

Article

« Gestion supervisée d'une unité de coagulation pour la potabilisation des eaux à partir d'une méthodologie d'apprentissage et d'expertise »

Bouchra Lamrini, Marie-Véronique Le Lann, El Khadir Lakhal et Ahmed Benhammou
Revue des sciences de l'eau / Journal of Water Science, vol. 20, n° 4, 2007, p. 325-338.

Pour citer cet article, utiliser l'information suivante :

URI: <http://id.erudit.org/iderudit/016908ar>

DOI: 10.7202/016908ar

Note : les règles d'écriture des références bibliographiques peuvent varier selon les différents domaines du savoir.

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter à l'URI <https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. Érudit offre des services d'édition numérique de documents scientifiques depuis 1998.

Pour communiquer avec les responsables d'Érudit : info@erudit.org

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

BOUCHRA LAMRINI^{1*}, MARIE-VÉRONIQUE LE LANN^{2, **}, EL KHADIR LAKHAL¹,
AHMED BENHAMMOU¹

¹Laboratoire d'Automatique, de l'Environnement et de Procédés de Transfert (LAEPT), Faculté de Sciences Semlalia,
BP 2390, 40000 - Marrakech, Maroc

²LAAS-CNRS, Université de Toulouse, 7, Avenue du Colonel Roche, 31077 Toulouse Cedex 4, France

Reçu le 7 mai 2005, accepté le 2 mai 2007

RÉSUMÉ

Le travail présenté propose une méthodologie de classification par apprentissage qui permet l'identification des états fonctionnels sur une unité de coagulation impliquée dans le traitement des eaux de surface. La supervision et le diagnostic de ce procédé ont été réalisés en utilisant la méthode de classification LAMDA (Learning Algorithm for Multivariate Data Analysis). Cette méthodologie d'apprentissage et d'expertise permet d'exploiter et d'agréger toutes les informations provenant du procédé et de son environnement ainsi que les connaissances de l'expert. L'étude montre qu'il est possible d'ajouter aux informations issues des capteurs classiques (température, matières en suspension, pH, conductivité, oxygène dissous), la valeur de la dose de coagulant calculée par un capteur logiciel développé dans une étude antérieure afin d'affiner le diagnostic. Le site d'application choisi pour l'identification des états fonctionnels est la station

de production d'eau potable Rocade de la ville de Marrakech, Maroc.

Mots clefs : Procédé de coagulation, classification, apprentissage supervisé, apprentissage non supervisé, reconnaissance de formes, logique floue.

ABSTRACT

The present work proposes a learning classification method to identify the functional states of a coagulation process for the treatment of surface water and production of drinking water. Supervisory control and diagnosis were performed using the LAMDA (Learning Algorithm for Multivariate Data Analysis) classification technique. This expert learning method involves

*Auteur pour correspondance :
Télécopieur : +212 24 43 74 10
+33 5 61 33 69 36
Courriel : blamrini@yahoo.fr
mvlelann@laas.fr

the processing and aggregation of all information stemming from an environmental process, and it allows the incorporation of the user's knowledge. The study shows that it is possible to refine the diagnosis by taking into account the information obtained from common sensors (e.g., temperature, suspended solids, pH, conductivity, dissolved oxygen) together with the predicted coagulant dosage, as computed with an intelligent software sensor developed previously. The Rocade drinking water plant located at Marrakech, Morocco was chosen to test the method.

Key words: *Coagulation process, classification, supervised learning, unsupervised learning, pattern recognition, fuzzy logic.*

1. INTRODUCTION

Le contrôle et la conduite des procédés industriels sont délicats et difficiles notamment lorsque les lois physiques décrivant le système sont trop complexes ou incomplètes. Les dérives et les défauts sont alors difficiles à détecter et à localiser sans utiliser de modèle représentatif du système réel. Certes, les procédés de coagulation sont fortement non linéaires et non stationnaires et leur fonctionnement reste difficilement modélisable à cause de la connaissance partielle des phénomènes mis en jeu. La conduite supervisée de ces procédés nécessite la définition de méthodes avancées de contrôle et de supervision qui intègrent la connaissance insuffisante que l'on a de ces procédés. Jusqu'à présent, il n'existe pas de modèle mathématique ou structurel permettant d'identifier de façon certaine l'état de fonctionnement d'un procédé de coagulation. C'est généralement au moyen d'analyses hors-ligne que les opérateurs contrôlent indirectement cet état et interviennent en retour sur le processus avec un délai important. Cependant, l'opérateur dispose de nombreuses informations issues du procédé lui-même et de son environnement, souvent précises, dont il est facile de faire une acquisition en ligne. Ces informations traitées de façon individuelle ne permettent toujours pas de caractériser l'état du procédé.

Depuis plusieurs décennies, les études publiées sur la coagulation ont mis en évidence des mécanismes fortement complexes dépendant d'un nombre important de variables. Ainsi, d'importants efforts et différentes approches ont été développés pour construire et mettre en œuvre des modèles de contrôle automatique de la dose de coagulant (BAZERBACHI *et al.*, 1986; EVANS et ENOCH, 1998; GAGNON *et al.*, 1997; LAMRINI *et al.*, 2005; VALENTIN *et al.*, 1999) et des outils de supervision permettant d'optimiser ce procédé

de traitement (KAN et HUANG, 1998; LIU et WU, 1997; STANDEN *et al.*, 1997). En ce qui concerne les outils dits de supervision, dans le cas des travaux de KAN et HUANG (1998), même si on parle de supervision ou de suivi, l'essentiel du développement porte sur la détermination automatique de la dose de coagulant grâce à un suivi optique. De même dans le cas des travaux de STANDEN *et al.* (1997), la modification automatique de la dose de coagulant se fait en rapport à la mesure de la turbidité et au nombre de particules colloïdales mesurées au préalable par diffraction laser lors de jar-tests. Les travaux de LIU et WU (1997) sont quant à eux directement reliés au développement d'un contrôleur flou allié à un détecteur de courant permettant d'améliorer l'opération de coagulation. On voit donc que, plus qu'un suivi de fonctionnement d'une opération, il s'agit plutôt d'un contrôle automatique de la dose de coagulant, l'objectif étant de déterminer la dose à appliquer en fonction des paramètres de l'eau brute observés. Cependant, cette structure de contrôle automatique ne permet pas de diagnostiquer une dérive ou un mauvais fonctionnement à elle seule. Pour cela, il faut lui adjoindre une structure de supervision/diagnostic permettant de déterminer l'état de fonctionnement de l'unité (cet état de fonctionnement pouvant être une défaillance).

Pour assurer une bonne gestion et contribuer au bon fonctionnement de ce procédé, il semble nécessaire d'exploiter simultanément toutes les informations disponibles sur le procédé (les caractéristiques de l'eau brute et leurs évolutions provenant des anomalies imprévues), et de son environnement, ainsi que les connaissances de l'expert. Pour ces raisons, nous proposons de réaliser la conduite supervisée de ce procédé en utilisant une technique de classification par apprentissage, la technique LAMDA. Une telle technique permet d'exploiter et d'agréger toute information sur le procédé afin de reconnaître des situations spécifiques et agir, si possible, sur ce procédé. L'idée générale de la classification est l'évaluation des signaux significatifs du système (caractéristiques de l'eau brute, taux de coagulant, etc.), afin de reconnaître les facteurs liés à une quelconque situation et d'aider l'opérateur à prendre une décision lors de l'apparition de dysfonctionnement.

Le travail présenté dans le cadre de cet article est structuré de la manière suivante : la première section présente le site d'application choisi pour identifier les états fonctionnels. Dans la deuxième section, on présente une brève description de la méthode LAMDA. Finalement, dans la troisième section, des résultats expérimentaux seront présentés et discutés. Comme le montrent ces résultats obtenus lors de l'application de la méthodologie LAMDA aux données expérimentales couvrant 4 années de fonctionnement de la station de production d'eau potable de Marrakech, ils ouvrent la possibilité d'appliquer cette méthode à la surveillance en temps réel de ce procédé.

2. PRÉSENTATION DU SITE D'APPLICATION

L'installation de traitement concernée par cette étude est la station de production d'eau potable Rocade de la ville de Marrakech. Elle alimente en eau potable, de façon variable dans le temps, une population de 1,5 millions habitants. L'eau à l'entrée de la station provient d'une prise située dans le canal Rocade. En cas de défaillance de la ressource (crues, pollution accidentelle, travaux de maintenance, etc.), la station peut être alimentée à partir d'un réservoir de mise en charge situé près de la station de pompage Takerkoust. En se référant aux historiques de la station, soixante pour cent des besoins de la ville de Marrakech sont assurés par la station Rocade et les quarante pour cent restants sont fournis par l'eau des ressources souterraines (puits, forages, etc.). À noter que la capacité maximale de traitement de la station étudiée est d'environ $1\,400\text{ L s}^{-1}$.

Compte tenu des caractéristiques de l'eau brute acheminée vers la station, les décideurs (Office National de l'Eau Potable, ONEP) utilisent le procédé physico-chimique classique illustré par la figure 1. Ce procédé comporte les phases suivantes : la pré-chloration, la coagulation-floculation, la décantation, la filtration et la désinfection. Cependant, la station présente des limites de fonctionnement au niveau du traitement des matières en suspension (MES) présentes dans l'eau brute. Ainsi, les limites de fonctionnement se présentent comme suit :

-2 g L^{-1} en MES : limite d'un fonctionnement normal de la station;

$-2-3\text{ g L}^{-1}$ en MES : diminution de la production de la station;

$\rightarrow 3\text{ à }4\text{ g L}^{-1}$ en MES : arrêt du fonctionnement de la station.

3. LA MÉTHODE DE CLASSIFICATION LAMDA

La méthode LAMDA est une technique de classification développée par PIERA *et al.* (1989) suite à l'idée originale d'AGUILAR-MARTIN *et al.* (1980). Elle fut appliquée à plusieurs processus (AGUADO, 1998; AGUILAR-MARTIN et LÓPEZ DE MÁNTARAS, 1982; CHAN *et al.*, 1989; PIERA CARRETÉ *et al.*, 1990; WAISSMAN, 2000). L'objectif de cette méthode repose sur la représentation de classes ou de concepts au moyen de la connexion logique de toutes les informations marginales disponibles. L'une des caractéristiques les plus importantes de la technique LAMDA réside dans sa capacité à modéliser naturellement l'indistingabilité totale ou l'homogénéité au travers d'une classe non informative appelée NIC « Non-Informative Class ». Cette classe accepte tous les objets présents dans l'espace de description avec le même degré d'adéquation. De cette façon, elle représente un seuil minimum de classification, ce qui montre que la classe NIC est toujours présente dans l'espace des classes déjà établies. De plus, cette méthode offre la possibilité de réaliser indifféremment un apprentissage séquentiel supervisé et non supervisé ainsi que de manipuler des informations de natures différentes : de natures numérique (quantitative), symbolique (qualitative) ou mixte.

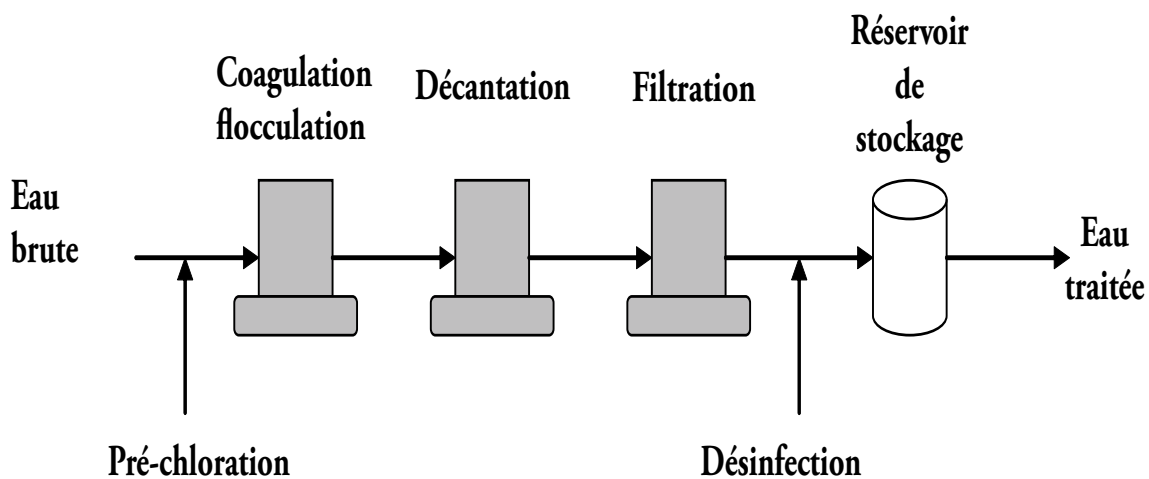


Figure 1. Chaîne de traitement de la station Rocade.
Water treatment operation of the Rocade plant

Ces deux caractéristiques ne se rencontrent pas simultanément dans les techniques de classification telles que celles par réseaux de neurones ou logique floue.

Ces informations sont des collections d'objets, décrits par un nombre η d'attributs appelés descripteurs. Chaque objet est décrit par un nombre fini et fixe de descripteurs représentés par le vecteur $\vec{x} = (x_1, x_2, \dots, x_n)$ où x_i est la valeur normalisée associée au descripteur i . Afin d'obtenir une confrontation entre \vec{x} et les différentes classes C_j , une fonction de vraisemblance, appelée Degré d'Adéquation Marginale (DAM), est calculée pour chaque classe et chaque espace de description. L'information donnée par le calcul des η DAMs est ensuite utilisée pour déterminer le Degré d'Adéquation Globale (DAG).

La figure 2 montre le schéma représentatif du DAG obtenu au moyen d'un opérateur logique d'agrégation appelé « connectif mixte de compensation linéaire », introduit et étudié par PIERA CARRETÉ et AGUILAR-MARTIN (1991). Ce connectif, issu de la théorie des ensembles flous, effectue une interpolation entre l'opérateur logique d'intersection (T-norme, T) et celui de l'union (T-conorme, S), par le biais de l'indice d'exigence, $\alpha \in [0 - 1]$. Le calcul du DAG se fait selon la formule :

$$\begin{aligned} \text{DAG}_\alpha(\text{DAM}_1, \text{DAM}_2, \dots, \text{DAM}_\eta) = \\ \alpha T(\text{DAM}_1, \text{DAM}_2, \dots, \text{DAM}_\eta) \\ + (1 - \alpha) S(\text{DAM}_1, \text{DAM}_2, \dots, \text{DAM}_\eta) \end{aligned} \quad (1)$$

La méthode LAMDA propose six algorithmes permettant de calculer les DAMs (KEMPOWSKY, 2004). Les résultats qui seront présentés dans cette étude sont obtenus en utilisant l'algorithme Lamda3 :

$$\begin{aligned} \text{DAM} \left[x_i \mid \rho_{j,i} \right] = \rho_{j,i}^{1-d_{j,i}} - (1 - \rho_{j,i})^{d_{j,i}} \\ \text{avec } d_{j,i} = \left| x_i - c_{j,i} \right| \end{aligned} \quad (2)$$

où $\rho_{j,i}$ est la probabilité de succès pour que l'événement stochastique $x_i \in C_j$ et $c_{j,i} \in D_i$ soit appelé centre de la classe C_j .

En fait, ceci n'est qu'une extension floue de la loi de probabilité binomiale à laquelle a été ajoutée la possibilité de créer une partition autour des centres de classes (WAISSMAN, 2000).

L'actualisation des paramètres associés aux descripteurs quantitatifs de la classe C_j se fait de la façon suivante :

$$\rho_{j,i} = \rho_{j,i} + \frac{x_i - \rho_{j,i}}{N + 1} \quad (3)$$

où N est le nombre d'objets attribué à cette classe. Pour procéder séquentiellement, il est nécessaire de connaître le nombre d'éléments ayant servi au calcul des paramètres de la classe correspondante.

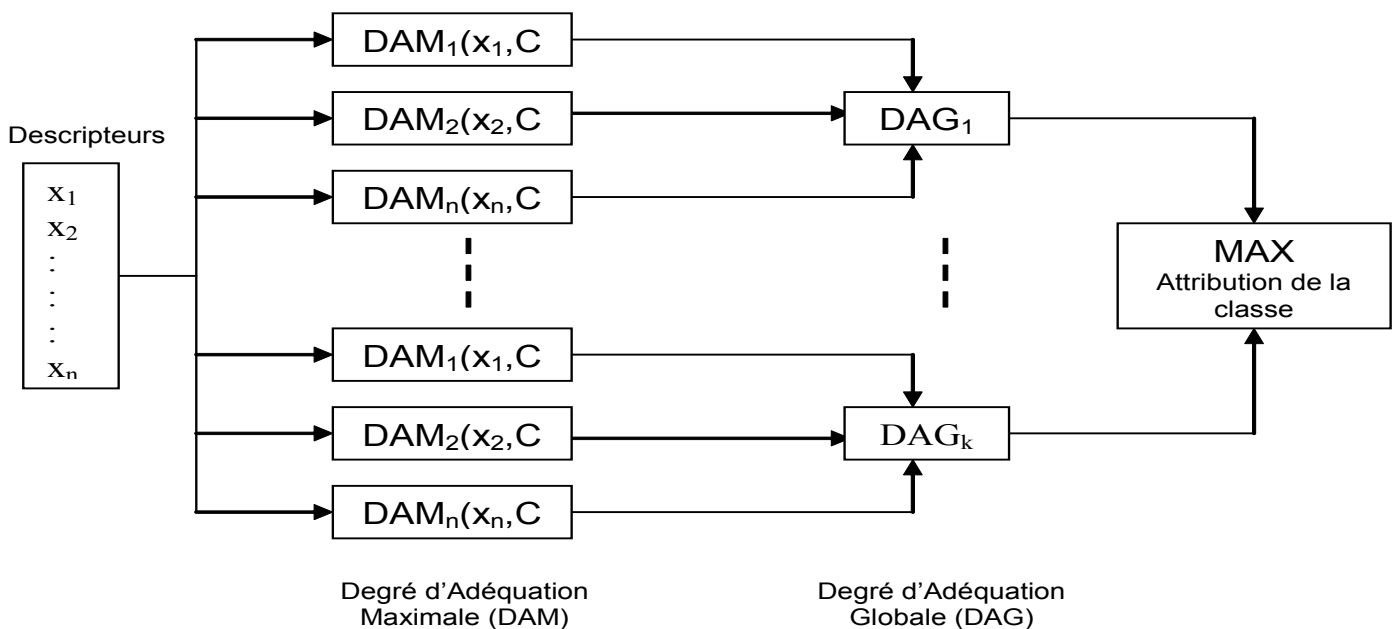


Figure 2. Degrés d'Adéquation Marginale et Globale. Degrees of marginal and global adequacy.

On peut créer aussi, à l'aide de la classe (NIC), une nouvelle classe qui va être caractérisée par l'affectation d'un élément à cette classe. X est le premier élément d'une nouvelle classe C_{j+1} et la représentation de cette nouvelle classe dépendra de cet élément. On prendra un paramètre fictif correspondant au nombre d'éléments de la classe NIC.

$$\rho_{j,0} = \rho_{j,0} + \frac{x_i - \rho_{j,0}}{N_0 + 1} \quad (4)$$

Dans le cas de l'auto-apprentissage, toute classe a dû être initialisée par la classe NIC, c'est pourquoi la formule de la mise à jour doit contenir ce paramètre fictif N_0 et elle devient :

$$\rho_{j,i} = \rho_{j,i} + \frac{x_i - \rho_{j,i}}{N_0 + N + 1} \quad (5)$$

Le paramètre $N_0 > 0$ détermine l'initialisation de l'apprentissage. Sa valeur peut être choisie arbitrairement mais elle influe sur le pouvoir d'absorption de chaque classe nouvelle; plus N_0 est grand, moins la classe nouvelle sera influencée par le premier élément. Par contre, en apprentissage dirigé, ce paramètre n'a pas d'influence sur le résultat de la classification.

Dans le cas de la reconnaissance de formes, l'élément est attribué à une classe significative ou rejeté dans la classe résiduelle NIC. Dans le cas de l'apprentissage, s'il est affecté à une classe significative, il y a modification des paramètres de cette classe. En revanche, si la classe vide a la plus grande adéquation, une nouvelle classe doit être créée pour contenir cet élément. Il y a rejet si la classe vide est la plus proche et qu'il n'y a pas possibilité de création de nouvelles classes parce que le nombre maximum de classes créées est atteint.

4. RÉSULTATS DE LA CLASSIFICATION

Dans cette étude, la classification des états fonctionnels est basée sur l'application itérative de la méthode LAMDA dont les étapes se résument comme suit :

1. Initialisation de la méthode par un ensemble de données (sans prendre en compte une quelconque connaissance préalable du procédé). À noter que les données à classer sont normalisées dans la gamme [0 - 1];
2. Application d'un apprentissage non supervisé (auto-apprentissage) : les objets ne sont pas affectés à des classes prédéfinies et seule la classe NIC existe initialement;

3. Identification, par l'expert, de l'ensemble des classes en rapport à un ensemble d'états significatifs. Ainsi, trois situations sont possibles : une classe est équivalente à un état, un ensemble de classes est équivalent à un état ou une classe n'a d'équivalence avec aucun état connu;

4. S'il existe au moins une classe dans la dernière situation, on doit retourner à l'étape 2 en considérant uniquement les données comprises dans cette classe;

5. Identification de toutes les séquences possibles d'états une fois que toutes les données sont classifiées en états connus;

6. Application d'un apprentissage supervisé à partir des données comprises entre chaque transition d'états possible afin d'obtenir toutes les conditions de changement. Dans ce cas, l'expert, selon ses propres critères, peut définir un certain nombre de classes significatives par un ensemble de données, et ainsi mettre en place une classification de référence.

Les données utilisées pour l'apprentissage sont issues de la station Rocado. Une base de données a été établie à partir de quatre années de fonctionnement de la station. Nous avons considéré comme descripteurs : la température (T), les MES, le pH, la conductivité (COND) et l'oxygène dissous (OD). Le choix de ces descripteurs a été dicté par l'extrême influence de leurs valeurs sur le procédé de coagulation. En effet, dans une étude précédente (LAMRINI *et al.*, 2005), nous avons développé un capteur logiciel, à base de réseaux de neurones, permettant de prédire la dose optimale de coagulant nécessaire dans une station de production d'eau potable. Le calcul de la dose est basé sur les caractéristiques de l'eau brute. Une étude par analyse en composantes principales (ACP) a permis de déterminer les cinq paramètres principaux de l'eau brute à prendre en considération lors du développement du capteur logiciel. De plus, l'étude effectuée sur le fonctionnement de la station Rocado nous a montré que la qualité de l'eau brute se caractérise par une turbidité très variable (très élevée à certaines périodes) et dès qu'elle dépasse la valeur 1 000 NTU, l'opérateur passe à la mesure des MES. Pour cette raison, et afin de tenir compte de toutes les variations de la qualité de l'eau brute, nous avons utilisé les MES au lieu de la turbidité. La dose de coagulant prédite par le capteur logiciel sera considérée comme un descripteur supplémentaire à utiliser dans la classification.

La figure 3 représente les caractéristiques d'eau brute utilisées pour l'apprentissage. Les données montrent que la qualité de l'eau brute est caractérisée par des MES très variables. Ainsi, on constate de fortes variations de MES durant la saison hivernale qui deviennent parfois très élevées pendant certaines périodes. Les valeurs de MES mesurées dépassent parfois 900 mg L⁻¹, et ce, pendant les crues. Le pH

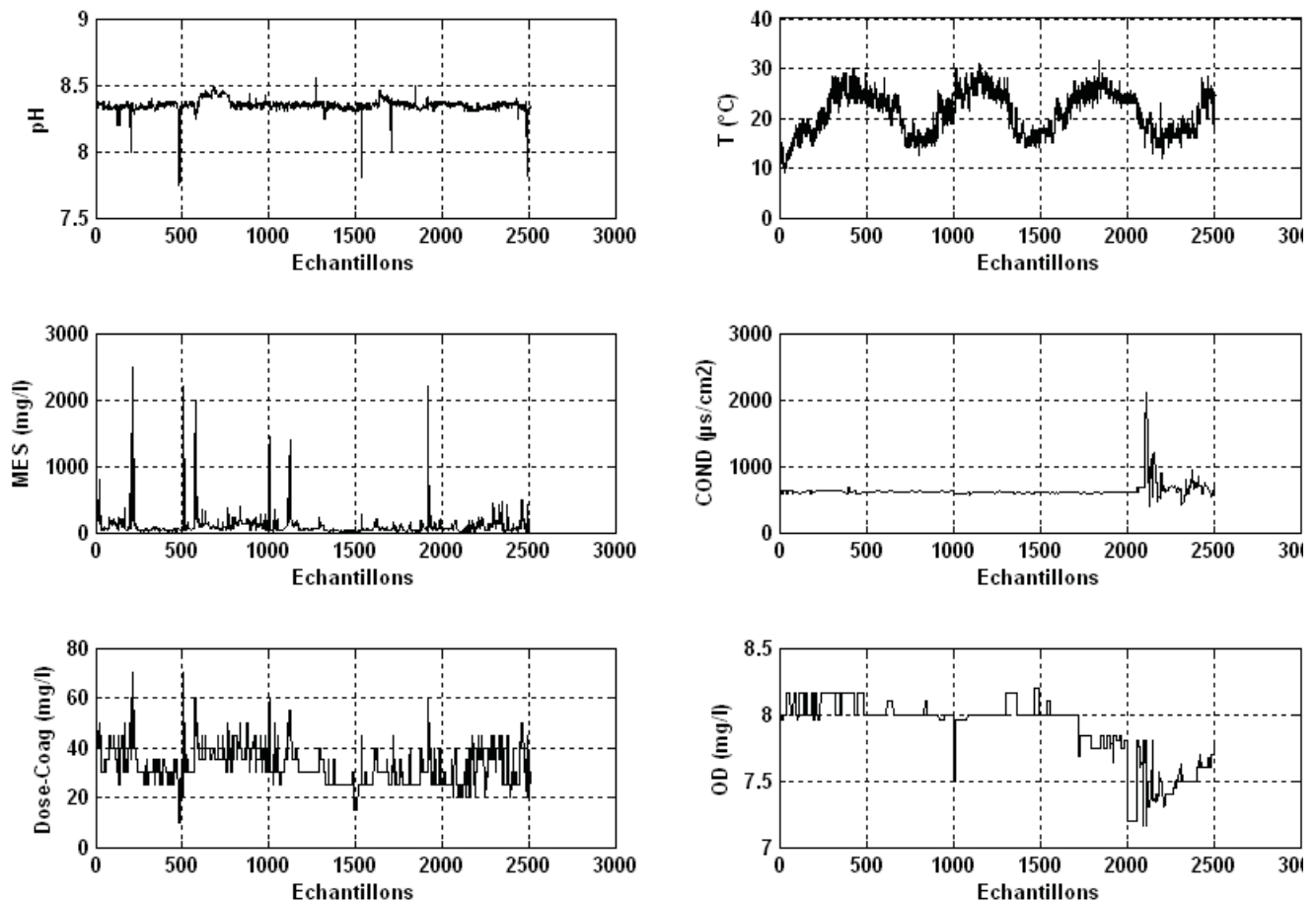


Figure 3. Caractéristiques de l'eau brute.
Raw water characteristics.

est élevé et atteint la valeur maximale 8,5. On constate aussi une forte variation (jusqu'à 2 200 $\mu\text{s cm}^{-2}$) de la conductivité, due à la présence des chlorures dans l'eau brute. Dans les conditions d'opération normale, l'oxygène dissous varie dans des proportions négligeables (8 à 8,2 mg L^{-1}). Cependant, une forte variation de l'oxygène dissous est détectée en parallèle avec des valeurs très élevées de la conductivité. En basse saison (saison hivernale), la qualité de l'eau ne nécessite que de faibles taux de coagulant (10 à 45 mg L^{-1}).

Afin de garder et de bien distinguer les signaux relatifs aux variations brusques et significatives de chaque descripteur, ces données brutes sont filtrées (Figure 4a) au moyen d'un filtre passe bas d'ordre 2 autour d'une fréquence de 0,5 Hz. Ensuite, pour mieux extraire les informations pertinentes pour la classification et connaître les conditions d'opération de chaque descripteur, nous avons utilisé un analyseur spectral en

sélectionnant une composante fréquentielle autour de laquelle le spectre est défini par les paramètres suivants :

- Fréquence sélectionnée : 0,01 Hz;
- Marge de tolérance (% Nyquist) : 10;
- Largeur de la fenêtre : 7;
- Déplacement de la fenêtre : 1;
- Période d'échantillonnage : 1.

Ce filtrage est effectué au moyen d'un outil d'abstraction des signaux appelé ABSALON (SARRATE, 2002).

Dans un premier temps, nous avons réalisé une classification sans recours à une quelconque connaissance préalable (classification avec auto-apprentissage). Dans ce cas, les objets ne sont pas déjà affectés à des classes prédéfinies et seule la classe NIC existe.

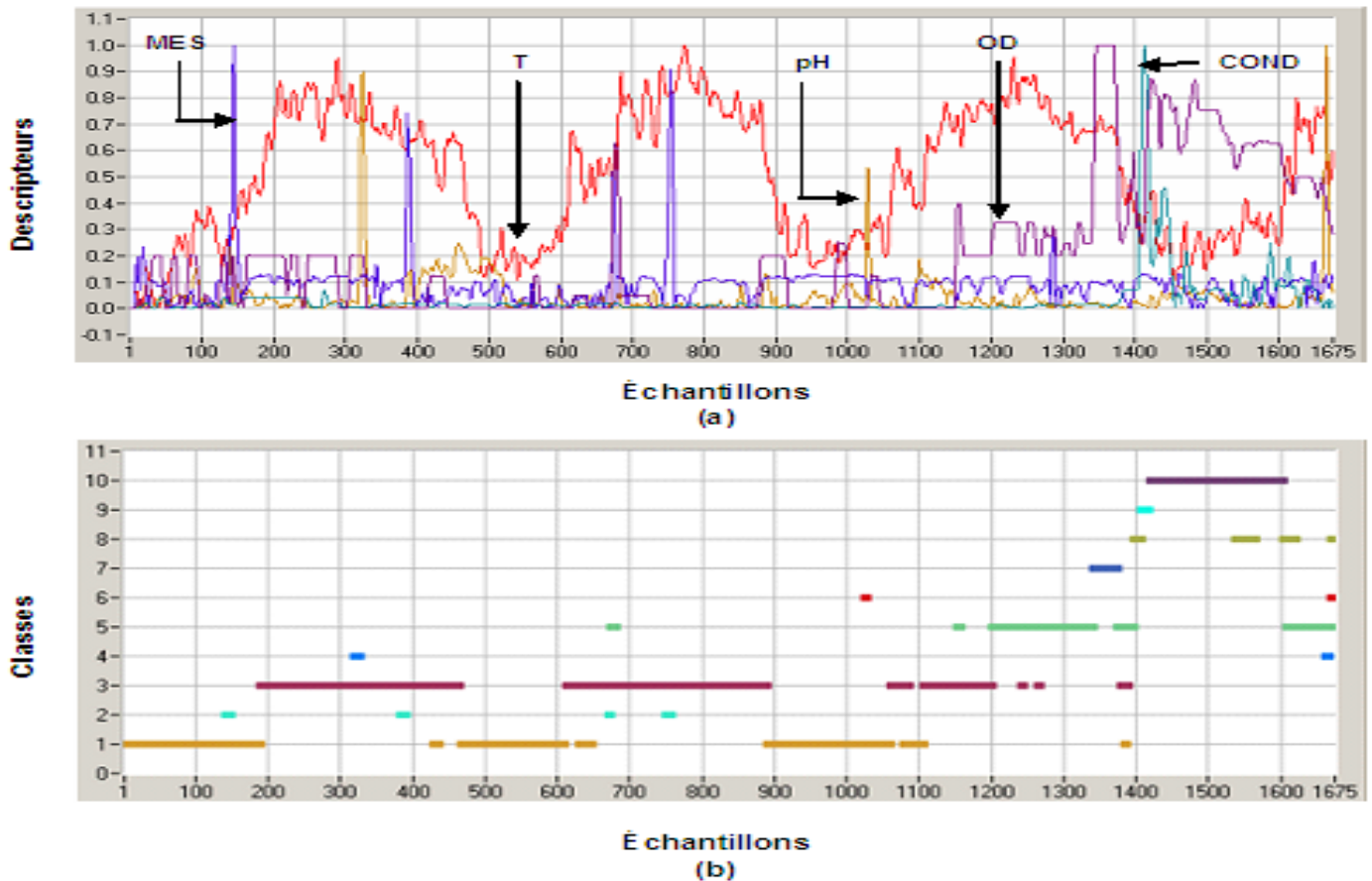


Figure 4. (a) Descripteurs normalisés, (b) Classes issues de l'auto-apprentissage.
(a) Normalised descriptors, (b) Self-learning classification results.

Nous avons procédé à des essais (LAMRINI *et al.*, 2005) pour les six algorithmes (KEMPOWSKY, 2004) et différents indices d'exigence. Les meilleurs résultats ont été obtenus avec l'algorithme Lamda3 (Éq. 3) comme fonction de présence utilisée pour calculer les adéquations marginales et le connectif Minimum/Maximum avec une exigence : $\alpha = 0,83$, pour le calcul des adéquations globales. Il est à noter que plus cet indice d'exigence est élevé, plus l'appartenance d'un objet à une classe est stricte, et plus le nombre d'objets non classés devient important. De ce fait, le nombre de nouvelles classes créées augmente et il est donc évident que le choix de ces différents paramètres, nécessaires à la réalisation d'un auto-apprentissage, doit être guidé par les connaissances de l'expert. La figure 4b illustre les 10 classes issues de l'auto-apprentissage. À noter que pour chaque descripteur i de l'élément $x = (x_1, \dots, x_n)$, la normalisation a été faite avec la formule suivante :

$$x_{i\text{norm}} = \frac{x_i - x_{i\text{min}}}{x_{i\text{max}} - x_{i\text{min}}} \quad (6)$$

La figure 5 présente le profil de chaque classe, i.e. la fréquence de contribution de chaque descripteur aux classes créées (Tableau 1). Les caractéristiques de chaque classe dans l'espace de description permettent à l'expert d'identifier les classes significatives et celles ayant des caractéristiques très proches. Ainsi, l'expert peut décider d'agréger ces classes dans un seul et même état. Sur cette figure, les descripteurs des classes apparaissent dans l'ordre suivant : Température, pH, MES, conductivité et oxygène dissous.

Décrivons brièvement ce profil de classes :

- Classes 1 et 3 : les descripteurs se trouvent dans des valeurs optimales de fonctionnement suivant deux saisons : la basse saison (saison hivernale) et la haute saison (saison estivale), cas de fonctionnement normal de la station. Ainsi, ces deux classes peuvent être associées à l'état de fonctionnement « Normal ». Les 549 éléments propres à la basse saison plus les 617 éléments propres à la haute saison ont été associés à ce même état de fonctionnement normal.

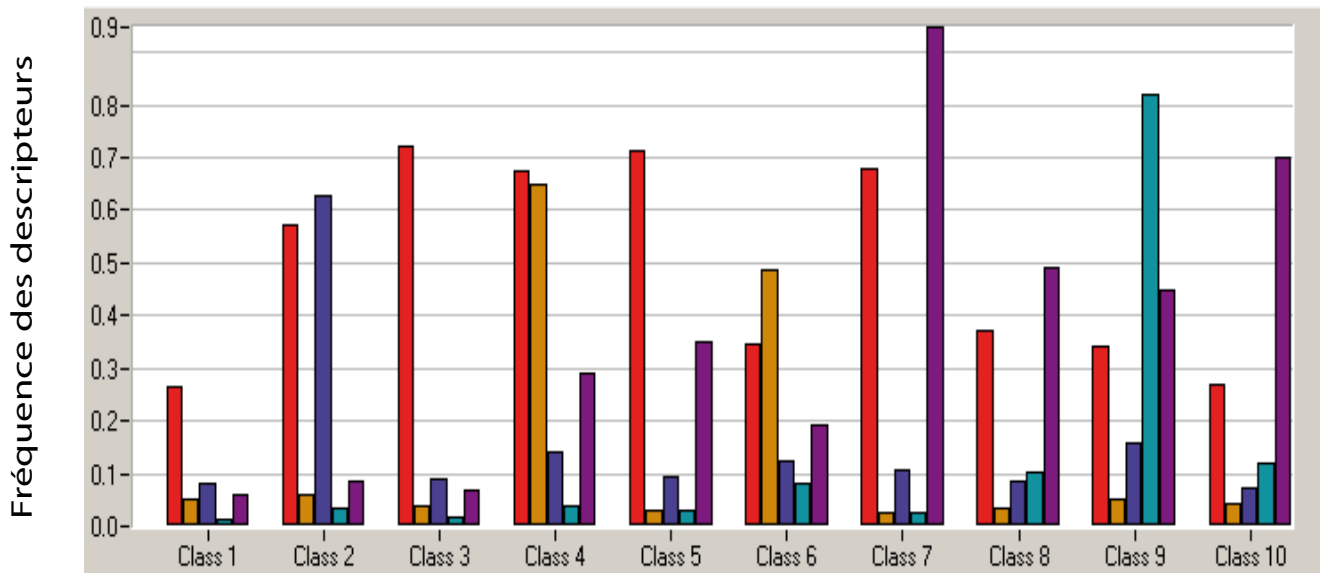


Figure 5. Profil des classes issues de l'auto-apprentissage.
Class profiles stemming from the self-learning procedure.

Tableau 1. Fréquence de contribution de chaque descripteur aux classes créées.
Table 1. Contribution frequency of each descriptor.

Classes	T	pH	MES	COND	OD
Classe 1	0,27	0,05	0,07	0,06	0,01
Classe 2	0,57	0,06	0,63	0,08	0,03
Classe 3	0,72	0,04	0,09	0,06	0,01
Classe 4	0,67	0,65	0,14	0,30	0,04
Classe 5	0,71	0,03	0,1	0,35	0,03
Classe 6	0,35	0,49	0,12	0,20	0,08
Classe 7	0,68	0,02	0,1	0,95	0,03
Classe 8	0,37	0,03	0,08	0,50	0,10
Classe 9	0,34	0,05	0,16	0,45	0,82
Classe 10	0,27	0,04	0,07	0,70	0,12

- Classe 2 : cette classe permet d'identifier toutes les variations élevées des MES ($MES_{norm} > 0,5$) ainsi que les limites de fonctionnement de la station, i.e. les variations qui peuvent provoquer soit l'arrêt ou le ralentissement de la station. La classe 2 peut être associée à l'état « MESAlarme+ » et les 26 éléments correspondants y ont été associés.
- Classes 4 et 6 : ces deux classes permettent de dégager successivement les 22 variations (propres à la basse et à la haute saison) du pH (vers un état très bas). Ces variations ont été identifiées simultanément avec une évolution de l'oxygène dissous vers son état optimal. En effet, le pH a plus d'influence dans la caractérisation de cette classe. Les deux classes peuvent être agrégées dans un seul état, soit « pHAlarme ».
- Classes 5 et 8 : ces deux classes permettent de détecter successivement, suivant les deux saisons, les variations de l'oxygène dissous vers un état bas. Les deux classes sont associées à l'état « ODAlarne » et les (206 + 44) éléments ont été intégrés à cet état.
- Classe 7 : cette classe permet d'isoler les 33 variations de l'oxygène dissous vers un état très bas (toutes les variations sont dans la gamme [0,7 - 1]). Nous associons cette classe à l'état « ODAlarne ».

- Classe 9 : cette classe a été générée à partir de 11 éléments. Elle identifie toutes les variations élevées de la conductivité. Nous constatons aussi une faible variation simultanée de l'oxygène dissous vers un état bas constant. La classe 9 est affectée à l'état « CONDAlarne ».
- Classe 10 : 167 éléments constituent la classe 10. Cette classe permet d'identifier l'évolution de l'oxygène dissous, en basse saison, vers un état très bas. Cette variation correspond à une faible variation de la conductivité. La classe 10 est associée à l'état « ODAlarne ».

Les résultats de classification issus de l'auto-apprentissage indiqués dans le tableau 2 montrent qu'on peut identifier cinq états fonctionnels. Ces états permettent d'identifier, sur l'unité de coagulation, les principales dérives de fonctionnement.

Cette classification optimisée a permis de définir la classification de référence. L'étape suivante du développement de l'outil de supervision/diagnostic a constitué à réaliser une classification passive, i.e. une reconnaissance pour valider cette affectation. Lors de l'affectation, les paramètres des classes ne sont pas modifiés, ce qui est normal puisqu'il s'agit d'une exploitation des connaissances et non d'un apprentissage. Les éléments non classés sont affectés à une classe appelée « classe résiduelle » (classe 0) ayant les mêmes caractéristiques que la classe NIC dans le cas de l'apprentissage. Les résultats sont présentés sur la figure 6. Sur cette figure, nous constatons que certains éléments (très peu) ont été soit mal reconnus (l'élément 138 est affecté à la classe 6 et les éléments 676,

677 et 678 sont affectés à la classe 8), soit non reconnus (c'est-à-dire n'appartenant à aucune classe prédéfinie); c'est le cas des éléments 674 et 675. Le tableau 3 représente le nombre d'éléments associés aux classes issues de l'auto-apprentissage et lors de la reconnaissance.

À partir de cette classification, il est possible de réaliser un apprentissage supervisé pour créer de nouvelles classes ou faire évoluer les classes déjà formées et ainsi arriver à une partition satisfaisante. Cet apprentissage supervisé a été effectué avec les mêmes valeurs pour les différents paramètres (fonction de présence, connectif, indice d'exigence) que celles utilisées lors de l'auto-apprentissage. Les figures 7 et 8 présentent les résultats de cet apprentissage supervisé. Nous avons pu détecter les mêmes états fonctionnels avec l'identification d'un nouvel état. La nouvelle classe (classe 11) identifie bien la transition des cinq descripteurs vers leur état optimal de fonctionnement. Cette classe est associée à l'état « Transition » et 10 éléments lui ont été associés. Il n'y a plus d'éléments non reconnus.

Le tableau 4 présente la comparaison des résultats propres aux deux classifications (résultats issus de l'auto-apprentissage et ceux issus de l'apprentissage supervisé). La compatibilité entre ces deux classifications peut être effectuée en calculant l'indice de dissemblance (LOPEZ DE MANTARAS, 1977) à partir de la matrice de contingence. Si on considère deux partitions F et G d'un même ensemble d'objets (Ω), les éléments de la matrice de contingence sont donnés par : $m_{ij} = \frac{\text{card}(F_i \cap G_j)}{\text{card}(\Omega)}$.

Tableau 2. Les états fonctionnels associés aux classes issues de l'auto-apprentissage.
Table 2. Functional states associated with the classes stemming from the self-learning procedure.

Classes	Nom de la classe	État fonctionnel	Nombre d'éléments associés
Classe 1	BSaison	Normal	549
Classe 2	MES_TrèsElevées---->limites	MESAlarne+	26
Classe 3	HSaison	Normal	617
Classe 4	pH_TrèsBas_HSaison	pHAlarne	16
Classe 5	OD_Bas_HSaison	ODAlarne	206
Classe 6	pH_Bas_BSaison	pHAlarne	6
Classe 7	OD_TrèsBas_HSaison	ODAlarne	33
Classe 8	OD_Bas_BSaison	ODAlarne	44
Classe 9	COND_TrèsBas_BSaison	CONDAlarne	11
Classe 10	OD_TrèsBas_BSaison	ODAlarne	167

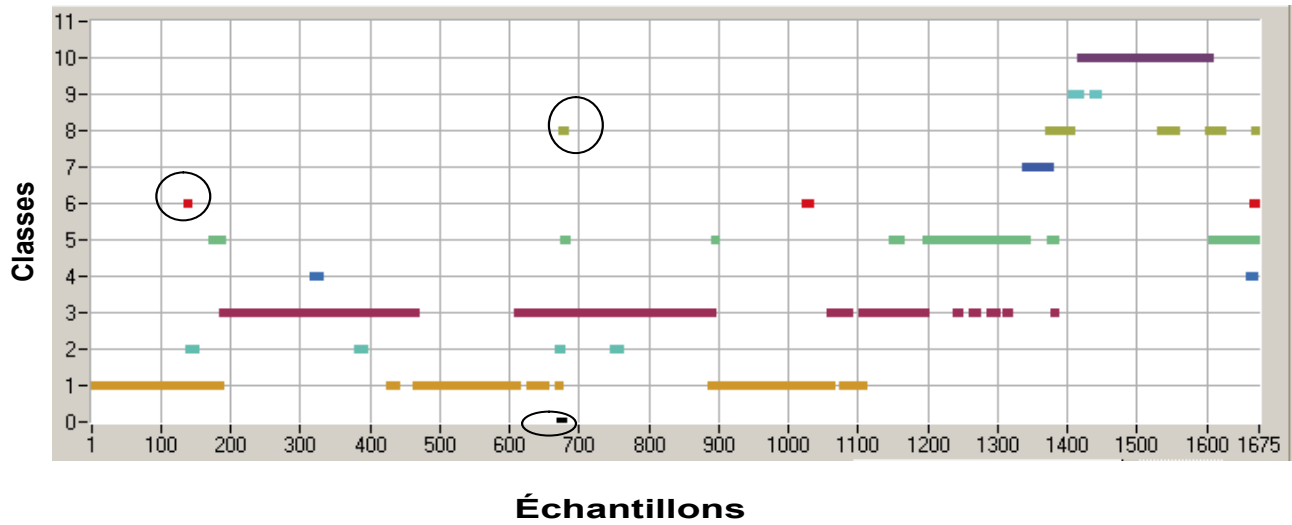


Figure 6. Résultats de la reconnaissance de formes.
Pattern recognition results.

Tableau 3. Les éléments associés aux classes issues de l'auto-apprentissage et de la reconnaissance de formes.

Table 3. Elements associated with the classes stemming from the self-learning procedure and pattern recognition.

Classes	Nombre d'éléments associés	
	Auto-apprentissage	Reconnaissance
Classe 1	549	536
Classe 2	26	27
Classe 3	617	615
Classe 4	16	15
Classe 5	206	192
Classe 6	6	10
Classe 7	33	33
Classe 8	44	64
Classe 9	11	17
Classe 10	167	164

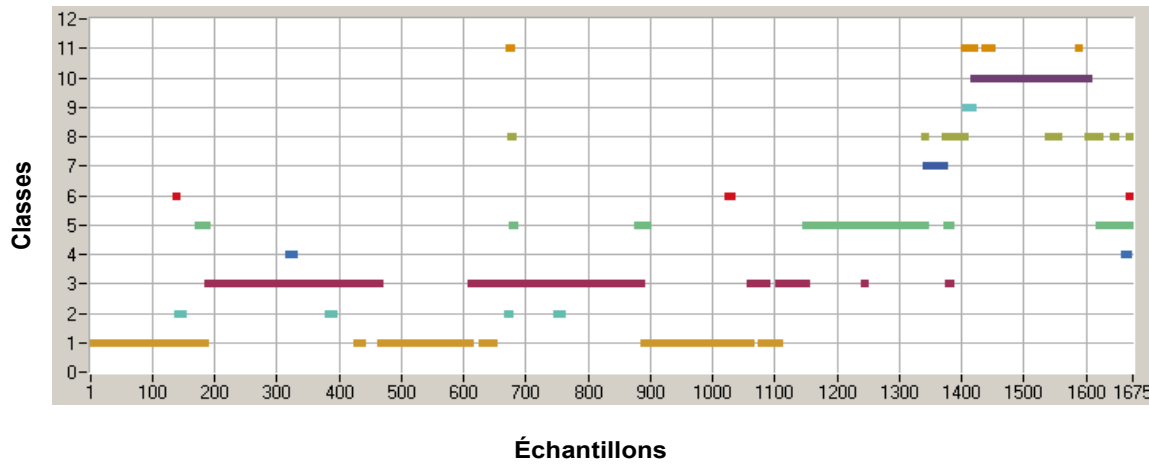


Figure 7. Création et modification des classes par apprentissage supervisé.
Design and modification of classes by supervised learning.

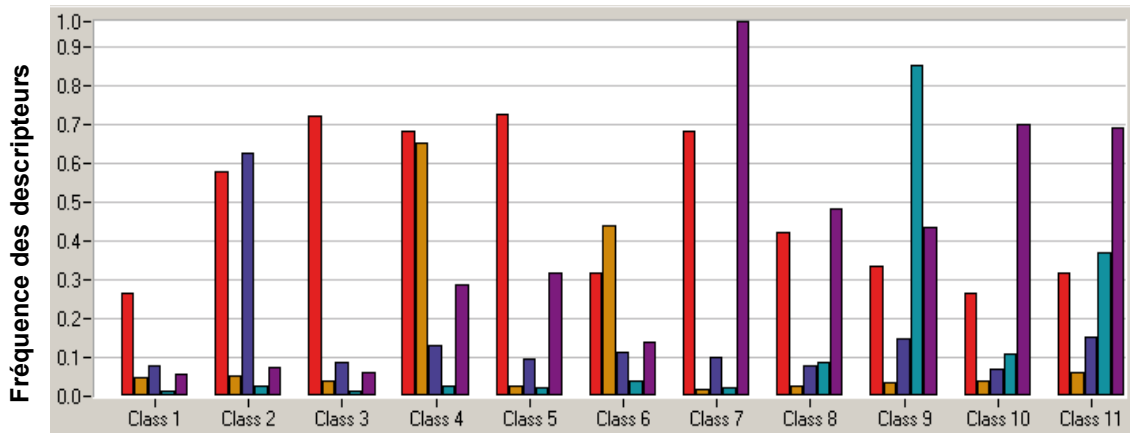


Figure 8. Profil de classes issues de l'apprentissage supervisé.
Class profiles stemming from supervised learning.

Tableau 4. Matrice de contingence (distance entre les deux classifications de 0.26).
Table 4. Contingency matrix (distance between the two classifications = 0.26).

Classes		Classification de référence									
		Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7	Classe 8	Classe 9	Classe 10
Classes issues de l'apprentissage supervisé	Classe 1	530	0	0	0	0	0	0	0	0	0
	Classe 2	0	26	1	0	0	0	0	0	0	0
	Classe 3	1	0	554	0	0	0	0	0	0	0
	Classe 4	0	0	0	16	0	0	0	0	0	0
	Classe 5	9	0	61	0	181	0	0	0	0	0
	Classe 6	4	0	0	0	0	6	0	0	0	0
	Classe 7	0	0	0	0	0	0	31	0	0	0
	Classe 8	5	0	1	0	23	0	2	33	0	0
	Classe 9	0	0	0	0	0	0	0	0	10	0
	Classe 10	0	0	0	0	0	0	0	9	0	154
	Classe 11	0	0	0	0	2	0	0	2	1	13

Les indices i et j représentent respectivement les nombres de classes des partitions F et G . Plus l'indice de dissemblance est faible, plus la compatibilité est bonne. Dans notre cas, cet indice est de 0,26, ce qui démontre une bonne compatibilité.

Dans ce qui suit, nous donnons les résultats relatifs à l'introduction d'un sixième descripteur. Ce descripteur supplémentaire est la dose de coagulant prédite par le capteur logiciel (LAMRINI *et al.*, 2005), précédemment développé sur la base de réseaux de neurones. En effet, dans les conditions d'opération normales de la station, la dose de coagulant appliquée évolue de façon non linéaire, en fonction des cinq descripteurs de la qualité de l'eau brute. Trois réseaux de neurones ont été développés suivant la variable utilisée en entrée, en plus du pH, de la température, de l'oxygène dissous et de la conductivité :

- Le réseau 1 qui utilise directement les MES;
- Le réseau 2 qui admet en entrée des valeurs de la turbidité limitées (inférieure à 1 000 NTU);

- Le réseau 3 qui utilise la turbidité avec une interpolation de celle-ci pour des valeurs supérieures à 1 000 NTU à l'aide d'une fonction d'interpolation des MES.

Il a été démontré que ces trois réseaux donnaient des résultats similaires et qu'il était possible de passer du réseau 2 au réseau 1 par exemple.

La démarche adoptée a été d'effectuer un apprentissage supervisé et d'adapter la classification réalisée en fonction de l'évolution des nouvelles données. La figure 9 présente les résultats de la classification obtenus après un apprentissage supervisé en adoptant les mêmes paramètres de classification que ceux initialement utilisés. Les résultats de classification montrent que l'introduction de la dose de coagulant comme descripteur supplémentaire fait apparaître un nouvel état de fonctionnement dégradé. Les classes 7 et 8 identifiant cet état de dysfonctionnement ont été associées à l'état « CoagulantAlarme ». Il correspond à un fonctionnement en basse saison où la qualité de l'eau ne nécessite que de faibles taux de coagulant (10 à 45 mg L⁻¹). De plus, l'introduction de

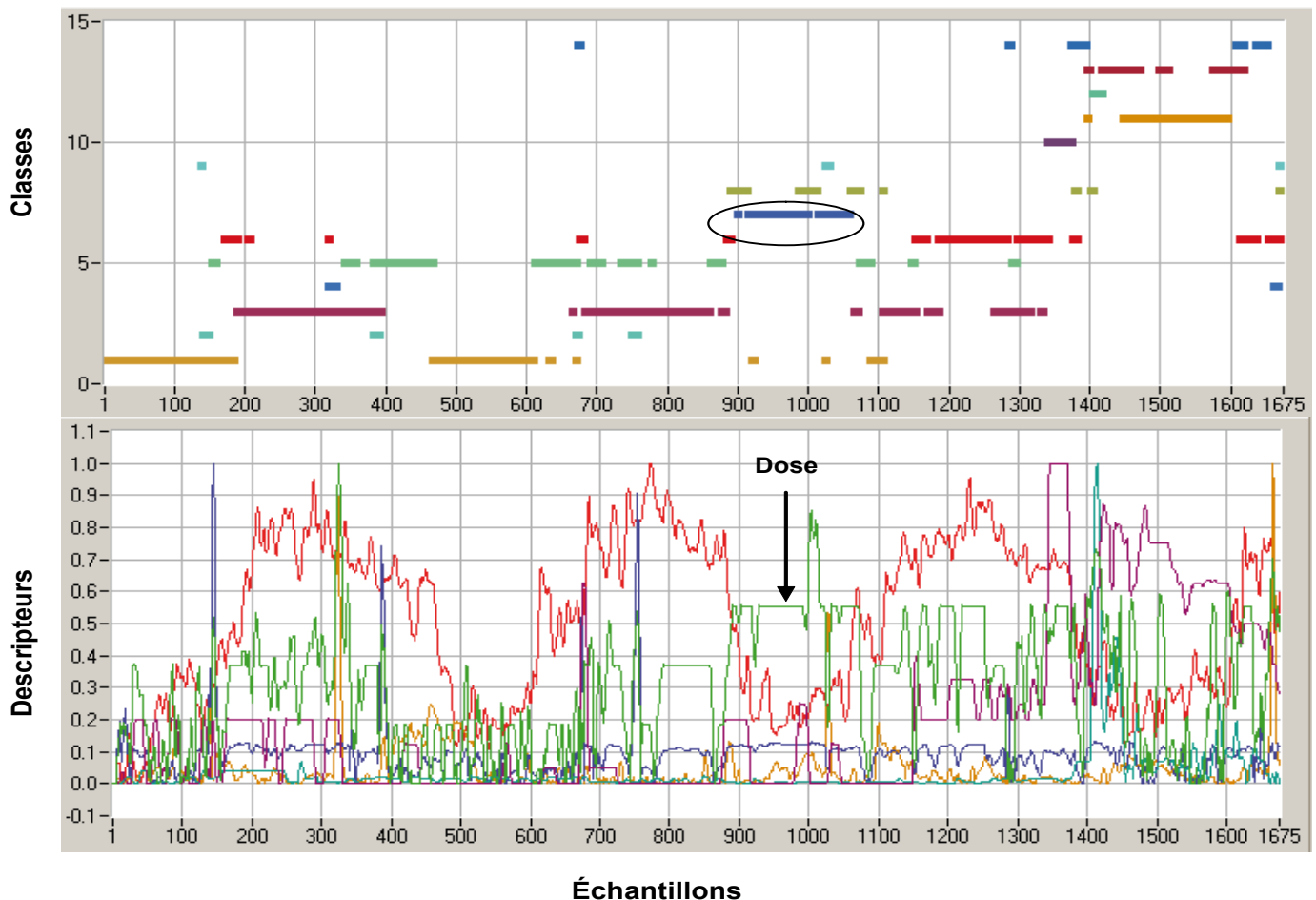


Figure 9. Classes issues de l'apprentissage supervisé avec la dose de coagulant
Classification stemming from supervised learning with coagulant dosage.

la dose de coagulant a permis d'affiner l'état de fonctionnement normal par la création d'une classe supplémentaire.

Les mêmes états fonctionnels identifiés précédemment sont retrouvés par les classes suivantes :

- Les classes 1, 3 et 5 caractérisent l'état de fonctionnement « Normal » suivant trois saisons : la basse saison, la mi-saison et la haute saison;
- La classe 2, qui permet d'identifier toutes les fortes variations des MES, est alors associée à l'état « MESAlarme+ »;
- Les variations du pH (vers un état très bas) identifiées par les classes 4 et 9 sont associées à l'état « pHAlarme »;
- L'état « ODAIarme » est associé aux quatre dérives de fonctionnement dégradé de l'oxygène dissous par les classes 6, 10, 11 et 13;
- La classe 12, caractérisée par de fortes variations de la conductivité, est affectée à l'état « CONDAIarme ».

5. CONCLUSION

Dans ce travail, la technique de classification par apprentissage LAMDA a été appliquée dans le but de réaliser une conduite supervisée du procédé de coagulation. Nous avons montré les performances de la méthode pour détecter les différentes séquences d'états fonctionnels décrivant le comportement de ce procédé. Les principaux avantages de cette méthode sont de permettre un apprentissage séquentiel supervisé et non supervisé et d'adapter la classification réalisée en fonction de l'évolution des données à classer. En outre, cette méthode se différencie des autres techniques par la notation de l'adéquation nulle et l'incidence de celle-ci sur la classification. Dans la base de données, trois types d'états fonctionnels ont été reconnus : des états de fonctionnement normal (la station se trouve la majorité du temps dans des conditions d'opération normales), des états de fonctionnement dégradé identifiés hors opération normale et d'autres états fonctionnels vers lesquels il est possible de transiter, en revenant toujours après un temps plus ou moins long, à l'état d'opération normal. À noter que les résultats obtenus peuvent servir de base pour le suivi et la détection des défaillances au niveau des procédés de coagulation. Une reconnaissance en ligne peut être envisagée à partir de ce modèle de référence pour déterminer l'état fonctionnel représentant le procédé une fois qu'une nouvelle observation se présente. En effet, le modèle de référence n'est pas toujours exhaustif du fait que, pour certaines situations, soit le recueil des données n'a pas été possible ou qu'il existe des situations dont l'occurrence est peu fréquente. Il est donc possible de faire évoluer la base d'apprentissage au fur et à mesure de l'apparition de nouveaux cas, en réalisant des apprentissages dits spécialisés

permettant de créer de nouvelles classes à partir d'éléments non reconnus.

Des travaux sont actuellement réalisés pour permettre la validation automatique de la classification obtenue. Jusqu'à présent, seul l'expert était juge de la validité de la classification obtenue. À l'aide d'indices de mesure de compacité et de séparation des classes exploitant le concept d'entropies intraclasse et interclasse, il est possible de faire évoluer l'indice d'exigence, par exemple, afin d'obtenir de manière quasi automatique la classification avec les meilleures caractéristiques en matière de compacité et de séparation des classes. De plus, pour éviter des transitions vers des classes non représentatives (lorsque l'élément en question a des DAM à deux classes très peu différents par exemple), des études sont menées actuellement pour ne prendre en compte que des transitions significatives. Ces deux modifications permettront à terme d'obtenir plus facilement une classification plus compacte et d'éviter d'avoir des transitions non représentatives.

En se référant aux travaux réalisés dans ce domaine, il semble que l'application de l'approche, réalisée dans le cadre de cette étude, est la première de son genre qui ait pu montrer l'utilité des techniques de classification dans le suivi et la surveillance de ce type de procédé. Il est clair que l'objectif final de l'étude est de pouvoir étendre cette surveillance à l'ensemble de la station de traitement Rocado, de traiter toutes les informations issues des différents procédés de traitement (décantation, filtration, etc.) et de détecter le plus tôt possible les dérives de fonctionnement. De telles informations peuvent aider à identifier les pannes sur une unité située en amont de la station mais aussi servir à la maintenance préventive des appareils comme les filtres ou les décanteurs.

RÉFÉRENCES BIBLIOGRAPHIQUES

- AGUADO J.C. (1998). *A mixed qualitative-quantitative self-learning classification technique applied to situation assessment in industrial process control*. Thèse de Doctorat, Univ. Politècnica de Catalunya, Terrassa, Espagne, 160 p.
- AGUILAR-MARTIN J., M. BALSSA. et R. LOPEZ DE MANTARAS (1980). *Estimation récurrente d'une partition. Exemples d'apprentissage et auto-apprentissage dans R^n et I^n* , LAAS-CNRS, Toulouse, France, Rapport scientifique N° 880139, 7 p.
- AGUILAR-MARTIN J. et R. LÓPEZ DE MANTARAS (1982). *The process of classification and learning the meaning of linguistic descriptors of concepts. Approximate reasoning in decision analysis*. Approximate reasoning in decision

- analysis, MM. GUPTA et E. SANCHEZ (Éditeurs), Amsterdam, Pays-Bas, 165-175.
- BAZER-BACHI A., E. PUECH-COSTE, R. BEN AIM et J.L. PROBST (1990). Modélisation mathématique du taux de coagulant dans une station de traitement d'eau. *Rev. Sci. Eau*, 3, 377-397.
- CHAN M., J. AGUILAR-MARTIN, N. PIERA CARRETÉ, P. CELSIS et J.P. MARC-VERGNES (1989). Classification techniques for feature extraction in low resolution topographic evaluative images: Application to cerebral blood flow estimation. *12th conference GRESTI*, 12-16 juin, Juan-les-Pins, France.
- EVANS J. et C. ENOCH (1998). Investigation of advanced control methods applied to coagulation dosing at a WTW. *UK Water Industry Research Limited/Sensors for Water Interest Group (UKWIR/SWIG) Workshop*, 9-11 décembre, National Motor Cycle Museum, Birmingham, Grande-Bretagne.
- GAGNON C., B.P.A. GRANDJEAN et J. THIBAUT (1997). Modelling of coagulant dosage in a water treatment plant. *Artif. Intell. Eng.*, 11, 401-404.
- KAN C. et C. HUANG (1998). Coagulation monitoring in surface water treatment facilities. *Water Sci. Technol.*, 38, 237-244.
- KEMPOWSKY T. (2004). *Surveillance de procédés à base de méthodes de classification : conception d'un outil d'aide pour la détection et le diagnostic des défaillances*. Thèse de Doctorat, Institut National des Sciences Appliquées, Toulouse, France, 170 p.
- LAMRINI B., A. BENHAMMOU, M.V. LE LANN et A. KARAMA (2005). A neural software sensor for on-line prediction of coagulant dosage: Application to a drinking water treatment plant. *Trans. Inst. Meas. Control (TIMC)*, 27, 195-213.
- LAMRINI B., M.V. LE LANN et A. BENHAMMOU (2005). Detection of functional states based on interactivity between expertise and LAMDA learning technique. *Information and Communication Technologies International Symposium (ICTIS'05)*, 3-6 juin, Tetouan, Maroc, pp. 105-110.
- LAMRINI B., M.V. LE LANN et A. BENHAMMOU (2005). Conduite supervisée d'une station de traitement d'eau potable par la méthode de classification LAMDA. *10^e congrès de la société française des génies de procédés (SFGP)*, 20-22 septembre, Toulouse, France, Auteurs I3.
- LIU J.C. et M.D. WU (1997). Fuzzy control of coagulation reaction through streaming current monitoring. *Water Sci. Technol.*, 36, 127-134.
- LOPEZ DE MANTARAS R. (1977). *Auto-apprentissage d'une partition. Application au classement itératif de données multidimensionnelles*. Thèse de Doctorat, Université Paul Sabatier, Toulouse, France, 150 p.
- PIERA CARRETÉ N. et J. AGUILAR-MARTIN (1991). Controlling selectivity in non-standard pattern recognition algorithms. *IEEE Trans. Syst. Man Cybern.*, 21, 71-82.
- PIERA CARRETÉ N., P. DESROCHES et J. AGUILAR-MARTIN (1989). *LAMDA: An incremental conceptual clustering system*. LAAS-CNRS, Toulouse, France. Rapport scientifique N° 89420, 10 p.
- PIERA CARRETÉ N., P. DESROCHES et J. AGUILAR-MARTIN (1990). Variation points in pattern recognition. *Patt. Recognit. Lett.*, 11, 519-524.
- SARRATE R. (2002). *Supervisió intelligent de processos dinàmics basada en esdeveniments (ABSALON: Abstraction analysis on-line)*. Thèse de Doctorat, Univ. Politècnica de Catalunya, Terrassa, Espagne, 145 p.
- SHEPHERD A. et L. ORTOLANO (1996). Water-supply operations: Critiquing expert-system approach. *J. Water Resour. Plann. Manage.*, 122, 348-355.
- STANDEN G., P.J. INSOLE, K.J. SHEK et R.A. IRWIN (1997). The use of particle monitoring in the performance optimisation of conventional clarification processes. *Water Sci. Technol.*, 36, 191-198.
- VALENTIN N., F. FOTOOHI, T. DENOEU (1999). Modeling of coagulant dosing in a water treatment plant. *Proceedings of EANN'99*, septembre, Warsaw, Pologne, pp.165-170.
- WAISSMANN J. (2000). *Construction d'un modèle comportemental pour la supervision de procédés : application à une station de traitement des eaux*. Thèse de Doctorat, Institut National Polytechnique de Toulouse, France, 119 p.
- ZHU X. et A.R. SIMPSON (1996). Expert system for water treatment plant operation. *J. Environ. Eng.*, 122, 822-829.