MODELLING OF COAGULANT DOSAGE IN A WATER TREATMENT PLANT

N. Valentin[1,2], T. Denoeux[2], F. Fotoohi[1]

(1) CITI - Suez Lyonnaise des Eaux, Technopolis, ZAC de Mercières, 14, rue du Fonds Pernant, 60471 Compiègne, France
Email: Nicolas.Valentin, Farrokh.Fotoohi @citi.suez-lyonnaise-eaux.fr

(2) Heudiasyc, UMR CNRS 6599, Université de Technologie de Compiègne, BP 20529, 60205 Compiègne, FRANCE
Email: Thierry.Denoeux@hds.utc.fr

Abstract: Artificial Neural Network (ANN) techniques are applied to the control of coagulant dosing in a drinking water treatment plant. Coagulant dosing rate is non-linearly correlated to raw water parameters such as turbidity, conductivity, pH, temperature, etc. An important requirement of the application is *robustness* of the system against erroneous sensor measurements or unusual water characteristics. The hybrid system developed includes raw data validation and reconstruction based on a Kohonen self-organizing feature map, and prediction of coagulant dosage using multilayer perceptrons. A key feature of the system is its ability to take into account various sources of uncertainty, such as atypical input data, measurement errors and limited information content of the training set. Experimental results with real data are presented.

I. INTRODUCTION

The water industry is facing increased pressure to produce higher quality treated water at a lower cost. The coagulation-flocculation process is a major step in the production of potable water, allowing the removal of colloidal particles. The main difficulty is to determine the optimum coagulant dosage related to the influent of raw water. Excessive coagulant overdosing leads to increased treatment costs and public health concerns, while underdosing leads to a failure to meet the water quality targets and less efficient operation of the water treatment plant. For the moment, both manual and automatic methods are available to predict optimum coagulant dosage rate [1-2]. Manual methods mainly include jar testing. Jar testing involves taking a raw water sample and applying different quantities of coagulant to each sample. After a short period of time each sample is assessed for water quality and the dosage that produces the optimal result is used as a set point. Operators change the dose and make a new jar test if the quality of treated water changes. Disadvantages associated with jar testing are the necessity to perform manual intervention, and the limitation to feedback control. Automatic coagulant control is ensured mainly by streaming current detectors (SCD) [3], which measure the residual charge on colloidal colour and turbidity particles in the water. As these particle have a negative charge and the coagulant ions have a positive charge, the amount of coagulant added dictates the magnitude and sign of the net electrical charge. The system controls this net charge at a set point which has been shown by jar testing to provide close to optimum coagulation under a certain range of raw water conditions. Disadvantages associated with the SCD are its operation cost and its lack of adaptation to all types of raw water quality.

This paper addresses the problem of on-line determination of optimal coagulant dosage from raw water characteristics such as turbidity, pH, conductivity, etc. Some previous studies [4-5] have shown the potential effectiveness of such an approach based on ANN's. The innovative aspect of this work resides in the integration of various techniques in a global system including data validation and reconstruction, modelling of coagulant dosage and analysis of uncertainties. Given the high variability of the inputs and the low reliability of available sensors, an important requirement of the application is *robustness* of the system, running without human supervision, against erroneous sensor measurements or unusual water characteristics, due to accidental pollution for instance. Special attention has also to be paid to the automatic design and training of such a system from learning data (including the phases of data validation, input selection and model choice), which should allow the portability of the system at low cost from one site to another.

This paper is organized as follows. The water treatment operation is first explained in Section II. The methodology used to build ANN models for the prediction of coagulant dosage is then described in Section III. Finally, experimental results are presented and discussed in Section IV.

II . WATER TREATMENT OPERATION

Water treatment involves physical, chemical and biological processes that transform raw water into drinking water. However, contrary to most industrial processes, for which the quality of the input raw material is under control, the quality of the given raw water source may fluctuate due to natural perturbation or occasional pollution. The Viry-Chatillon water treatment plant, which was used as an application site for this study, provides water to more than 300 000 inhabitants and has a nominal capacity to process 120 000 m3 of water per day. Figure 1 presents a schematic overview of the various operations necessary to treat the water, the available measurements, and the coagulant dosing point.

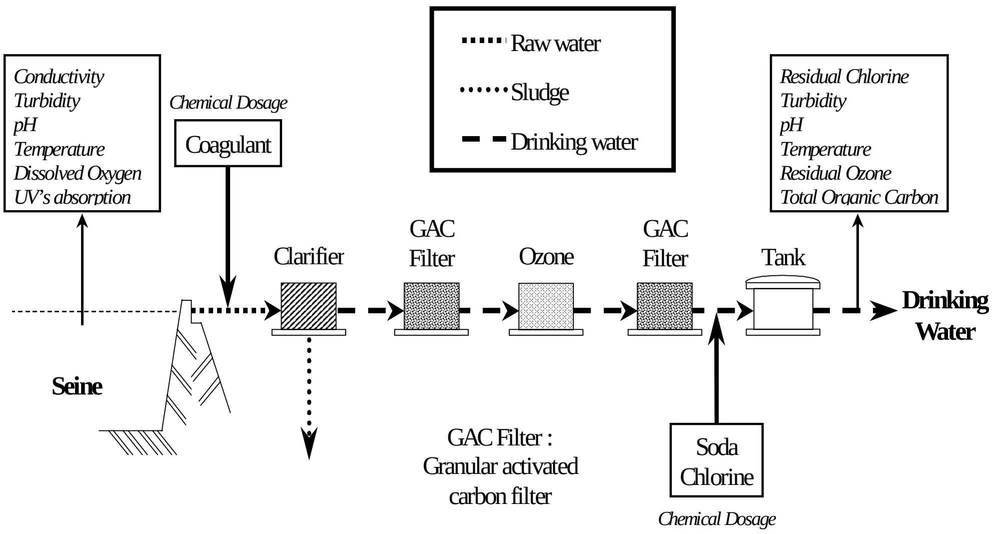

Figure 1 - Simplified synopsis of the water treatment plant.

Raw water is extracted from the river Seine and pumped to the treatment works. The treatment consists in coagulation-flocculation, settling, filtration, ozonation, filtration and final disinfection. The water is then stored in a tank and ready to be transported through the water supply network. The coagulation-flocculation step, which requires the addition of a chemical coagulant, is the critical process to remove colloidal solids.

III. METHODOLOGY

The system developed was divided in three modules : single-parameter data validation, multi-parameter data validation and reconstruction, and modelling of coagulant dosage. Figure 2 illustrates the structure of the system.

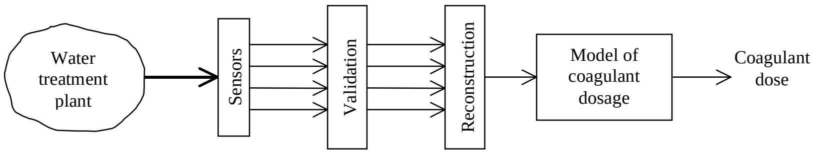

Figure 2 - Structure of the system for the prediction of the coagulant dosage.

3.1 *Single-parameter data validation*

As the system has to operate on-line with unreliable input data coming from a real process, the *data validation* step is fundamental. The objective of this first step, referred to as single-parameter sensor data validation, is to associate to each raw data item a confidence level. Confidence is measured by a real number belonging to [0,1]. A zero confidence means "no confidence, the data is invalid" while a confidence equal to one means "the data is perfectly valid". At this stage, the validation procedure is quite simple and is based on a comparison of each signal and its derivative to a range of values typically obtained in the absence of erroneous measurements. Raw data whose confidence level is less than a given threshold are declared as invalid data. Although this simple approaches proves to be sufficient in most cases, the detection of inconsistences in the data involving more than one parameter requires the use of more sophisticated techniques such as Kohonen maps.

3.2 *Multi-parameter data validation and reconstruction*

We propose an approach based on the use of a Self-Organizing Map (SOM) [6] for multi-parameter data validation and reconstruction of input data. The process in which the SOM is formed is an unsupervised learning process. The SOM defines a mapping from the input data space $\Re^n$ (raw water quality parameters) onto a regular two-dimensional array of nodes. A reference vector, or prototype, $m_i \in \Re^n$ is associated to every node i. Each input vector $x \in \Re^n$ is compared with the m_i, and the best match defines the winning prototype. The input is then mapped onto the corresponding location on the grid.

With this technique, the evolution of raw water quality can be visualized in two dimensions, and atypical data or outliers can be detected by measuring the distance between each input vector and its closest reference vector. For reconstruction, each missing value of a given input variable is estimated by the value of the corresponding component in the winning prototype.

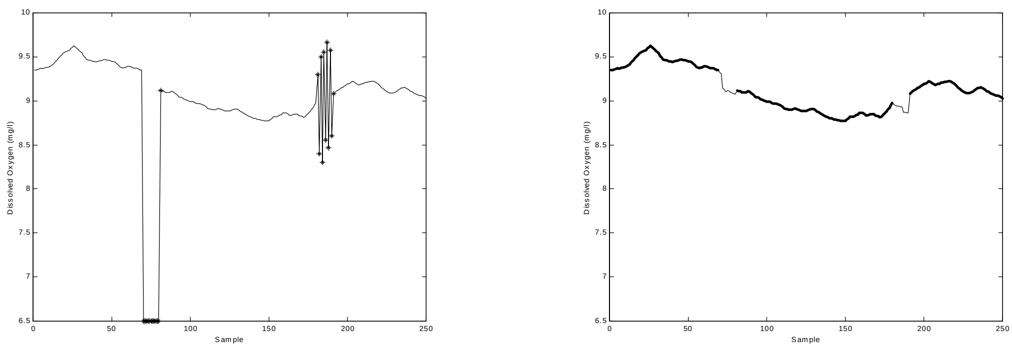

Figure 3: Example of data validation (*: invalid data) and reconstruction (thin line: reconstructed signal) on dissolved oxygen.

Figure 3 illustrates data validation and reconstruction for dissolved oxygen data. At this stage, the pre-processing phase is terminated and the data is ready to be processed by the coagulant dosage model.

3.3 *Modelling of coagulant dosage*

For the modelling of coagulant dosage, a sigmoid multilayer perceptron (MLP) was trained using the Optimal Brain Damage (OBD) [7] learning and pruning algorithm. This approach is based on the following general procedure.

First, a relatively large network (Figure 4) is trained using the back-propagation algorithm. The network is then examined to assess the relative importance of the weights, and the least important are deleted. The implementation of this technique consists of the following steps (Figure 5):

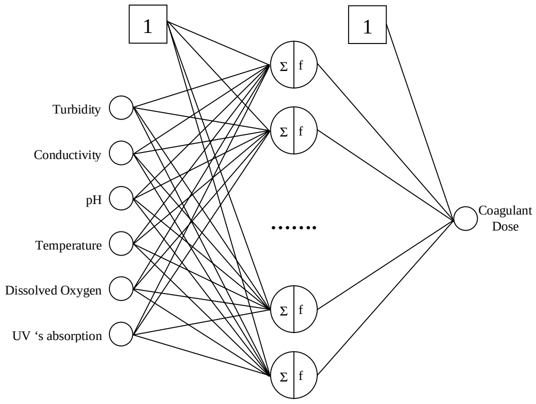

Figure 4: Initial network

1. Choose a relatively large initial network architecture.
2. Train the network using a back-propagation algorithm applied to a sum of squares error function.
3. Compute the second derivatives H_{ii} for each of weights and evaluate the saliencies.
4. Sort the weights by saliency and delete some of the low-saliency weights.
5. Go to step 2 and repeat until some overall stopping criterion is reached.

For practical use, a software system to predict the optimal coagulant dosing rate should not only provide point estimates but also confidence intervals. Bootstrap sampling [8] was used to generate confidence intervals for the system outputs [9]. As shown in Figure 6, 50 bootstrap sets of training data were created from the original data training data by resampling with replacement. These bootstrap training sets were used to train 50 bootstrap MLP models using the same architecture and training procedure described previously. Lower and upper limit confidence bounds for any input were obtained by sorting these outputs and selecting the 10% and 90% cumulative levels. The confidence interval provides upper and lower limits on the prediction.

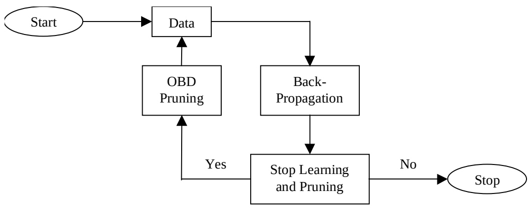

Figure 5 : Learning and pruning algorithm

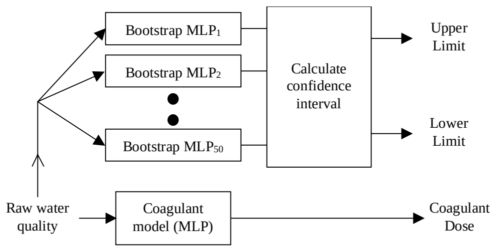

Figure 6 : Bootstrap sampling to generate confidence interval.

IV. RESULTS

The water treatment plant of Viry-Chatillon has been well instrumented for several years. Various process variables (see Figure 1) such as conductivity, turbidity, pH, temperature, dissolved oxygen, UV absorption of raw water and coagulant dosage are available. The raw database is made of 100000 measurements of each variable during a period of 12 month (November 97 – November 98) sampled every 5 minutes.

Many sensor faults were detected and only 1600 measurements of each parameters were available for training. Figure 7 shows the wide range of raw water conditions that can exist on the treatment plant.

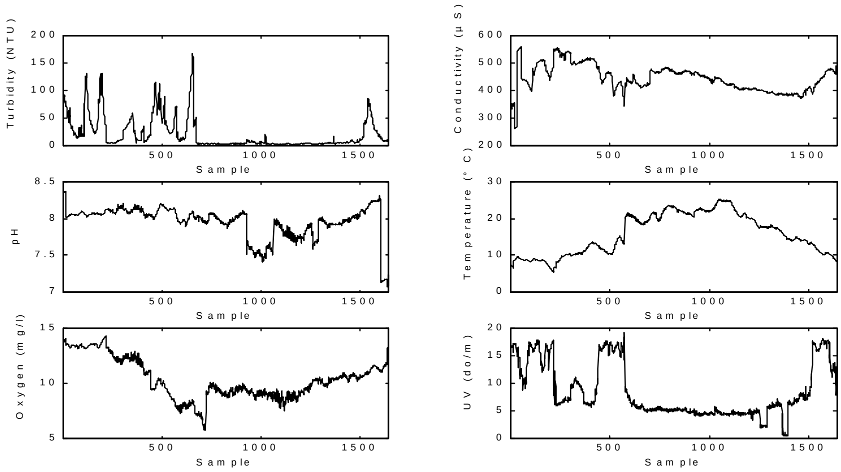

Figure 7 : Raw Water Data

We used 70% of the data set to develop the model, find the best structure of the ANN and estimate the prediction accuracy by bootstrap. The other part of the data set was used to validate the resulting model. The prediction accuracy and confidence interval of the ANN are shown in Figure 8 on the validation set.

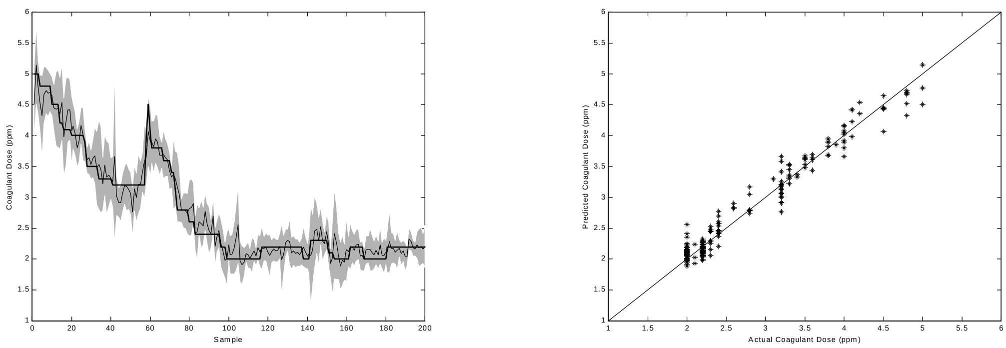

Figure 8: Actual (thick line) versus predicted (thin line) coagulant dosage with ANN model on test data and confidence interval (shaded region).

A linear regression model was also developed for comparison with the neural network model. Figure 9 shows the outputs of the linear model whose parameters were estimated with the same data as the ANN. The prediction accuracy is clearly inferior to that of the ANN model.

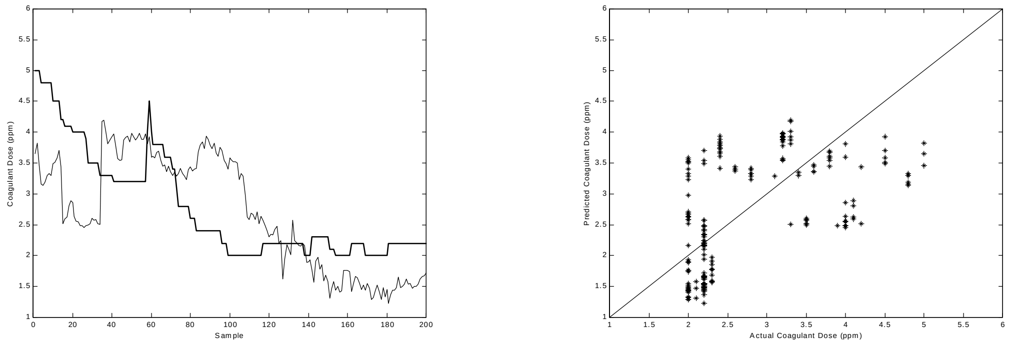

Figure 9 : Actual (thick line) versus predicted (thin line) coagulant dosage with linear regression model.

V. CONCLUSIONS

An integrated coagulant dosing system based on unsupervised and supervised neural network models, as well as various statistical techniques, has been described. Experimental results using real data have demonstrated the efficiency and soundness of this approach. Field testing is currently underway to fully validate the system before its widespread dissemination to other sites. More robustness, if required, of the system could be provided several techniques such as (1) input probability density estimation using, e.g., the EM algorithm with a Gaussian mixture approach (allowing to determine the validity domain of the system), and (2) propagation of the imprecision of reconstructed inputs using sensitivity analysis and interval arithmetic. The performance of the network is dependent on the quality and completeness of data provided for system training. As such, continuous updating of training data would certainly improve the performance of the system. This model, however, is only based on the previous behaviour of operators and jar-test results. Further work is needed to develop a model taking into account the dynamics of the system, and to predict treated water parameters (mainly turbidity) at the output of the clarification process.

REFERENCES

[1] C. Lind (1994). Coagulation Control and Optimization : Part One. *Public Works for October*, 56-57.
[2] C. Lind (1994). Coagulation Control and Optimization : Part Two. *Public Works for November*, 32-33.
[3] F. Bernazeau, P. Pierrone, J.P. Duguet (1992). Interest in using a streamline current detector for automatic coagulant dose control. *Water Supply* **10** (4), 87-96.
[4] A. Mirsepassi, B. Cathers, H.B. Dharmappa (1995). Application of Artificial Neural Networks to the Real Time Operation of Water Treatment Plants. In *Proceedings International Joint Conference on Neural Networks*, Volume 1, pp. 516-521.
[5] J. Evans, C. Enoch, M. Johnson, P. Williams (1998). Intelligent based auto-coagulation control applied to a water treatment works. In *Proceedings of International Conference on Control*, pp. 141-145.
[6] T. Kohonen (1995). *Self-Organizing Maps*. Heidelberg: Springer Verlag.
[7] Y. Le Cum, J.S. Denker, S.A. Solla (1990). Optimal Brain Damage. In D.S. Touretzky (Ed.), *Advances in Neural Information Processing Systems,* Volume 2, pp. 598-605. San Mateo, CA: Morgan Kaufman.
[8] B. Efron, R.J. Tibshirani (1993). *An Introduction to the Bootsrap*. New York: Chapmann & Hall.
[9] R. P. Lippmann, L. Kukolich, D. Shahian (1997). Predicting of Complications in Coronary Artery Bypass Operations using Neural Networks. In *Neural Information Processing System* **7**, 1055-1062.