eaux superficielles en utilisant les réseaux de neurones artificiels et la régression linéaire multiple, *Int. J. Environ. Res. Public Health.* **3** (2012) 31–36.
 36. H. Tahraoui, A. E. Belhadj, A. E. Hamitouche, Prediction of the Bicarbonate Amount in Drinking Water in the Region of Médéa Using Artificial Neural Network Modelling, *Kem. Ind.* **69** (2020) 595–602, doi: <https://doi.org/10.15255/KUI.2020.002>.
 37. D. A. Belsley, E. Kuh, R. E. Welsch, *Regression diagnostics: identifying influential data and sources of collinearity*, Wiley, 1980, doi: <https://doi.org/10.1002/0471725153>.

38. S. H. Hong, M. W. Lee, D. S. Lee, J. M. Park, Monitoring of sequencing batch reactor for nitrogen and phosphorus removal using neural networks, *Biochem. Eng. J.* **35** (2007) 365–370, doi: <https://doi.org/10.1016/j.bej.2007.01.033>.
39. O. R. Dolling, E. A. Varas, Artificial neural networks for streamflow prediction, *J. Hydraul. Res.* **40** (2002) 547–554, doi: <https://doi.org/10.1080/00221680209499899>.
40. I. Manssouri, M. Manssouri, B. El Kihel, Fault Detection by K-Nn Algorithm and Mlp Neural Networks in A Distillation Column: Comparative Study, *J. Inform. Intell. Knowledge.* **3** (2011) 201.
41. I. Manssouri, A. El Hmaid, T. E. Manssouri, B. El Mounni, Prediction levels of heavy metals (Zn, Cu and Mn) in current Holocene deposits of the eastern part of the Mediterranean Moroccan margin (Alboran Sea), *IOSR J. Comput. Eng.* **16** (2014) 117–123, doi: <https://doi.org/10.9790/0661-1618117123>.
42. A. Hammoudi, K. Moussaceb, C. Belebchouche, F. Dahmoune, Comparison of artificial neural network (ANN) and response surface methodology (RSM) prediction in compressive strength of recycled concrete aggregates, *Constr Build Mater.* **209** (2019) 425–436, doi: <https://doi.org/10.1016/j.conbuildmat.2019.03.119>.
43. D. C. Montgomery, Design and analysis of experiments, 8th Ed, Wiley, Hoboken, NJ, 2013.
44. E. Ghafari, H. Costa, E. Júlio, RSM-based model to predict the performance of self-compacting UHPC reinforced with hybrid steel micro-fibers, *Constr Build Mater.* **66** (2014) 375–383, doi: <https://doi.org/10.1016/j.conbuildmat.2014.05.064>.
45. F. Dahmoune, H. Remini, S. Dairi, O. Aoun, K. Moussi, N. Bouaoudia-Madi, N. Adjeroud, N. Kadri, K. Lefsih, L. Boughani, L. Mouni, B. Nayak, K. Madani, Ultrasound assisted extraction of phenolic compounds from *P. lentiscus* L. leaves: Comparative study of artificial neural network (ANN) versus degree of experiment for prediction ability of phenolic compounds recovery, *Ind. Crop. Prod.* **77** (2015) 251–261, doi: <https://doi.org/10.1016/j.indcrop.2015.08.062>.
46. H. Ousmana, A. E. Hmaid, M. Berrada, B. Damnati, I. Etabaai, A. Essahlaoui, Development of a Neural Network approach for predicting nitrate and sulfate concentration in three lakes: Ifrah, Iffer and Afourgagh, Middle Atlas Morocco, *Mor. J. Chem.* **6** (2) (2018) 245–255, doi: <https://doi.org/10.48317/IMIST.PRSM/morjchem-v6i2.5939>.
47. M. Sediri, S. Hanini, H. Cherifi, M. Laidi, S. A. Turki, Dynamic Adsorption Modelling of P-nitrophenol in Aqueous Solution Using Artificial Neural Network, *J. Mater. Environ. Sci.* **8** (7) (2017) 2282–2287, url: https://www.jmaterenvironsci.com/Document/vol8/vol8_N7/245-JMES-Sediri.pdf.
48. O. Bousquet, S. Boucheron, G. Lugosi, Introduction to statistical learning theory, in: Summer School on Machine Learning, Springer, 2003: pp. 169–207, doi: https://doi.org/10.1007/978-3-540-28650-9_8.
49. E. Poubasheer, S. Riahi, M. R. Ganjali, P. Norouzi, Application of genetic algorithm-support vector machine (GA-SVM) for prediction of BK-channels activity, *Eur. J. Med. Chem.* **44** (2009) 5023–5028, doi: <https://doi.org/10.1016/j.ejmech.2009.09.006>.
50. S. J. AK, V. J. PL, Least squares support vector machines, World scientific, 2002.
51. V. Vapnik, S. E. Golowich, A. J. Smola, Support vector method for function approximation, regression estimation and signal processing, in: *Adv. Neural. Inform. Process Sys.* 1997, pp. 281–287.
52. X. Wang, J. Chen, C. Liu, F. Pan, Hybrid modeling of penicillin fermentation process based on least square support vector machine, *Chem. Eng. Res. Des.* **88** (2010) 415–420, doi: <https://doi.org/10.1016/j.cherd.2009.08.010>.
53. H. Benimam, C. Si-Moussa, M. Laidi, S. Hanini, Modeling the activity coefficient at infinite dilution of water in ionic liquids using artificial neural networks and support vector machines, *Neural. Comput. Appl.* **32** (2019) 8635–8653, doi: <https://doi.org/10.1007/s00521-019-04356-w>.
54. J.-S. Jang, ANFIS: adaptive-network-based fuzzy inference system, *IEEE Trans. Syst. Man Cyber.* **23** (1993) 665–685, doi: <https://doi.org/10.1109/21.256541>.
55. M. Bouhedda, S. Lefnaoui, S. Rebouh, M. M. Yahoum, Predictive model based on Adaptive euro-Fuzzy Inference System for estimation of Cephalexin adsorption on the Octenyl Succinic Anhydride starch, *Chemom. Intell. Lab. Syst.* **193** (2019) 103843, doi: <https://doi.org/10.1016/j.chemo-lab.2019.103843>.
56. R. Adeline, R. Gouriveau, N. Zerhouni, Pronostic de défaillances: Maîtrise de l’erreur de prédiction., in: 2008.
57. M. Naoual, A. Abdelaziz, E. H. Abdellah, Use of Artificial Neural Networks Type MLP for the Prediction of Phosphorus Level from the Physicochemical Parameters of Sediments, (n.d.) 10.
58. S. Lefnaoui, N. Moulai-Mostefa, Investigation and optimization of formulation factors of a hydrogel network based on kappa carrageenan–pregelatinized starch blend using an experimental design, *Colloids and Surfaces A: Colloids Surf. A Physicochem. Eng. Asp.* **458** (2014) 117–125, doi: <https://doi.org/10.1016/j.colsurfa.2014.01.007>.
59. R. Desjardins, Le traitement des eaux, Presses inter Polytechnique, 1997.
60. B. Lakhdari, Effet de la coagulation-floculation sur la qualité des eaux épurées de la STEP de Ain El Houtz, 2011.
61. F. Sekiou, Effets de la nature des particules solides dispersées et des conditions de formation du floc sur l’efficacité de la floculation, Mémoires de Magister. (2014).
62. S. Bouranene, N. Sedira, P. Fievet, N. Attia, Treatment of paint wastewater by coagulation process, *Filtr. Sep.* **52** (2015) 42–45, doi: [https://doi.org/10.1016/S0015-1882\(15\)30184-1](https://doi.org/10.1016/S0015-1882(15)30184-1).

SAŽETAK

Optimizacija i predviđanje doze koagulantna za uklanjanje organskih mikrozagađivala na temelju podataka o zamućenju

Hichem Tahraoui,^{a} Abd-Elmouneïm Belhadj,^a Nassim Moula,^b Saliha Bouranene,^c and Abdeltif Amrane^d*

Četiri različita matematička modela primijenjena su za predviđanje doze koagulantna u svrhu uklanjanja zamućenja: model odzivne površine (RSM), umjetna neuronska mreža (ANN), model potpornih vektora (SVM) i model prilagodljivog sustava neizrazitog zaključivanja zasnovanog na neuronskoj mreži (ANFIS). Rezultati su pokazali da svi modeli točno opisuju eksperimentalne podatke, iako je ANN model bio nešto bolji. SVM model imao je sličnu podudarnost kao i ANN model no razlika je bila u validaciji modela gdje je ANN model ostvario vrlo visoke vrijednosti koeficijenta korelacije te niske vrijednosti statističkih pokazatelja. No s ekonomskog gledišta, SVM model je prikladniji od ANN modela, budući da je njegov broj parametara 22 što je gotovo upola manje od broja parametara ANN modela (43 parametra), dok su rezultati bili slični. Dodatno smanjenje ekonomskih troškova može se ostvariti primjenom RSM modela koji je ostvario visoke vrijednosti koeficijenata s obzirom na svega 13 parametara. Uz to, RSM model imao je prihvatljive statističke pokazatelje.

Ključne riječi

Koagulacija, fizikalno-kemijska analiza, metodologija odzivnih površina, umjetne neuronske mreže, metoda potpornih vektora, ANFIS

^a *Laboratory of Biomaterials and Transport Phenomena (LBMPT), Nouveau Pôle Urbain, University of Médéa, 26 000 Médéa, Alžir*

^b *Fundamental and Applied Research in Animal and Health (FARAH), Department of Veterinary Management of Animal Resources, Faculty of Veterinary Medicine, University of Liege, Liege 4000, Belgija*

^c *University of Souk Ahras, Dept. Process Engineering, STEE Lab., Rue d'annaba, BP 1553, 41000 Souk-Ahras, Alžir*

^d *University of Rennes, Ecole Nationale Supérieure de Chimie de Rennes, CNRS, ISCR – UMR6226, F-35 000 Rennes, Francuska*

Izvorni znanstveni rad
Prispjelo 8. siječnja 2021.
Prihvaćeno 18. ožujka 2021.