MAITRISE DES RISQUES DANS LE DOMAINE DE L'EAU POTABLE

Walter SCHÖN[1,2], Khaled ODEH[1,3], Thierry DENOEUX[1,2], Farrokh FOTOOHI[1,3]

(1) : Laboratoire SIME (Systèmes Intelligents pour les Métiers de l'Eau)
Université de Technologie de Compiègne (UTC) BP 20529
60205 COMPIEGNE CEDEX FRANCE

(2) : Laboratoire Heudiasyc (Heuristique et Diagnostic de Systèmes Complexes)
Université de Technologie de Compiègne BP 20529
60205 COMPIEGNE CEDEX FRANCE

(3) : Centre International de Recherche sur l'Eau et l'Environnement
Pôle Informatique Technique - CITI
59, Avenue Emile Thièbaut
78110 LE VESINET FRANCE

Résumé

La maîtrise des risques dans le domaine des métiers de l'eau est bien sûr en premier lieu la maîtrise de la qualité sanitaire de l'eau potable distribuée, objet d'une réglementation très stricte des pouvoirs publics. C'est également la prise en compte d'une exigence de disponibilité très élevée de la part des usagers, comparable à celle concernant la distribution d'électricité. C'est enfin l'intégration de tous ces aspects dans une démarche globale de maîtrise des coûts d'exploitation et d'investissement dans un contexte où la demande et les niveaux d'exigence ne cessent de croître. Cette communication présente les travaux communs dans ce domaine entre l'UTC/CNRS et Suez Lyonnaise des eaux, dans le cadre d'une collaboration engagée de longue date et qui se prolonge aujourd'hui par la création d'un laboratoire de recherches commun.

Introduction

Le domaine de la distribution d'eau potable présente des caractéristiques bien spécifiques qu'il est indispensable de prendre en compte dans la construction d'une démarche globale de maîtrise des risques :

- L'aspect sécurité à savoir la qualité sanitaire de l'eau distribuée est évidemment fondamental, dans la mesure où tout incident dans ce domaine touche potentiellement une population importante à l'échelle d'une ville entière

- L'aspect disponibilité à savoir la satisfaction de la demande aux points de consommation est également primordial et se traduit par des exigences comparables à celles en usage dans le domaine de la distribution d'énergie électrique, soit des disponibilités de l'ordre de 99,9%.

La correcte prise en compte de ces aspects constituant un facteur clé de la maîtrise des coûts d'exploitation (dimensionnement du débit disponible, détermination des traitements adaptés…) ainsi que d'investissement (surdimensionnements, mise en place de redondances, planification des renouvellements…) est donc un enjeu majeur pour Suez Lyonnaise des Eaux et un thème central des travaux menés en commun avec le laboratoire Heudiasyc (Heuristique et Diagnostic de Systèmes Complexes, Unité Mixte de Recherches CNRS) de l'Université de Technologie de Compiègne. Cette collaboration, née il y a une dizaine d'années et concrétisée récemment par la création du laboratoire commun SIME (Systèmes Intelligents pour les Métiers de l'Eau), exploite l'adéquation entre les besoins de Suez Lyonnaise des Eaux et les pôles de compétences du laboratoire Heudiasyc dans les domaines tels que le diagnostic, la validation de données, la gestion des connaissances et la sûreté de fonctionnement.

Aspect sécurité : maîtrise de la qualité sanitaire

La qualité sanitaire de l'eau distribuée est un problème de santé publique et fait à ce titre l'objet de textes réglementaires de plus en plus stricts. Ainsi les lois imposent elles de nombreux paramètres à prendre en compte (actuellement 62) pour lesquels des concentrations maximales admissibles sont fixées à un seuil très bas prenant en compte des critères de toxicité à long terme. La surveillance continue de la qualité de l'eau brute (rivière ou nappe phréatique) ainsi que du fonctionnement des installations de traitement est également exigée.

Dans ce cadre, l'un des thèmes de recherches majeurs du laboratoire SIME concerne la mise en œuvre des technologies les plus récentes en matière de traitement de l'information, notamment la logique floue et les réseaux de neurones artificiels, afin d'améliorer la surveillance de la qualité de l'eau brute ainsi que du processus de traitement [1]. Les paramètres à prendre en compte sont en effet nombreux, en partie redondants (certains étant liés), peuvent éventuellement être manquants ou erronés en cas de défaillances de capteurs, de sorte que l'ensemble est difficilement directement exploitable par un opérateur. L'idée consiste d'une part à valider les données brutes en exploitant les corrélations afin de reconstruire les éventuelles données manquantes ou aberrantes, et d'autre part à élaborer une information synthétique permettant à l'exploitant de juger globalement de la qualité de l'eau, de déterminer le traitement adapté voire d'interrompre la distribution en cas de pollution importante.

La figure 1 illustre le principe d'un système de fusion de données tel qu'il a été développé dans le cadre du projet ESPRIT EM^{2}S (Environmental Monitoring and Management System) et implémenté sur plusieurs sites d'exploitation de Suez Lyonnaise des Eaux. L'objectif

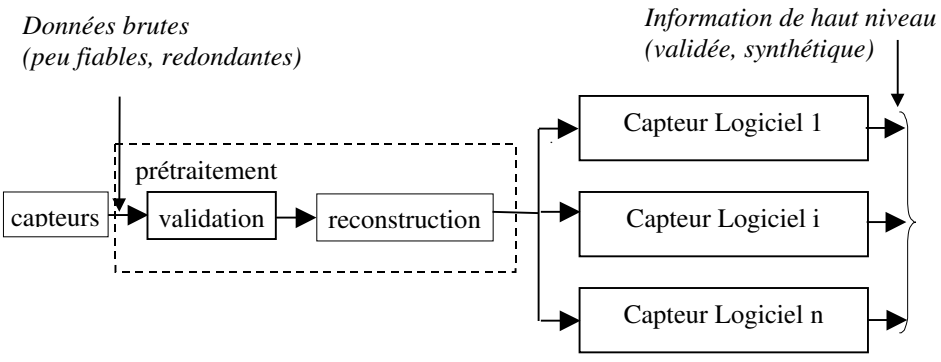

Figure 1 : Architecture d'un système de fusion de données

général d'un tel système est la transformation de données brutes, par essence peu fiables, peu compréhensibles et redondantes, en une information de haut niveau validée, synthétique et donnant à l'exploitant une vision précise de la qualité de l'eau en entrée de l'usine de production d'eau potable ainsi qu'aux différentes phases du traitement. Cette transformation peut être décomposée en trois étapes correspondant à la validation des données, à la reconstruction des valeurs manquantes, et à l'élaboration d'une information de haut niveau au moyen de capteurs logiciels.

- L'étape de validation a pour but d'améliorer la *robustesse* du système de traitement des données vis à vis de dysfonctionnements de la chaîne d'acquisition des informations. Son principe consiste à associer à chaque valeur mesurée un *indice de confiance* variant entre zéro et un, et traduisant la possibilité que la donnée soit invalide. Le calcul de cet indice s'appuie sur une *connaissance a priori* (un *modèle*) décrivant des contraintes supposées être vérifiées par les données valides. Le plus souvent, la connaissance disponible réside dans une base de données, validées visuellement par l'utilisateur. Le modèle construit est alors de type statistique.

- L'étape de reconstruction consiste à exploiter l'information statistique contenue dans la base de données de manière à estimer les composantes manquantes d'un vecteur de mesures. Ce problème peut être traité dans un cadre probabiliste : il s'agit alors d'estimer les paramètres d'une loi de probabilité à partir de données comportant éventuellement elles-mêmes des valeurs manquantes, en utilisant par exemple l'algorithme EM (Expectation-Maximisation) ; les valeurs manquantes sont alors remplacées par leur loi de probabilité conditionnelle estimée, ou par un indicateur de position de cette loi. Une autre approche [2, 3] consiste à s'appuyer sur un système neuro-flou construit par apprentissage à partir des données. Les valeurs manquantes sont alors remplacées par des nombres flous dont la fonction d'appartenance traduit l'incertitude associée à la reconstruction, et dont la propagation dans les calculs ultérieurs est plus facile que dans le cadre probabiliste. Dans les deux cas, la qualité de la reconstruction est évidemment fonction de la quantité d'information disponible ; elle doit impérativement être reflétée par la forme d'une distribution de valeur (de type probabiliste ou possibiliste), et non seulement par une valeur unique. La figure 2 illustre un exemple de données

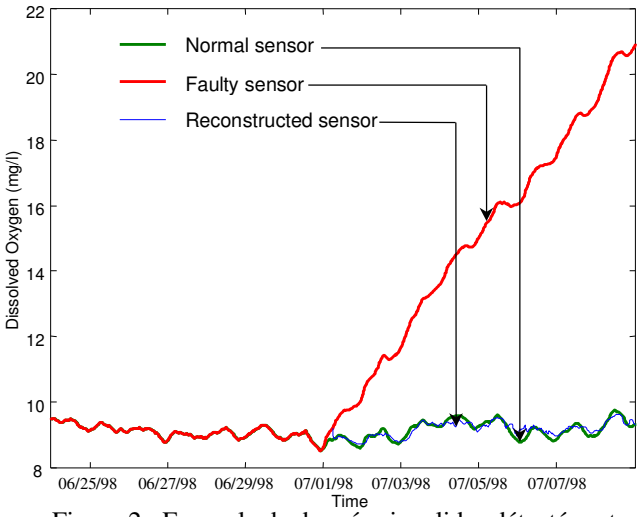

Figure 2 : Exemple de données invalides détectées et reconstruites. La dérive du capteur défaillant (faulty sensor) est détectée. La courbe reconstruite (reconstructed sensor) est très proche de la courbe normale (normal sensor).

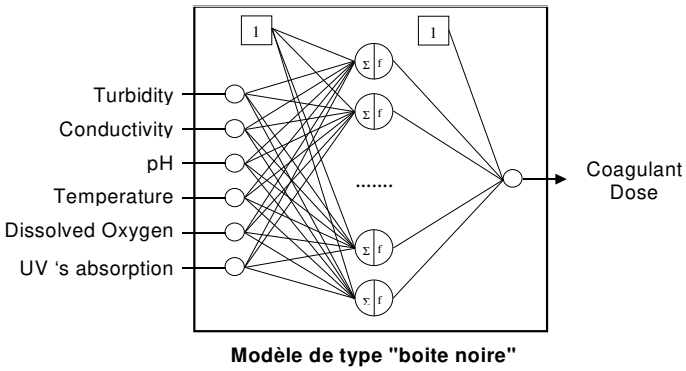

Figure 3 : Réseau de neurones pour la prédiction du taux de coagulants nécessaire à la clarification à partir de six paramètres physiques de l'eau

aberrantes générées artificiellement (dérive dans le temps d'un capteur d'oxygène dissous), ainsi que leur détection et leur reconstruction [4].

- Le processus d'élaboration, à partir de données validées et complétées, d'une information de haut niveau peut être structuré sous la forme de *capteurs logiciels* dont la fonction consiste à fournir à l'utilisateur une information complexe non observable directement, telle qu'un paramètre physico-chimique non mesurable en continu (ou à un coût prohibitif), un indice global de qualité ou de traitabilité de l'eau, ou une consigne pour l'ajustement des procédés de traitement (taux de coagulants). La construction d'un capteur logiciel peut être essentiellement vue comme un problème d'apprentissage de relations entre variables à partir de données observées, pour lequel les réseaux de neurones constituent un outil particulièrement performant.

A titre d'exemple, la figure 3 illustre un exemple de réseau de neurones pour la prédiction du taux de coagulants dans une usine de production d'eau potable [5, 6]. L'une des étapes les plus délicates dans le traitement des eaux de surface concerne en effet la clarification, qui consiste à éliminer les particules en suspension dans l'eau. Pour cela, on injecte des réactifs (sels d'aluminium, polymères) qui ont pour effet de provoquer l'agglutinement (coagulation, floculation) des particules en suspension. La détermination de la dose de réactifs à injecter est délicate, le dosage optimal des réactifs étant lié de manière complexe aux caractéristiques de l'eau brute. Il est le plus souvent effectué manuellement : une fois par jour, on prélève un échantillon d'eau brute, et on détermine le dosage optimal en laboratoire (jar test), afin de déterminer une consigne de traitement jusqu'au lendemain. En plus de son coût, cette méthode présente l'inconvénient de ne pas permettre une adaptation rapide à des modifications soudaines des caractéristiques de l'eau.

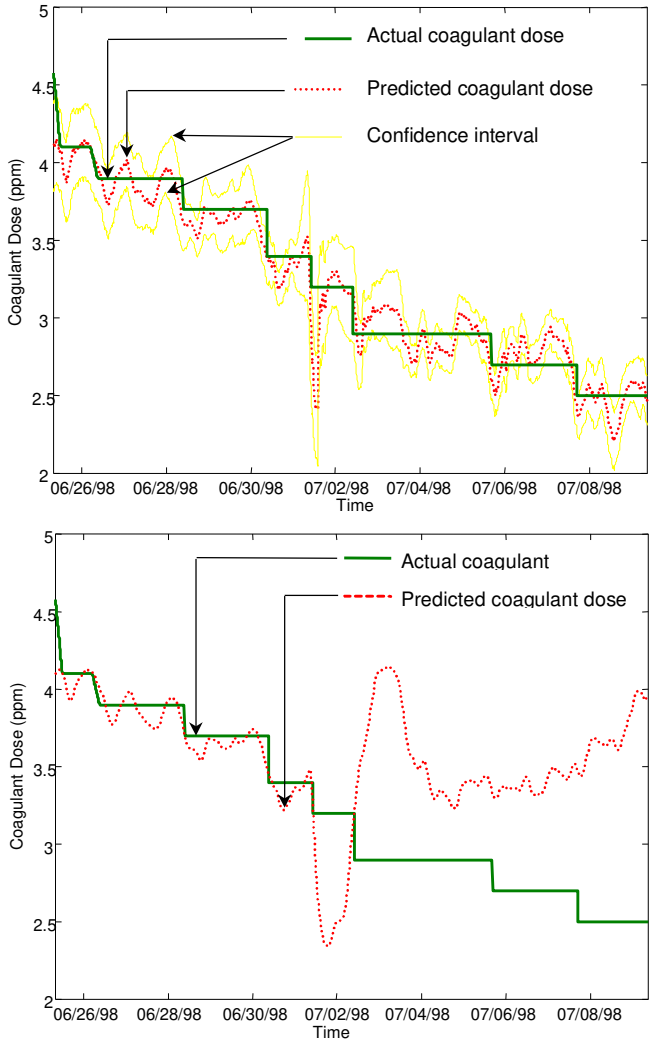

Figure 4 : Dose de coagulants prescrite par le système (predicted coagulant dose) et par la méthode de référence : jar test (actual coagulant), avec (en haut) et sans (en bas) validation des données d'entrée.

Le réseau de neurones illustré par la figure 3 et mis en place sur l'usine de Viry en Région Parisienne permet d'estimer automatiquement la dose de coagulants à injecter, à partir de mesures effectuées en continu sur l'eau brute. Une qualité fondamentale d'un tel système est sa robustesse par rapport à des erreurs de mesure ou à des modifications inattendues de la qualité de l'eau, d'où l'intérêt des méthodes de prétraitement (validation et reconstruction) mentionnées précédemment. La figure 4 montre en fonction du temps la dose de coagulant prescrite par le système avec et sans validation des variables d'entrée, démontrant ainsi la robustesse de la chaîne de traitement développée.

Aspect disponibilité : maîtrise de la qualité de service rendue à l'usager

Spécificités des métiers de l'eau

Un réseau d'eau potable (figure 5) est constitué de ressources naturelles (rivières ou forages), usines de traitement, stations de pompage, stockages à niveau ou en hauteur, points de consommation reliés entre eux par un réseau de canalisations et de vannes. Pour un tel système, de grande étendue géographique, en grande partie enterré et sur lequel les informations disponibles (provenant de capteurs répartis sur le réseau) sont peu nombreuses, l'obtention d'un niveau de disponibilité élevé représente une véritable gageure, nécessitant des études prévisionnelles approfondies prenant en compte les défaillances pouvant se produire dans des conditions d'environnement données.

De plus la notion même de disponibilité peut prendre plusieurs aspects selon le niveau d'exigence pris en compte. On distingue principalement :

- Disponibilité mécanique ou connexité du réseau : mesurée par la probabilité qu'à un instant donné on soit capable de relier tout point de consommation du réseau à au moins une source en état d'assurer une production [7].

- Disponibilité du flux de demande : mesurée par la probabilité que le réseau soit en état de satisfaire un niveau de demande donné en tout point de consommation [8]

- Disponibilité hydraulique : mesure la probabilité de satisfaire la demande en tout point à un niveau de pression défini [9].

Modélisation du réseau

Si la disponibilité mécanique peut être traitée par les méthodes classiques d'analyse de la sûreté de

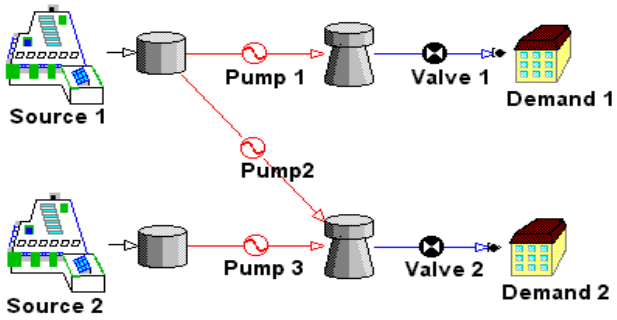

Figure 5 : un réseau de distribution d'eau potable

fonctionnement (construction du diagramme de fiabilité mettant en évidence les chemins de succès reliant sources et points de consommation) il n'en va pas de même pour la disponibilité du flux de demande et a fortiori pour la disponibilité hydraulique pour lesquelles il est nécessaire de prendre en compte l'impact des défaillances sur des paramètres physiques (débit ou pression). L'étude prévisionnelle de ce type d'indicateurs nécessite par conséquent une approche spécifique basée sur la fiabilité des réseaux à flot stochastique.

Pour ce faire, le réseau réel tel celui de la figure 5 est modélisé par un réseau reliant unique nœud origine (représentant les points de production) à un unique nœud extrémité (représentant les points de consommation). A chaque composant du réseau réel (en dehors des simples points de stockage) est alors associé un arc auquel est affecté une capacité en m3/h (voir figure 6). Chaque arc (en dehors des arcs correspondant aux consommations considérées comme parfaites, c'est-à-dire sans défaillance) peut alors subir des défaillances équivalentes à faire passer leur capacité de sa valeur nominale à zéro (les défaillances sont considérées comme soudaines et totales).

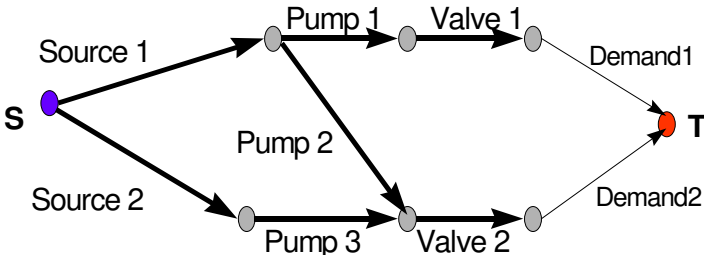

Figure 6 : Le modèle du réseau

La disponibilité du flux de demande est alors la probabilité pour que, compte tenu de toutes les défaillances possibles de leur taux d'occurrence et des taux de réparation associés, le flux à travers tous les arcs correspondant aux consommations soit égal à sa valeur nominale.

L'étude de la disponibilité hydraulique nécessiterait quant à elle une modélisation plus fine prenant en plus en compte la pression associée aux arcs représentant les productions et les consommations, et par conséquent la modélisations des pertes en charge introduites par les éléments passifs (valves et canalisations) ainsi que des différentiels de pression introduits par les éléments actifs (pompes). Une telle modélisation sort du cadre de la présente communication et ne modifie de toutes façons pas de manière notable l'ordre de grandeur des résultats obtenus.

Aspect qualitatif : Coupes minimales

A partir du modèle précédent, on effectue alors la recherche des coupes minimales et des chemins minimaux de succès, utilisant un algorithme basé sur la factorisation de Shannon [10] : la fonction de succès $\phi(X_1,\ldots X_N)$ du réseau (prenant la valeur logique 1 si toutes les demandes sont satisfaites et 0 sinon) qui dépend des N variables booléennes X_i traduisant le bon fonctionnement du composant i, peut en effet être décomposée en utilisant autant de fois que nécessaire la formule [11]:

$$\phi(X_1,\ldots,X_N) = X_i.\phi(X_1,\ldots,X_{i-1},1,X_{i+1},\ldots,X_N) + \overline{X_i}.\phi(X_1,\ldots,X_{i-1},0,X_{i+1},\ldots,X_N)$$

L'application de cette formule revient à construire un arbre binaire dont les nœuds représentent les combinaisons d'état possibles des différents composants et auxquels sont associées les valeurs de la fonction de succès. Pour ce faire, on utilise un simulateur de réseau, partie intégrante de l'application d'évaluation de la disponibilité (SAEP : Sûreté de l'Approvisionnement en Eau Potable), permettant de déterminer dans les différents cas de panne le flux disponible à chaque point de demande, et donc l'état de succès du réseau. L'exploration de l'arbre de Shannon et l'identification des chemins menant de la racine aux nœuds auxquels est associée la valeur logique 0 permet de déterminer les coupes puis d'extraire les coupes minimales et les chemins minimaux de succès.

Afin d'éviter de se heurter à des problèmes d'explosion combinatoire dans la construction de l'arbre de Shannon, des heuristiques sont utilisées dans le choix de l'ordre des variables utilisé. Comme il n'est à l'évidence pas utile de poursuivre la construction au delà du premier nœud de valeur logique 0 rencontré sur chaque branche, on choisit en priorité les variables associées aux composants ayant les capacités les plus importantes (dont on peut penser que la panne va rapidement conduire à la panne du système). Dans certains cas l'arbre peut être tronqué, et sa quantification conduira à une borne supérieure de la disponibilité avec une incertitude qui peut être bornée [12].

Aspect quantitatif

Connaissant la disponibilité asymptotique de chaque composant (fonction de son taux de défaillance et du taux de réparation associé) on obtient alors aisément la disponibilité asymptotique du réseau, ce qui revient à utiliser autant de fois que nécessaire le théorème des probabilités totales (Ai représentant la disponibilité du composant i) :

$$A(X_1,\ldots,X_N) = A_i.A(X_1,\ldots,X_{i-1},1,X_{i+1},\ldots,X_N) + (1-A_i).A(X_1,\ldots,X_{i-1},0,X_{i+1},\ldots,X_N)$$

Notons que l'obtention des coupes minimales se fait en parallèle lors du calcul de la disponibilité.

Le simulateur de réseau intégré dans l'outil SAEP permet également d'accéder au manque $M(C_i)$ généré par chaque coupe minimale C_i (différence entre le flux obtenu et la demande), puis à l'indicateur de *manque moyen* (moyenne pondérée par la probabilité d'occurrence de chaque coupe minimale).

Il est donc ainsi possible lors de la conception de futurs réseaux, de réaliser des études prévisionnelles de la disponibilité du flux de demande et de confronter les résultats aux objectifs fixés, afin de parvenir à un dimensionnement du système au juste nécessaire (évitant d'une part une conception trop faible pour tenir les objectifs, mais également une conception exagérément renforcée en regard de ces mêmes objectifs).

Résultats

La figure 7 représente un réseau d'eau potable au Sud de Paris dont les besoins moyens en consommation sont de l'ordre de 100000 m3/jour.

Le but de l'application de l'outil sûreté AEP sur ce réseau est de comparer deux scénarii d'investissement issus d'une étude de schéma directeur. Le premier scénario consiste en la construction d'une nouvelle usine (Source6 sur la figure); le deuxième est celui du doublement de la capacité de l'usine de traitement désignée par Source5 sur la figure. Les résultats présentés sur la figure 8 montrent l'évolution du manque moyen journalier par scénario en fonction de l'évolution annuelle de la consommation. La courbe correspondant au scénario 1 (construction d'une nouvelle usine) est toujours en dessous (manque moyen moins important) de celle correspondant au scénario 2 (doublement de capacité d'une usine existante).

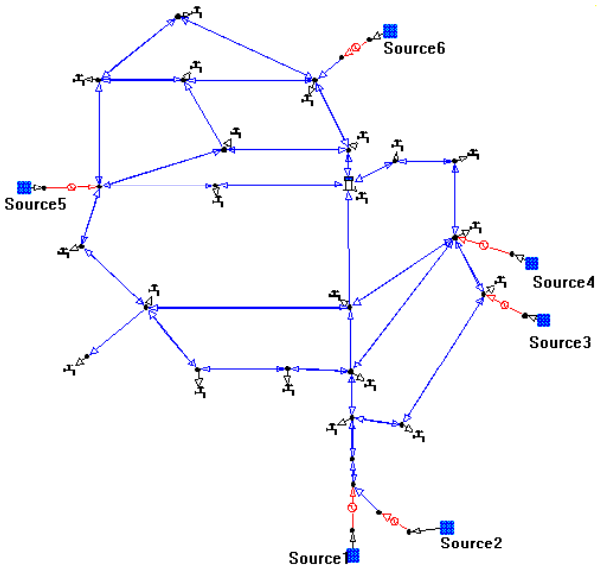

Figure 7 : Le réseau Sud de Paris

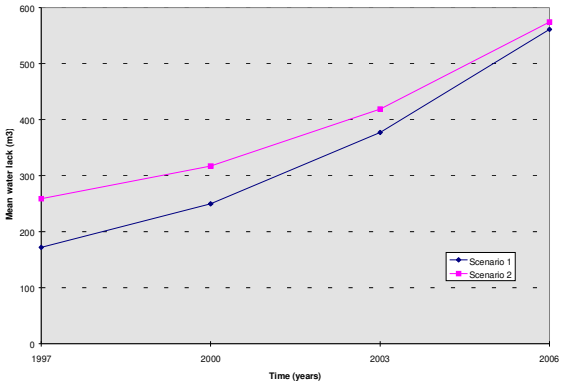

Figure 8 : Comparaison de deux investissements en termes d'impact sur la disponibilité

La comparaison entre les deux courbes montre l'apport en termes de disponibilité du premier scénario par rapport au deuxième. Les deux courbes se rejoignent en 2006, car la consommation atteint le niveau de production des ressources existantes, ce qui nécessite des nouveaux investissements de ressources.

Notons que les données de disponibilité ont été obtenues soit par une analyse statistique basée sur l'historique soit à partir des lois de probabilité génériques tenant compte des caractéristiques des composants (pour les canalisations : âge, longueur et diamètre)

Les résultats précédents montrent l'efficacité de l'outil SAEP pour les réseaux de taille raisonnable. Les indicateurs de sûreté (disponibilité, manque moyen) calculés le sont moyennant une approximation mais l'erreur induite qui peut être majorée est très faible (inférieure à 10^{-11}). De plus les temps de calcul restent très raisonnables de l'ordre de 30 secondes pour un réseau tel que celui présenté.

Une évaluation exacte de ces indicateurs pour un réseau de plus grande taille (plus d'une centaine de composants) nécessite l'utilisation de techniques différentes comme les diagrammes de décision binaire (BDD) [13] qui ont été utilisés dans les réseaux à connexité mécanique [14].

Conclusion

Les travaux objet de la présente communication permettent une maîtrise des risques liés à la production et à la distribution d'eau potable, avec un niveau de confiance très élevé ce qui est indispensable dans un domaine aussi critique. La démarche mise en place prend en compte les spécificités des métiers de l'eau :

- Pour l'aspect sécurité (qualité sanitaire de l'eau distribuée) par une surveillance continue de nombreux paramètres de l'eau de la ressource (rivière ou nappe phréatique), dont la qualité peut évoluer très rapidement en cas d'intempéries exceptionnelles ou de pollution accidentelle. Cette surveillance indispensable permet d'adapter en temps réel le traitement en fonction des caractéristiques de l'eau brute, voire dans les cas extrêmes d'interrompre la distribution. Elle est robuste aux défaillances de capteurs grâce à la mise en place d'un système très performant de détection de données aberrantes et de leur reconstruction. Elle est également une surveillance intelligente dans le sens où elle ne se borne pas à présenter les paramètres mesurés et corrigés, mais est capable de les exploiter pour obtenir des informations synthétiques et directement interprétables.

- Pour l'aspect disponibilité (satisfaction de la demande des abonnés), par la mise en place d'une démarche d'analyse prévisionnelle d'indicateurs spécifiques aux métiers de l'eau (prenant en compte le débit en tout point du réseau), ce qui nécessite une approche basée sur la fiabilité des réseaux à flot stochastique. L'outil actuel (Sûreté de l'Approvisionnement en Eau Potable : SAEP) permet un calcul approché mais avec une erreur résiduelle très faible des indicateurs liés à un réseau de taille raisonnable. Le calcul des très grands réseaux ainsi que l'affinement des modèles de panne utilisés (prise en compte en particulier des conditions d'environnement) font l'objet de travaux en cours.

Mots clés

Eau potable, environnement, santé publique, maîtrise des coûts d'investissement et d'exploitation, indicateurs de mesure des risques.

Références

[1] Trautmann T., Clément M., Denoeux T. and Wittig T. Application of intelligent techniques to river quality monitoring. Proceedings of EUFIT'95, Aachen, August 1995, vol. 3, pages 1603-1609

[2] Petit-Renaud S. and Denoeux T. A neuro-fuzzy system for missing data reconstruction. IEEE Workshop on Emerging Technologies, Intelligent Measurement and Virtual systems for instrumentation and measurement, Saint-Paul USA, May 1998.

[3] Petit-Renaud S. and Denoeux T. A Fuzzy-neuro system for reconstruction of multi-sensor information, Fuzzy-Neuro Systems'98, Ed. W. Brauer Munich March 1998, Infix pages 322—329.

[4] Böhme T. J., Valentin N., Cox C. S. and Denoeux. T. Comparison of autoassociative neural networks and Kohonen maps for signal failure detection and reconstruction. Proceedings of ANNIE'99 Saint-Louis, November 1999.

[5] Valentin N., Fotoohi F. and Denoeux T. Modeling of coagulant dosing in a water treatment plant. EANN'99 Varsovie, September 1999.

[6] Valentin N., Denoeux T.and Fotoohi F. A hybrid neural network based system for optimization of coagulant dosing in a water treatment plant. Proceedings of IJCNN'99 Washington, July 1999.

[7] Kansal. M.L, Kumar A., Sharma. P.B. Reliability analysis of water distribution systems under uncertainty, *Reliability Engineering.* 1995, Vol. 50, pp.51-59.

[8] Aggarwal K.K., Chopra. Y.C., Bajwa. J.S. Capacity consideration in reliability analysis of communication systems, *IEEE Trans. Reliab*. 1982, Vol. R-31, N°2, Jun., pp. 177-181.

[9] Wu. S-J., Yoon J-H., Quimpo R.G. Capacity-weighted water distribution system reliability, *Reliability Engineering.* 1993, Vol. 42, pp. 39-46.

[10] Theologou O., Contribution à l'évaluation de la fiabilité des réseaux, Thèse de doctorat, Université de Technologie de Compiègne 1990.

[11] Shneeweiss W.G. Disjoint Boolean product via Shannon's expansion, *IEEE Trans. Reliab*. Oct. 1984, Vol. R-33 N°4, pp. 239-332.

[12] Shneeweiss W.G. Approximation of fault-tree analysis without cut sets. *In : Proceedings of Annual Reliability and Maintainability Symposium*, 1992, pp. 370-375.

[13] Odeh K. Nouveaux algorithmes pour le traitement probabiliste et logique des arbres de défaillance. Thèse de doctorat, Université de Technologie de Compiègne, 1995

[14] Dutuit Y., Rauzy A., Signoret J-P, Computing network reliability with Réséda and Aralia