Desalination and Water Treatment
www.deswater.com

doi: 10.1080/19443994.2013.803327

52 (2014) 4079–4088
June

Taylor & Francis
Taylor & Francis Group

A decision support tool for technical processes optimization in drinking water treatment

Bouchra Lamrini[a,*], El Khadir Lakhal[a], Marie Véronique Le Lann[b,c]

[a]*Laboratoire d'Automatique, de l'Environnement et des Procédés de Transferts, URAC, Faculté des Sciences Semlalia, Université Cadi Ayyad, P.O. Box: 2390, 40000 Marrakech, Morocco*
Tel. +33 05 61 33 69 47; Fax: +33 05 61 33 69 36; email: blamrini09@gmail.com
[b]*CNRS, LAAS, 7 Avenue du Colonel Roche, F-31400, Toulouse, France*
[c]*Université de Toulouse, INSA, LAAS, F-31400, Toulouse, France*

Received 14 December 2011; Accepted 29 April 2013

ABSTRACT

In water treatment, the technical processes study aims generally to deal with problems that natural processes are unable or only inadequate to perform. The technical systems aim for a good control of process and therefore a good stability. This is the case of coagulation process in drinking water treatment by removing suspended particles. It requires a good knowledge of raw water characteristics to ensure adequate choice of the coagulant rate. Without the adequate coagulant, this method is not effective. The good coagulation control is therefore essential to guarantee the reliability of the water treatment and the final quality of water produced. This paper presents a neural approach in combination with a fuzzy methodology to study the impact of raw water characteristics on the coagulation process control. Using the concepts of evolutionary algorithms, we developed a decision support tool using fault detection, data validation-reconstruction, and predictive control methods to predict the optimum *coagulant dosage* to be used in a drinking water treatment plant. Simulation results using experimental data stemming from four treatment plants show the reliability of this system to optimize one of critical processes in drinking water treatment.

Keywords: Coagulation process; Artificial neural networks; Fuzzy logic; Fault detection; Data validation; Data reconstruction

1. Introduction

The coagulation process is a major and critical step during the surface water clarification [1]. The aim of applying coagulation treatment is to remove the colloidal suspension present in the raw water, e.g. to destabilize the particle charges by neutralizing the forces that keep them apart. Coagulants added provide positive electric charges to reduce negative charges (zeta potential) of the colloids. As a result, the particles collide to form larger particles (flocs). Rapid mixing is required to disperse the coagulant throughout the liquid. Care must be taken not to overdose the coagulants as this can cause a complete charge reversal and re-stabilize the colloid complex. The jar-test is a traditional laboratory procedure [1,2] that simulates coagulation-flocculation with different chemical doses. The purpose of the procedure is to estimate the optimal coagulant dosing required achieving certain water

*Corresponding author.

quality goals. Samples of raw water to be treated are placed in several jars, various amounts of chemicals are added to each jar, they are stirred, and the settling of solids is observed. The lowest dose of chemicals that provides satisfactory settling is the dose used to treat the water. Disadvantages associated with this procedure are the necessity to rely on manual intervention and the resulting problems of excess or an insufficient coagulant, particularly during periods of fast variations of raw water characteristics. In more recent years, a device called streaming current detector (SCD) has become widely used in water treatment as a method that offers online monitoring, measuring, and control functions for the coagulation process [3]. It was first introduced as a new technology in 1966. It has been experimentally demonstrated that a correlation exists between the SCD output and the measured zeta potentials [4]. However, drawbacks associated with this device are the lack of an exact quantitative model explaining its functioning, and its operation cost and limited efficiency for certain types of raw water quality (particularly when pH>/8).

Good control of this process is essential for the maintenance of satisfactory treated water quality and economic plant operation. Some variables can be measured online with simple physical sensors. In contrast, access to certain information requires long and expensive laboratory analysis, which cannot be executed online. Nevertheless, the possession of this information is generally a key point for control and monitoring of these processes. The development of such a model is still a problem under study: it needs to describe plainly the interactions between transfer mechanisms, chemical, and biological kinetics in a misunderstood environment. Given that the coagulation process is relatively complex and strongly nonlinear, an artificial neural network can be used to build a model for this type of process to replace the traditional methods, e.g. the jar-test and the SCD. Moreover, as we have not an explicit knowledge of the coagulation process, it is very difficult to model it in a determinist way. We are therefore interested to develop a decision support tool (DST) based on artificial neural networks, which describes intrinsically nonlinear relationships between raw water characteristics (such as the turbidity (TUR), the total suspended solids, the temperature (T), the pH, the conductivity (COND), the dissolved oxygen (DO), etc.) and the optimal coagulant dosage. The neural approach was the object of several works showing their efficiency and their potential for coagulation control [5–7]. It offers the advantage of very short computational times and it is likely to describe some nonlinear relationships between inputs and outputs system. Given the strong evolution of the raw water characteristics, an important property for such system is indeed the robustness with regard to the sensors failings or to the unexpected raw water characteristics, owing to accidental pollution for example. In this context, and in combination with the neural approach, we applied an expertise and fuzzy classification technique to identify the functional states of a plant towards the operating conditions. This technique has the advantage to aggregate expert knowledge to information stemming from process and its environment. In this way, we proposed an optimization method [8] that takes four models into account to build this DST (Fig. 1):

(1) FSI (functional states identification) classification model to identify the functional states (normal and abnormal) of treatment plant investigated. LAMDA (Learning Algorithm for Multivariate Data Analysis) [9] technique is used to detect these states.

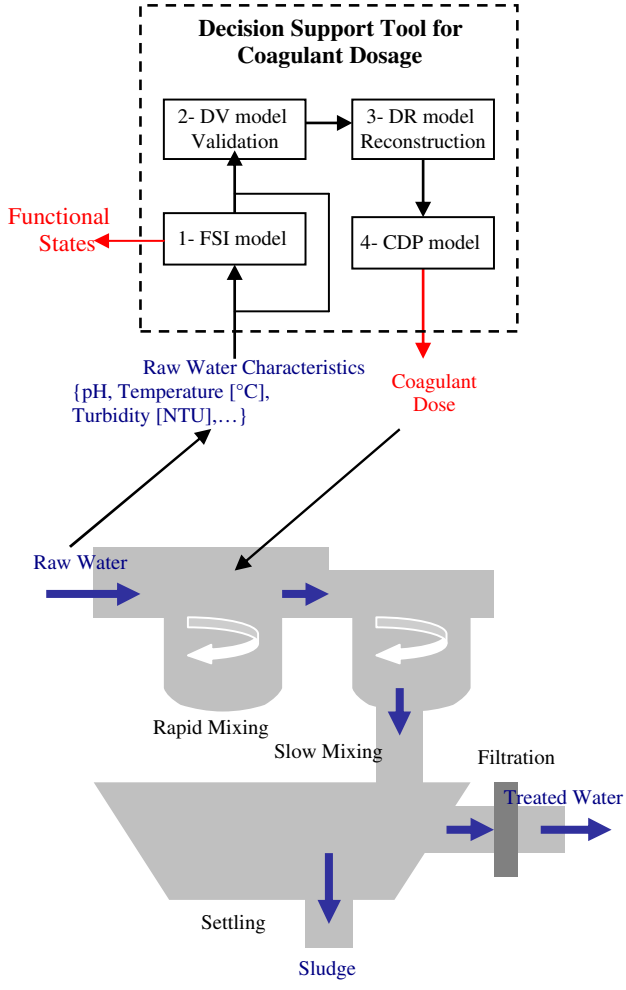

Fig. 1. The DST developed to predict coagulant dosage.

(2) DV (data validation) neural model to validate the invalid data and those stemming from abnormal states detected by LAMDA algorithm. Self-Organizing Maps of Kohonen (SOM) [10,11] are used to develop this model.
(3) DR (data reconstruction) neural model for missing DR built by means of SOM algorithm.
(4) CDP (coagulant dose prediction) neural model using a multilayer perceptron (MLP) [11] to predict the optimal coagulant dose to use in the plant.

For each modeling step, the methodological development and validation results by means of LAMDA technique, SOM, and MLP algorithms have been published in [12–14]. Experimental data to validate this part of work are stemming from *Rocade* plant located in Marrakech city of Morocco. In this paper, we present the prediction results using experimental data stemming from three other plants of drinking water treatment located in Tensift-Marrakech area. Our main purpose is to show the reliability of this optimization approach to build and provide a DST that allows optimizing the coagulant dosage used in a water purification plant, and this at different variations of raw water characteristics. Taking into account the operating state of the plant, we could develop a fast and simple tool that can substitute the coagulant dosage in reducing the operational costs. A description of algorithms used for system development is given in Section 2 and validation results are included in Section 3.

2. DST development

2.1. FSI using LAMDA technique – FSI model

LAMDA is a fuzzy methodology of conceptual clustering and classification [9] developed and applied in several studies [13,15–17]. It allows the representation of classes or concepts by means of the logic connection of all marginal information available. The formation and the recognition of classes are based on the attribution of each object to a class according to the heuristic rule of maximal adequacy. An object is then most likely to belong to the class which presents the greater adequacy degree (GAD). It models the total *indistinguishability* (chaotic homogeneity) or homogeneity inside the description space from which the information is extracted. This is carried by means of a special class called the noninformative class (NIC). This class accepts all items with the same adequacy; therefore, it introduces naturally a classification threshold. LAMDA has two fundamental steps: learning and recognition. At the first stage of learning step (self-learning or unsupervised learning), no previous information is given and LAMDA generates clusters or classes. In this case, it allows obtaining different classifications with the same data-set, by changing LAMDA parameters. Using this strategy on a known data-set, the expert proceeds to a knowledge-based interpretation of such classes. He modifies the LAMDA parameters in order to improve the quality of the final classification. The classes and updated learning parameters are the output of this initial learning stage. In the second stage (supervised learning), this learning allows performing a different number of choices, like learning from an initial set of classes, which can be modified by adding new classes or by updating their parameters or both. It has two alternatives, either the user allows unclassified individuals, meaning that an individual has not been recognized in any class (its adequacy degree is lower than the minimum threshold) and has been placed in the NIC class, or force every individual to be assigned to a class, in this last case the NIC is not taken into account for recognition.

The marginal adequacy degree (MAD) concept is a term related to how similar is one object descriptor to the same descriptor of a given class, and GAD is defined as a membership degree of one object to a given class. Classification process is performed according to a similarity criteria computed in two stages. First, MAD to each existing class is computed for each object descriptor. Second, these partial results will be aggregated in order to get a GAD of an individual to a class. Given that MAD depends on the nature of each descriptor, the algorithm uses general possibility functions. For quantitative descriptors there are several options introduced in [16] to compute the MAD. One possibility function applied is a fuzzy extension of the binomial probability function, which gives as result the following expression:

$$\mathrm{MAD}[x_i|\rho_{j,i}] = \rho_{j,i}^{1-d_{j,i}} - (1-\rho_{j,i})^{d_{j,i}} \tag{1}$$

where $d_{j,i} = |x_i - c_{j,i}|$; $\rho_{j,i}$ is the possibility of the observed element to belong to a class C_j; x is the normalized value of the quantitative descriptor for a particular element; $c_{j,i}$ is center of C_j.

According to Zadeh [18] concept, GAD calculation is performed as an interpolation between T-Norm and T-Conorm by means of the α parameter. $\alpha = 1$ represents the intersection and $\alpha = 0$ means the union.

$$\begin{aligned}\mathrm{GAD}_\alpha(\mathrm{MAD}_1,\ldots,\mathrm{MAD}_n) &= \alpha T(\mathrm{MAD}_1,\ldots,\mathrm{MAD}_n) \\ &\quad + (1-\alpha)S(\mathrm{MAD}_1,\ldots,\mathrm{MAD}_n)\end{aligned} \tag{2}$$

2.2. Data processing using self-organizing maps

The self-organizing feature maps [10] draw some inspiration from the way we believe the human brain works. There are several public domain implementations of SOM, of which we would like to highlight the SOM_PAK and Matlab SOM Toolbox, both developed by Kohonen's research group. In this study, Kohonen's SOM are used to failure data detection and reconstruction. The Kohonen's SOM is trained using unsupervised learning to produce low-dimensional representation of the training samples while preserving the topological properties of the input space. It performs a topology preserving mapping from high-dimensional space onto map units so that relative distances between data points are preserved. The map units, or neurons, form usually a two-dimensional regular lattice. The SOM can thus serve as a clustering tool of high-dimensional data. It also has capability to generalize, i.e. the network can interpolate between previously encountered inputs. Each neuron i of the SOM is represented by an N-dimensional weight $m_i = [m_{i1}, m_{i2}, \ldots, m_{iN}]$, where n is the dimensional of the input vectors. The weight vectors of the SOM form a codebook also called prototype vectors or referent vectors. The neurons of the map are connected to adjacent neurons by a neighborhood relation, which dictates the topology of the map. Usually rectangular or hexagonal topology is used. Immediate neighbors (adjacent neurons) belong to neighborhood N_i of the neuron i. In the basic SOM algorithm, the topological relations, and the number of neurons are fixed from the beginning. The number of neurons determines the granularity of the mapping, which affects accuracy and generalization capability of the SOM. In the training phase, a given training pattern x is presented to the network, and the closest unit is selected. This unit is called Best-Matching Unit (BMU), denoted here by b:

$$\|x - m_b\| = \min_i \{\|x - m_i\|\} \tag{3}$$

where $\|x - m_i\|$ is a distance measure, typically Euclidean.

After finding the BMU, the weight vectors of the SOM are updated. The BMU and its topological neighbors are moved closer to the input vector in the input space. The update rule [10] for the weight vector of unit i is:

$$m_i(t+1) = m_i(t) + \eta(t) \cdot h_{bi}(t) \cdot [(x - m_i(t)] \tag{4}$$

$\eta(t)$ is the learning rate and $h_{bi}(t)$ is the neighborhood function how much unit i is updated when unit b is the winner. Both parameters decrease with time in the learning phase.

2.2.1. Application to DV–DV model

Outliers have always been considered like a source of information distortion gotten from raw data. It is therefore necessary to highlight the diversity of available methods to interpret or to characterize these abnormal values, either while rejecting them in order to restore the data initial properties, or while adopting methods that decrease their impact during the statistical analysis [19,20]. The SOM model combines the goals of projection and clustering algorithms, and may be seen as a method for automatically arranging high-dimensional data. In our case, self-organizing maps allow not only to visualize the evolution of raw water characteristics in two dimensions, but also to detect atypical data by computing the distance between each input vector and its closest reference vector. The basic idea of DV approach consists of the determination of a confidence degree in every data sample, based on monitoring this distance. The validity of a characteristic measurement, for instance, may be put for different reasons: (1) the value is abnormally high or low; (2) the variation between two successive measurements is too important; (3) and the value is incompatible with other measurements of the same quantity obtained by an independent device, etc.

Given a N prototype vectors $\{m_i, \ldots, m_N\}$. Every prototype m_k represents a C_k_som class. The reference space is divided thus in N classes $N(C_k - \text{som})_{k=1}^{N}$. To determine the confidence degree involved in defining the activation of unit i for input x using a Gaussian kernel as:

$$h_i(x) = \exp\left(\frac{-1}{2\sigma_i^2}\|x - m_i\|^2\right) \tag{5}$$

σ_i^2 is a parameter defining the size of the influence region of unit i. σ_i^2 may be computed as the average empirical variance of the n input features, among the samples associated to unit i. More σ_i^2 is bigger; more the influence region of m_i is bigger and therefore more $h_i(x)$ is closer to one.

If the activation $h_b(x)$ of the winning prototype is smaller than a specified threshold, the current sample is considered as abnormal. The contributions of each of the components of vector x to the distance $\|x - m_b\|$ are then examined to determine more precisely which data should be declared as abnormal.

2.2.2. Application to DR – DM model

If vector prototypes provide a good data representation, each missing value of a given input variable can be estimated by the value of the corresponding component of the winning prototype. Given x a new vector, composed of two parts x_o and x_m, containing, respectively, observed and missing values. The main aim is to rebuild x_m from the information provided by Kohonen's map. The method proposed rests on similarity between this new vector $x = (x_o, x_m)$ and the reference vectors m_k. Given X_o and X_m the under-spaces, respectively, of x_o and x_m variables. m_o and m_m are the projections of these under-spaces. According to the activation defined by the Eq. (5), more x_o is closer to m_o, more we will have chance that x_m is closer.

$$h_i(x^o) = \exp\left(\frac{-1}{2\sigma_i^2}\|x^o - m_i^o\|^2\right) \tag{6}$$

The approaches of missing data estimation call for the various techniques, generally presupposing a probabilistic context. For instance, the heuristic methods (such as average and median replacing techniques) are often used and also constitute some simple and little expensive solutions. The parametric methods of maximization, as the expectation maximization (EM) algorithm [21], are extensively used and proved their efficiency, but they require the knowledge or the estimation laws of the variables probabilities. In our case, we can use a simple method that estimates missing data by the component value corresponding to winning prototype m_b:

$$\forall\, p \in M(x), \quad \hat{x}_p = m_{bp} \tag{7}$$

where $M(x)$ is the indexes set of missing values.

This method is very sensitive to the prototype change between two successive vectors of x.

To resolve this problem, we considered another method which takes into account the influence of the k nearest prototypes. Each missing or invalid value j is estimated by a combination of the corresponding component in the k nearest prototypes:

$$\hat{x}(j) = \frac{\sum_{i=1}^{k} h(i) m_i(j)}{\sum_{i=1}^{k} h(i)} \tag{8}$$

where $m_i(j)$ denotes component j of prototype i .

2.3. CDP – CDP model

For CDP model, the perceptron architecture was used with one hidden layer. The architecture of the neural network (number of hidden layers and number of neurons) has been fixed a priori: it is well known that, generally, only one hidden layer is sufficient to model any continuous system if this layer contains enough neurons. The number of neurons in the hidden layer has been optimized with a weight decay pruning method [22] in combination with the Levenberg–Marquardt algorithm [23], allowing the weak weights to be penalized (the connections with weak weight are eliminated). In this framework, the weights and biases of the network are assumed to be random variables with specified distributions. The regularization parameters are related to the unknown variances associated with these distributions. We can then estimate these parameters using statistical techniques. This approach consists of minimizing $Cf(w)$, defined in Eq. (9), using training and modifying this objective function to:

$$Cf'(w) = Cf(w) + \beta\Omega(w) \quad \text{and} \quad \Omega(w) = \frac{1}{p}\sum_{i=1}^{p} w_i^2 \tag{9}$$

p is the number of network weights and β is a parameter that determines the importance of the two terms in the new performance function $Cf'(w)$. Using this, performance functions will cause the network to have smaller weights and biases, and this will force the network response to be smoother and less likely to overfit. This method has the advantage of being simple to implement, since the gradient of Cf' can be easily calculated from the gradient of Cf and the network weights.

3. Results and discussion

3.1. Coagulant dose prediction

A first validation was performed on five descriptors of the raw water quality of *Rocade* plant. Descriptors used are: the turbidity [NTU], the temperature [°C], the pH, the DO [mg/l], and the conductivity [μs/cm^2]. DV and invalid DR were performed before the neural prediction step. The MLP is used with four tangents hyperbolic as transfer functions in the hidden layer. The experimental database (2,511 samples) was separated into two sets: a data-set (2/3 of all data) used for the learning step, in order to identify the model parameters (weights of connections) and a data-set for validation (1/3 of the all data) used to test the network but not used during learning. Simulation results (model structure, adjustment parameters, and validation results) are published in [12].

In the present paper, we tested the system developed with data stemming from other water drinking plants: *Safi* plant, Ouarzazate plant, and El-Kelaa

plant. But the use of data collected during our study encountered a difficulty stemming from the missing descriptors measurements necessary to test this predictive model. We used therefore the same model developed using only three descriptors: the temperature [°C], the pH, and the turbidity [NTU]. The other two inputs (conductivity [μs/cm^2] and DO [mg/l]) are disabled during processing. Data according to temperature, pH, turbidity, and coagulant dose variables are shown in Figs. 2–4. Treatment plants are all supplied by surface water deriving from dams and they are exposed to weather conditions. The temperature has significant variations. The results show a decrease in temperature variable during cold periods and increasing during warm periods. The pH evolution of raw water is almost stable for all samples collected from three plants. However, there was a decrease in pH caused by air pollutants during the winter period. We note fluctuations in turbidity during flood periods that provides charges resulting of the soil erosion and mineral substances dissolved. The water supplying El-Kelaa plant is more turbid (turbidity can happen up to 2,000 NTU). Indeed, the supply of water to the plant is via an open channel. They are therefore exposed to the contributions of storms and floods that cause very high values of turbidity. As for the injected dose of coagulant in the three stations, we see clearly that does not change necessarily with respect to the turbidity variable. For example, the coagulant rate used at the *Ouarzazate* plant is almost invariable with respect to turbidity changes. Table 1 presents

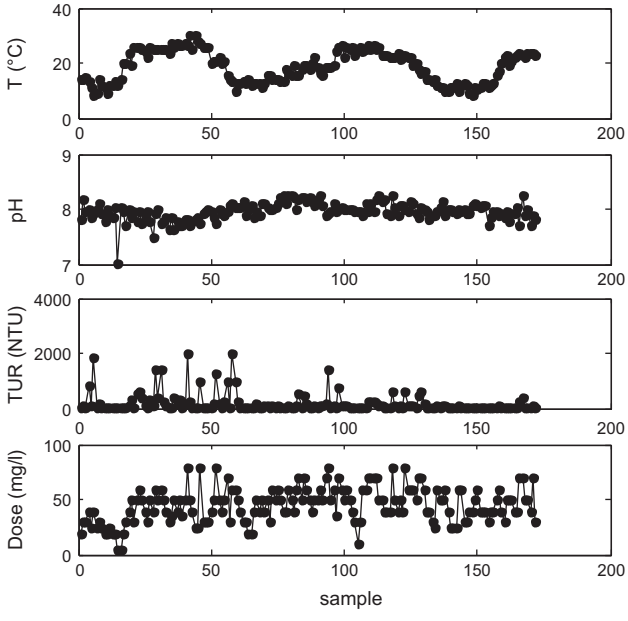

Fig. 2. Raw water characteristics evolution of El-Kelaa plant.

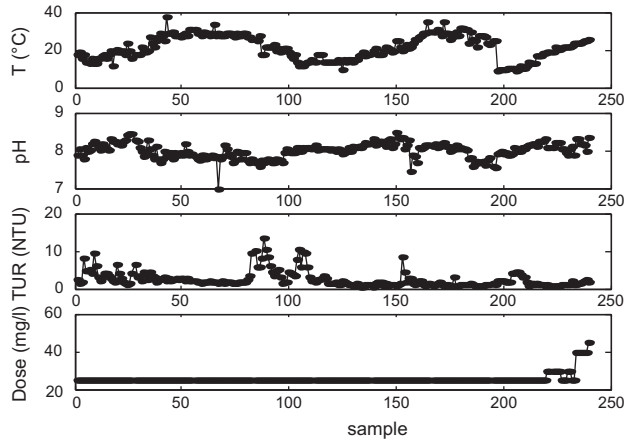

Fig. 3. Raw water characteristics evolution of Ouarzazate plant.

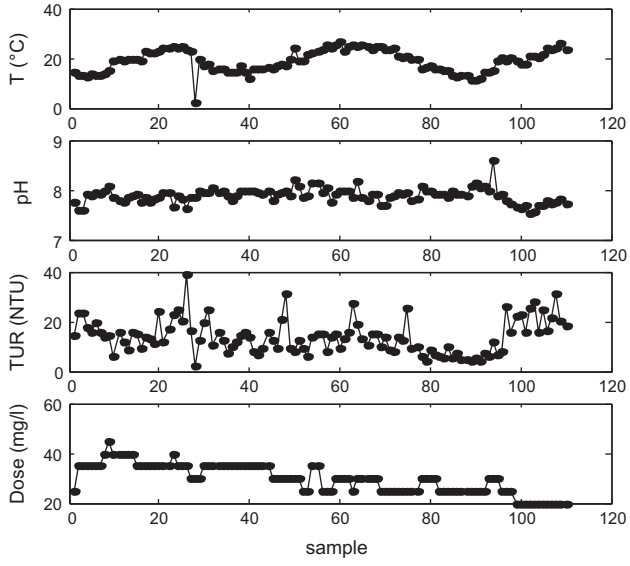

Fig. 4. Raw water characteristics evolution of Safi' plant.

parameters modeling used to identify FSI, DV, DR, and CDP models.

Fig. 5 shows the neural output computed by the CDP model. According to mean squared error computed for four plants (Table 2), we see that the network response is very close to the real output, and consequently, the model will generalize to new data. Figs. 6–8 present simulation results were obtained with experimental data stemming from Safi, Ouarzazate, and El-Kelaa plants. Invalid and missing data stemming from these three plants have been reconstructed by DM model. Given the current results and those previously obtained [12–14], the system developed seems completely adapted to possible variations of raw water

Table 1
Parameters modeling used to identifier FSI, DV, DR, and CDP models

FSI model	Lamda algorithm				Exigency parameter
	$\mathrm{MAD}\left[x_i \middle\rho_{j,i}\right] = \rho_{j,i}^{1-d_{j,i}} - (1-\rho_{j,i})^{d_{j,i}}$				$\alpha = 0.85$
DV + DR model	Map lattice	Map size	Neighborhood function	Neighborhood radius	Initial learning rate
	Hexagonal	25×10	Gaussian	$\sigma_{fin} = 1$	$\eta = 0.95$
CDP model	Hidden layer				Nbre of parameter connexions
	1				86

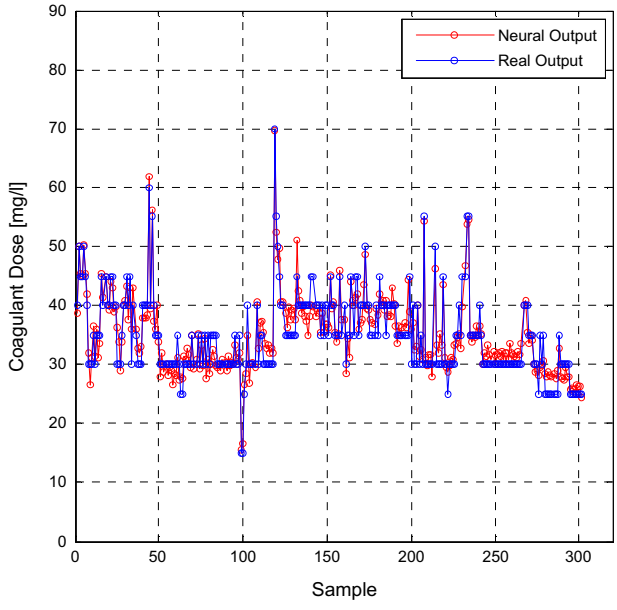

Fig. 5. Experimental coagulant dose (Real Output [mg/l]) and predicted (neural output) according to DV stemming from *Rocade* plant.

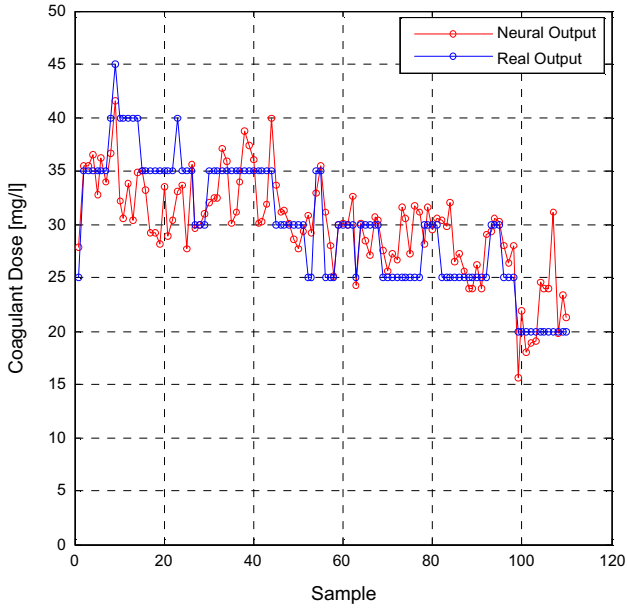

Fig. 6. Experimental coagulant dose (real output [mg/l]) and predicted (neural output) according to DV stemming from Safi plant.

Table 2
Mean-squared error (MSE) according to data plants

	Rocade plant	Safi plant	El-Kalaa plant	Ouarzazate plant
MSE	0.0076	0.13	0.042	0.03

characteristics of *Rocade* and *El-Kelaa* plants. The results obtained in this study clearly showed that the coagulant dose is relatively nonlinear characteristics of the raw water such as turbidity, temperature, and pH. These experimental results showed also the efficiency and reliability of this approach. Model performance depends on the quality and quantity of data available for learning. The importance of neural model is its ability to consider multiple sources of uncertainty, such as inputs atypical and limited information contained in the learning set. However, for *Safi* and *Ouarzazate* plants, we proposed a statistical study for example by means of principal components analysis to extract other key information needed for coagulant modeling.

Given that the systemic study performed about coagulation process and operating conditions of these various plants, the three variables were selected as key-information proof that their influence on the coagulant dose is not useless in the absolute, with the data-set considered for this modeling problem. Therefore, the results of this analysis indicate that, in the context of model optimization, it is possible to keep these variables (including the least influents, in our case conductivity and DO) without disrupting significantly the results of neural modeling. For all validations, the

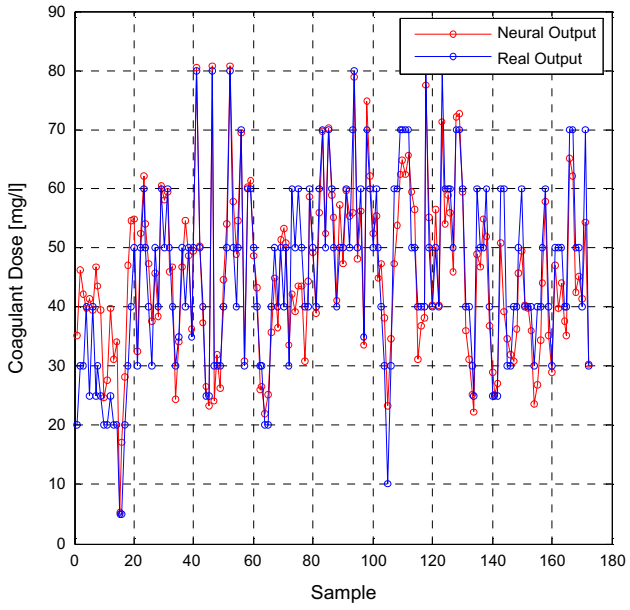

Fig. 7. Experimental coagulant dose (Real Output [mg/l]) and predicted (neural output) according to DV stemming from El-Kelaa plant.

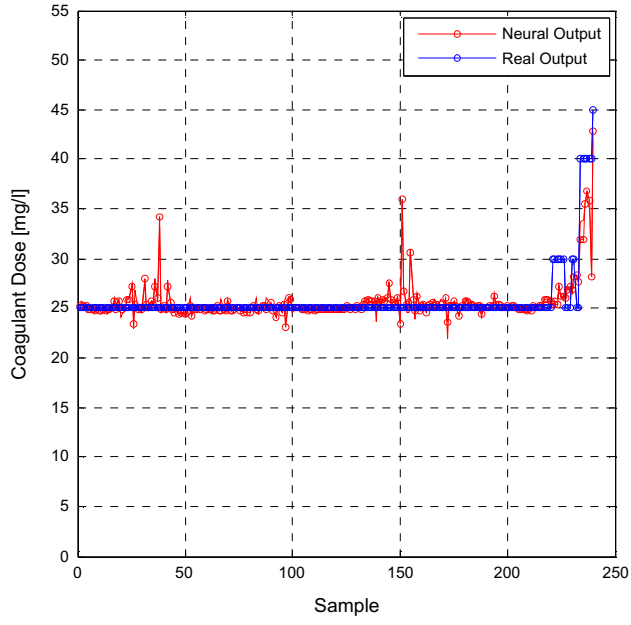

Fig. 8. Experimental coagulant dose (real-output [mg/l]) and predicted (neural output) according to DV stemming from Ouarzazate plant.

three descriptors, temperature, pH, and turbidity, have a greater influence than other two variables, conductivity, and DO. The three descriptors inform the model and report some knowledge about physical and chemical interactions generated in the coagulation process.

4. Conclusion

These development issues are mitigated in this work by using soft computing techniques: neural networks and fuzzy logic. At recent years, this type of approaches is contributed to various fields of water treatment research such as demand forecasting, purification performance prediction, and control processes [24–26]. Here, different models are improved by using a priori knowledge of raw water descriptors and by exploiting knowledge of the plant operator to achieve a flexible and reliable predictor as a DST for coagulating automatic control of the water purification plants of Tensift-Marrakech area as on as case study.

In this paper, we proved through the validation results of system developed that the coagulation process is relatively nonlinear with respect to raw water characteristics studied in this work. We initially built the neural model to predict the coagulant dose from five descriptors: temperature [°C], pH, turbidity [NTU], conductivity [μs/cm^2], and DO [mg/l]. A step of DV and DR has been evaluated in a previous study [12]. The experimental results showed the effectiveness and reliability of the fuzzy-neural approach for this type of complex process. The fuzzy-neural model has a prediction rate close to 93.6% towards data acquired from *Rocade* plant. Model performances depend on the quality and quantity of data available for learning. We tested this model on data stemming from other plants and the results were satisfactory according to database used for validation.

From the perspective of development time, it is clear that the approaches developed are advantageous compared to other techniques, where the deployment of other data requires the built of a new model that will be much easier to carry by learning. The model variables remain the same (in a first approximation and unless major evolution of technique), so that learning of a new neural model boils down to elaborate a representative database, and to implement usual learning algorithms by varying the model complexity (varying the number of hidden neurons). It is always desirable and often possible to use for network design, mathematical knowledge or physical laws, which are available for the phenomenon to model. As a first perspective of this work, we plan to expand this type of modeling approach to optimize others process of water treatment. Another aim is to emerge this methodology to optimize other processes and develop other DSTs for control and supervision plants of water purification.

Acknowledgments

The authors gratefully thank all members of the ONEP (Office National de l'Eau Potable, Marrakech, Morocco) for their constructive comments and their assistance during this study. We would like also to thank *Louis Wehenkel* from the Modeling and Systems Research Unit (Montefiore Institute, Liege-Belgium) for his cooperation in this project.

Symbols

$Cf\ (w)$	—	cost function optimized with Levenberg–Marquardt algorithm
$Cf'\ (w)$	—	cost function optimized with weight decay and Levenberg–Marquardt algorithm
$c_{j,i}$	—	center of C_j
$d_{j,i}$	—	| {x_i - c_{j,i} } |
GAD	—	global adequacy degree ($\mathrm{GAD}\alpha\,(\mathrm{MAD}_1,\ldots,\mathrm{MAD}_n) = \alpha T(\mathrm{MAD}_1,\ldots,\mathrm{MAD}_n) + (1-\alpha)\, S(\mathrm{MAD}_1,\ldots,\mathrm{MAD}_n)$)
$h_{bi}\ (t)$	—	neighborhood function how much unit i is updated when unit b is the winner $m_i(t+1) = m_i(t) + \eta(t)\cdot h_{bi}(t)\cdot[(x - m_i(t)]$; $h_i(x) = \exp\left(\frac{-1}{2\sigma_i^2}\|x - m_i\|^2\right)$
m_b	—	best matching unit ($\|x - m_b\| = \min_i\{\|x - m_i\|\}$)
MAD	—	marginal adequacy degree $\left(\mathrm{MAD}[x_i|\rho_{j,i}] = \rho_{j,i}^{1-d_{j,i}} - (1-\rho_{j,i})^{d_{j,i}}\right)$
x	—	normalized value of the quantitative descriptor for a particular element
$\|x - m_i\|$	—	Euclidean distance
α	—	exigency parameter of LAMDA algorithm. $\alpha = 1$ represents the intersection and $\alpha = 0$ means the union
β	—	parameter that determines the importance of the two terms in the new performance function
$\eta\ (t)$	—	learning rate
$\rho_{j,i}$	—	the possibility of the observed element to belong to a class C_j
σ_i^2	—	neighborhood radius defining the size of the influence region of unit i

References

[1] W.J. Masschelein, H. Chevolet, Processus unitaires du traitement de l'eau potable [Unit processes in drinking water treatment], Cebedoc, Paris, 1997.
[2] Degrémont. Mémento technique de l'eau [Water treatment handbook] (Tomes 1 et 2, 10 Ed.), Tec & Doc (Ed), Paris, 2007.
[3] W.F. Gerdes, A new instrument-the streaming current detector, in: Proceedings of the 12th National ISA analysis Instrument Symposium, Houston, TX, USA, May 1966, pp. 181–198.
[4] C.A. Walker, J.T. Kirby, S.K. Dentel, The streaming current detector: A quantitative model, J. Colloid Interface Sci. 182 (1996) 71–81.
[5] C. Gagnon, B.P.A. Grandjean, J. Thibault, Modelling of coagulant dosage in a water treatment plant, Eng. Appl. Artif. Intell. 11 (1997) 401–404.
[6] A. Adgar, C.S. Cox, P.R. Daniel, A. Lowdon, Chemical dosing philosophies for a water treatment plant: Results of some pilot plant experimentation, In: Proceedings of the IEE Conference Publication 427 (1996) 1052–1057.
[7] N. Valentin, T. Denoeux, A neural network-based software sensor for coagulation control in a water treatment plant, Intell. Data Anal. 5 (2001) 23–39.
[8] B. Lamrini, M.V. Le Lann, A., E.K. Lakhal, Control, fault detection, and supervision in technical processes–Application to surface water treatment, in: 1st Water and Environment International Conference, October 26–29, Marrakech, 2011.
[9] J. Aguilar-Martin, M. Balssa, R. Lopez De Mantras, Recursive estimation of partitions: Examples of learning and self teaching in RN and IN, Questiio (Quaderns d'Etadisticia, Siestems, Informatica i Investigacio Operativa) 5 (1981) 150–172.
[10] T. Kohonen, Self-organizing Maps, vol. 30, Springer Series in Information Sciences, Berlin, 1995.
[11] G. Dreyfus, J.M. Martinez, M. Samuelides, M.B. Gordon, F. Badron, S. Thiria, L. Hérault, Réseaux de neurones: Méthodologie et applications [Neural networks: Methodology and applications], Eyrolles, Paris, 2005.
[12] B. Lamrini, A. Benhammou, M.V. Le Lann, A. Karama, A neural software sensor for on-line prediction of coagulant dosage: Application to a drinking water treatment Plant, Trans. Inst. Meas. Control (TIMC) 27 (2005) 95–213.
[13] B. Lamrini, M.V. Le Lann, A. Benhammou, E.K. Lakhal, Gestion supervisée d'une unité de coagulation pour la potabilisation des eaux à partir d'une méthodologie d'apprentissage et d'expertise [Coagulation monitoring through a learning and expertise methodology for drinking water], J. Water Sci. 20 (2007) 325–338.
[14] B. Lamrini, E.K. Lakhal, M.V. Le Lann, L. Wehenkel, Data validation and missing data reconstruction using self organizing map for water treatment, Neural Comput. Appl. 20 (2011) 575–588.
[15] J. Waissmann-Vilanova, Building a behavioural model for process supervision: Application to a wastewater treatment plant [PhD thesis], LAAS-CNRS, Institut National Polytechnique (INP), Toulouse, 2000.
[16] T. Kempowsky, Surveillance des procédés à base de méthodes de classification: conception d'un outil d'aide pour la détection et le diagnostic des défaillances Process monitoring based on classification methods: A design tool for fault detection and diagnosis [PhD thesis], LAAS-CNRS, Institut National des Sciences Appliquées (INSA), Toulouse, France, 2004.
[17] L. Hedjazi, J. Aguilar Martin, M.V. Le Lann, Similarity-margin based feature selection for symbolic interval data, Pattern Recognit. Lett. 32(4) (2011) 578–585.
[18] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Set. Syst. 1 (1978) 3–28.
[19] V. Barnett, T. Lewis, Outliers in Statistical Data, Wiley Series in Probability & Mathematical Statistics, third ed, Wiley, New York, 1994.
[20] D. Cousineau, S. Chartier, Outliers detection and treatment: A review, Int. J. Psychol. Res. 3(1) (2010) 58–67.
[21] C. Biernacki, G. Celeux, J.F. Si Abdallah, G. Govaert, F. Langrognet, MIXMOD User's Guide (MIXture MODelling software: High Performance Model-based Cluster and Discriminant Analysis, Univ. Franche-Comté, France. 2009. http://www-math.univ-fcomte.fr/mixmod/index.php.

[22] D.J.C. MacKay, A practical Bayesian framework for backpropagation networks, Neural Comput. 4 (1992) 448–472.
[23] F.D. Foresee, M.T. Hagan, Gauss-Newton approximation to Bayesian regularisation, in: Proceedings of the Int. Joint Conf. on, Neural Networks, Houston, TX, 1997, pp. 1930–1935.
[24] H.R. Tashaouie, G.B. Gholikandi, H. Hazrati, Artificial neural network modeling for predict performance of pressure filters in a water treatment plant, Desalin. Water Treat. 39 (2012) 192–198.
[25] J.M. Grosso, C. Ocampo-Martínez, V. Puig, Adaptive multilevel neuro-fuzzy model predictive control for drinking water networks, in: 20th Mediterranean Conference on Control and Automation, IEEE, Barcelona, 2012. pp. 1548–1553.
[26] J.A. Hernández, D. Colorado, Uncertainty analysis of COP prediction in a water purification system integrated into a heat transformer using several artificial neural networks, Desalin. Water Treat. 52(7–9) (2012) 1443–1456.