

Méthodologie de validation des données hydrométriques en temps réel dans un réseau d'assainissement urbain

Real time validation of hydrometric data in a sewer network

S. BENNIS^{1*}, F. BERRADA^{2*}, F. BERNARD^{3*}

Reçu le 13 mars 2000, accepté le 22 décembre 2000**.

SUMMARY

We developed an automated methodology for real-time validation of hydrometric data in a sewer network. Our methodology uses real-time validated data to optimise system management and non-real-time data to evaluate day-to-day performance.

Two approaches can be used to validate and correct hydrometric data; the choice depends on the number of level gauges present in a system. In single gauge systems, univariate filtering is used to smooth data. For example, frequency filtering systematically eliminates values corresponding to frequencies higher than a predetermined threshold frequency. In systems with several gauging stations — duplex, triplex, or multiplex systems — the multivariate filtering method proposed here can be used to validate data series from each gauge. Material redundancy in duplex or higher order systems makes it possible to detect a deficient gauge, using a decision rule to set aside erroneous readings before averaging accepted values. Part of the underlying principle of this methodology is heavier reliance on gauges that give readings consistent with previous and subsequent validated values in a given series. Thus isolated positive or negative variations within a series are eliminated if corresponding variation values at other gauges are more consistent. To evaluate persistence, a reading is compared to a value predicted by an autoregressive (AR) model calibrated by the previous validated reading.

This filtering technique constitutes an intelligent alternative to the frequency filtering method mentioned above. In more practical terms, it compares the deviation of an AR model prediction from a measured value with the deviation of the same AR model prediction from a value estimated by a regressive

1. Département génie de la construction, École de technologie supérieure, 1100, Notre-Dame Ouest, Montréal, Québec.

2. Département de physique, Faculté des sciences Ain Chock, Université Hassan II, Casablanca, Maroc

3. Responsable du service technique, Ville de Verdun, Québec.

* Correspondance. E-mail : sbennis@ctn.etsmtl.ca

E-mail : faouzi_berrada@hotmail.com

E-mail : france.bernard@ville.verdun.qc.ca

** Les commentaires seront reçus jusqu'au 14 septembre 2001.

model at other stations in the network. Among the values measured and estimated by the regressive model, the one nearest the AR model prediction is retained.

Our methodology also relies on analytical redundancy generated by direct measurement of flow and hydrological simulation. More precisely, the deviation of the AR model prediction from the measured value is compared with the deviation of the same AR model prediction from a value obtained from a hydrological simulation model. Among measured and simulated values, the one nearest the AR model prediction is retained. To allow consideration of nonstationary models and to avoid the well-known bias of the least squares method, the KALMAN filter is used to identify the parameters of the AR model.

The methodology we propose employs three models. The first generates analytical redundancy using hydrological modelling. An autoregressive model is then used to predict future runoff rate values. Finally, a voting process model is used to compare measured and simulated values.

The proposed methodology was tested on the Verdun sewer system in Quebec with successful results. Two types of artificial disturbance of the measured hydrograph were created: white noise was added to measured values and disturbances of large amplitude and various forms were introduced. The methodology produced the initial values and performance criteria were conclusive. Thus on-site testing confirms that this approach allows completely automated detection and correction of most anomalies. Flood peaks were neither underestimated nor overestimated, and total runoff volumes were retained.

Key-words: *validation, redundancy, flow, measurement, KALMAN filter, autoregressive, real time, sewer.*

RÉSUMÉ

L'objectif du présent travail est l'élaboration d'une méthodologie de validation des données hydrométriques mesurées dans un réseau d'assainissement. L'information validée est utilisée aussi bien en temps réel, pour optimiser les consignes de gestion, qu'en temps différé, pour poser le véritable diagnostic et évaluer, sur une base quotidienne, l'efficacité des systèmes d'assainissement.

Le principe de base de la méthodologie proposée repose sur la redondance analytique de l'information provenant d'une part de la mesure directe du débit sur le terrain et d'autre part du débit simulé à partir des variables météorologiques. On compare ainsi d'une part, l'écart entre la valeur prévue par un modèle autorégressif (AR) et la valeur mesurée et d'autre part, l'écart entre la valeur prévue par ce même modèle AR et la valeur simulée par un modèle hydrologique. Parmi les valeurs, mesurée et simulée, celle qui se rapproche le plus de la valeur prévue est retenue. Afin de considérer des modèles non stationnaires et d'éviter le biais d'estimation des paramètres de régression par la méthode standard des moindres carrés, le filtre de KALMAN est utilisé pour identifier les paramètres du modèle AR.

La méthodologie proposée a été testée avec succès sur un bassin urbain de la municipalité de Verdun. L'hydrogramme mesuré a été bruité artificiellement à la fois par un bruit blanc et par un certain nombre de perturbations de grandes amplitudes et de différentes formes. Le processus de validation a permis de retrouver pratiquement les mesures initiales, non bruitées. Les critères de performance introduits sont largement concluants.

Mots-clés : *validation, redondance, débit, mesure, filtre de KALMAN, autorégressif, temps réel, assainissement.*

1 – INTRODUCTION ET PROBLÉMATIQUE

La gestion dynamique des réseaux d'assainissement s'est récemment développée grâce à l'essor des technologies, notamment en terme de transfert de l'information. C'est ainsi que la connaissance, en temps réel, de la qualité et la quantité des flux, à différents endroits du réseau, permet d'utiliser, de façon optimale, la capacité de transport et de stockage et minimiser, par exemple, l'impact des déversements et des inondations (PLEAU *et al.*, 2000).

Malheureusement, les données recueillies sont toujours incertaines et parfois fausses ou manquantes. Dès lors, il est difficile de faire un bon diagnostic ou une bonne détection d'anomalie, surtout lorsque survient un grave problème d'inondation.

Les sources d'erreurs lors de mesures des écoulements dans les réseaux sont très nombreuses. On peut citer, à titre d'exemple, les erreurs qui :

- produisent un biais dans la mesure : encrassement et colmatage de la tête d'un capteur ou d'une crépine, dérive ou simplement, mauvais étalonnage, etc. ;
- génèrent des données manquantes : panne totale ou problème de transmission ;
- sont dues à la méconnaissance du phénomène qu'on mesure. On parle dans ce cas d'erreurs de contexte : curage, arrosage, déviation à l'amont, rejet accidentel ou accumulation de dépôt créant ainsi un barrage.

Une façon efficace de s'assurer de la validité des mesures est de multiplier (système multiplex) les chaînes de mesure. Cette redondance, dite matérielle, permet de détecter, voire de localiser le capteur défaillant (BENNIS et KANG, 2000 ; BERRADA *et al.*, 1996 ; PERRAULT *et al.*, 1995 ; NGUYEN et BISSON, 1998 ; BLANCHET *et al.*, 1998).

Une étude statistique des différents écarts entre les mesures, par utilisation de la méthode des composantes principales, pourrait, moyennant un seuil de tolérance, localiser le capteur défaillant (MCCUEN et SNYDER, 1986). L'inconvénient de cette approche réside dans la contrainte d'imposer un seuil arbitraire de tolérance pour détecter les mesures notoirement aberrantes. En réalité, les aberrations ne sont pas toujours évidentes et de grandes amplitudes.

Une autre méthode envisageable est la méthode des niveaux virtuels (ROY *et al.*, 1992). Quoique intéressante, cette méthode sous-estime légèrement les pointes des apports naturels calculés par bilan hydrique en fonction de ces niveaux fictifs (BERRADA, 1998). Par ailleurs, elle ne constitue pas une méthode de validation proprement dite car, la mesure virtuelle n'est représentative d'aucun capteur et s'écarte largement des mesures effectuées.

Malheureusement, compte tenu des contraintes budgétaires, la métrologie des réseaux d'assainissement urbain reste encore très limitée (ENEL, 1987). Ainsi, la redondance matérielle sur laquelle reposent les méthodes précitées est quasiment inexistante et dans le meilleur des cas, une seule station de mesure existe à un endroit donné du réseau.

Dans ces conditions, l'objet de ce papier est de développer une méthodologie complète de validation des données mesurées en temps réel basée sur le

principe de la redondance analytique. Celle-ci consiste à estimer la valeur mesurée Z_t qu'on cherche à valider à partir d'autres valeurs mesurées et qui lui sont reliées par des modèles physiques, statistiques ou conceptuels. La méthodologie présentée dans cet article s'inscrit dans cette démarche en restant assez simple dans son emploi. Elle utilise deux modèles :

- un modèle hydrologique (paragraphe 3.2) qui permet de transformer les données météorologiques en débits aux différents nœuds de drainage. Une redondance analytique, sera alors générée pour se substituer à la redondance matérielle qui fait généralement défaut ;
- un modèle temporel, ARMA (BOX et JENKINS, 1976) autorégressif de type AR(2). Il fournit les tendances de variation et permet de détecter les décrochages dans les signaux mesurés ou simulés. Pour remédier au problème d'autocorrélation des résidus (paragraphe 3.1), on adapte à ce modèle la technique du filtre de KALMAN.

Cette méthodologie (paragraphe 2) a l'avantage de ne pas transformer systématiquement le signal de base, seules les mesures suspectes sont corrigées.

Le site d'étude, les performances de la méthodologie, ainsi que les résultats obtenus sont présentés au paragraphe 4.

2 – MÉTHODOLOGIE DE VALIDATION

Lorsqu'on veut valider, en temps réel, une donnée mesurée, il est nécessaire de disposer, au même instant, de plusieurs valeurs de la même mesure. Un processus de vote est alors utilisé pour décider, à chaque pas de temps, laquelle parmi ces valeurs est la plus représentative du phénomène observé. Une bonne connaissance de l'évolution de ce phénomène ainsi qu'une analyse des résidus entre mesures que l'on cherche à valider et les autres mesures directes ou résultant d'une redondance analytique sont, à cet effet, essentielles pour poser le véritable diagnostic et détecter les anomalies.

Afin d'illustrer la méthodologie proposée, on considère un cas assez simple où la mesure de débit à l'instant présent « t » qu'on cherche à valider est estimée par un modèle de simulation hydrologique. Il devient évident dans ce cas que la validité du diagnostic est tributaire de la performance du modèle de simulation hydraulique et hydrologique. Par ailleurs, un bassin de drainage, urbain ou rural, joue naturellement le rôle d'un filtre mathématique dont le signal d'entrée est la pluie et la sortie est le débit. Ainsi, les valeurs successives du débit à des instants rapprochés sont fortement autocorrélées. Comme le modèle autorégressif est basé principalement sur la fonction d'autocorrélation, il est utilisé comme indicateur de la tendance de variation à court terme pour détecter les décrochages éventuels dans l'un ou l'autre des signaux, mesuré ou simulé.

Il faut être conscient que les perturbations dans les signaux mesurés ne proviennent pas seulement de la défaillance des appareils de mesure ou des erreurs de contexte énumérées précédemment. Des instabilités peuvent prove-

nir lors du passage d'un écoulement à surface libre à un écoulement en charge. En effet, la forme de l'hydrogramme à l'aval d'une zone en charge peut être extrêmement perturbée. Le débit manifeste alors une grande variabilité avec une alternance de périodes de blocage où le débit est très faible et de périodes de vidange du réseau (CHOCAT, 1997). Bien que les modèles actuels de simulation soient capables de reproduire ce phénomène, il est préférable de choisir le point de mesure de façon à minimiser l'impact de ces effets sur le signal mesuré. Pour le débit simulé, ces perturbations peuvent aussi être dues à des instabilités numériques dans le processus de résolution des équations aux dérivées partielles de Saint-Venant. Un choix approprié de la méthode de résolution et des pas de discrétisation spatial et temporel peut éliminer ces perturbations.

La méthode de validation proposée consiste à comparer, d'une part, l'écart entre la valeur prévue par un modèle autorégressif (AR) et la valeur mesurée et, d'autre part, l'écart entre la valeur prévue par le même modèle AR et la valeur simulée par un modèle hydrologique. Parmi les valeurs mesurée et simulée, celle qui se rapproche le plus de la valeur prévue est retenue. Le principe de cette approche est illustré à la *figure 1*. Mathématiquement, les résidus e_{1t} et e_{2t} sont donnés respectivement par les valeurs absolues des écarts suivants :

$$e_{1t} = |Q_t^{Prévu} - Q_t^{Mesuré}| \tag{1}$$

$$e_{2t} = |Q_t^{Prévu} - Q_t^{Simulé}| \tag{2}$$

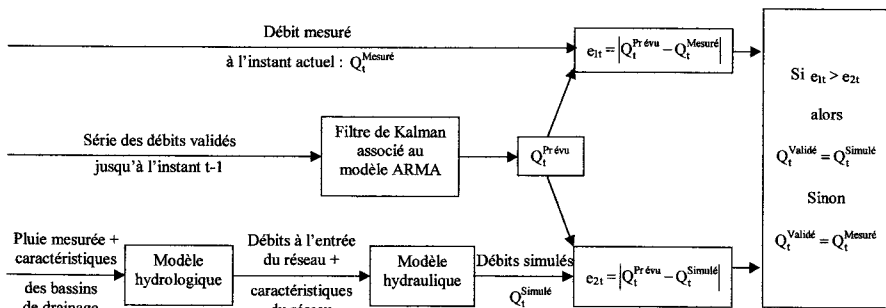


Figure 1 Processus de validation du débit mesuré à l'instant « t ».
Validation process of measured flow at time « t ».

Afin de considérer des modèles non stationnaires et d'éviter le biais d'estimation des paramètres de régression par la méthode standard des moindres carrés, le filtre de KALMAN est utilisé pour identifier les paramètres du modèle AR utilisé (BENNIS et BRUNEAU, 1993a ; 1993b). L'utilisation de ce filtre permet, en particulier, d'obtenir des résidus indépendants dont la fonction d'autocorrélation s'approche de zéro (BENNIS, 1987 ; KALMAN, 1960 ; KALMAN et BUCY, 1961). Ainsi, le modèle autorégressif utilisé pour la prévision est réajusté à chaque pas de temps à l'aide du débit mesuré et validé.

3 – FORMULATION DES DIFFÉRENTS MODÈLES UTILISÉS

3.1 Le modèle de prévision ARMA

Un modèle ARMA de type AR(2) est utilisé pour modéliser la tendance de variation du débit à un instant donné « t ». Ce modèle fournit les tendances de variation, permet de détecter les décrochages dans les signaux, mesuré ou simulé, et s'avère très supérieur à une méthode de lissage exponentiel des données, utilisée par NGUYEN et BISSON (1998). Lorsque utilisé pour la prévision, l'ordre optimal du modèle AR peut être choisi en utilisant le critère d'AKAIKE (1974). Comme ce modèle est utilisé ici uniquement comme indicateur de la tendance de variation du débit et que nous sommes intéressés par une méthode automatisée de validation, l'ordre a été fixé arbitrairement à deux. La présentation de ce modèle a été adaptée à la formulation générale de l'algorithme de KALMAN. Ce dernier permet, à chaque pas de temps, de filtrer les mesures et d'identifier les paramètres.

On utilise ainsi la technique MISP (Mutually Interactive State-Parameters estimation) (TODINI, 1978). Elle consiste à estimer, à partir d'un premier filtre de KALMAN, les mesures de la série chronologique. On utilise pour cela, des coefficients autorégressifs eux-mêmes estimés par un second filtre de KALMAN. À son tour, cette dernière estimation se basera sur les mesures estimées par le premier filtre. Cela revient à accorder une meilleure confiance aux valeurs estimées avec un maximum de vraisemblance plutôt qu'aux valeurs mesurées.

Le premier filtre se formule comme suit :

$$\begin{cases} \text{Équation d'état : } X_t = \Phi_{t/t-1} X_{t-1} + W_{t-1} \\ \text{Équation de mesure : } Z_t = H_t X_t + V_t \end{cases} \quad (3)$$

$$\text{où : } X_t = \begin{bmatrix} Q_t^{\text{Prévu}} \\ Q_t^{\text{Valid}} \end{bmatrix} ; X_{t-1} = \begin{bmatrix} Q_{t-1}^{\text{Valid}} \\ Q_{t-2}^{\text{Valid}} \end{bmatrix} ; \Phi_{t/t-1} = \begin{bmatrix} \theta_1 & \theta_2 \\ 1 & 0 \end{bmatrix}$$

$\Phi_{t/t-1}$ est la matrice de transition contenant les paramètres du modèle autorégressif ;

$$\text{et : } W_t \text{ est le bruit du modèle : } W_t = \begin{bmatrix} W_t \\ 0 \end{bmatrix}$$

$Q_t^{\text{Prévu}}$ représente le débit prévu pour l'instant « t », $H_t = [1 \ 0]$ est la matrice d'échelle ; V_t est le bruit de mesure et Z_t représente le débit à l'exutoire du bassin de drainage ou dans une section de collecteur.

Le second filtre se formule comme suit :

$$\begin{cases} \text{Équation d'état : } \Theta_t = \Theta_{t-1} + W_{t-1}^* \\ \text{Équation de mesure : } Z_t = H_t^* \Theta_t + V_t^* \end{cases} \quad (4)$$

$$\text{où : } \Theta_t^T = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, V_t^* \text{ et } W_t^* \text{ sont respectivement, les bruits de mesure et du}$$

modèle associés au second filtre de KALMAN. $H_t^* = \hat{X}_{t-1/t-1}^T$ est par ailleurs, sa nouvelle « matrice » de transition.

L'estimation *a priori* ou prévue, $\hat{X}_{t/t-1}$ connaissant seulement les données précédentes l'instant actuel « t » et l'estimation *a posteriori* $\hat{X}_{t/t}$ sont déterminées à partir de l'algorithme général de KALMAN.

3.2 Description du modèle physique de génération de la redondance

Si la valeur mesurée Z_t représente le débit ou le niveau d'eau à l'exutoire d'un bassin de drainage ou dans une section d'un collecteur, il est possible d'estimer cette même grandeur à partir de modèles hydrologiques et hydrauliques. La première famille de modèles permet de transformer les hydrogrammes prévus ou mesurés en hydrogrammes à l'exutoire des bassins de drainage. La seconde famille de modèles hydrauliques permet de propager dans le réseau les hydrogrammes calculés précédemment pour calculer les débits et niveaux aux différents nœuds de ce réseau.

Les modèles hydrologiques sont généralement basés sur la notion de fonction de transfert utilisant des hydrogrammes unitaires, réels ou synthétiques. Les modèles hydrologiques peuvent aussi être conceptuels à l'instar du réservoir non linéaire incorporé dans XP_SWMM (HUBERT et DICKINSON, 1988) qui a été utilisé dans ce projet. Dans ce modèle, le ruissellement est idéalisé par un réservoir non linéaire représenté à la figure 2.

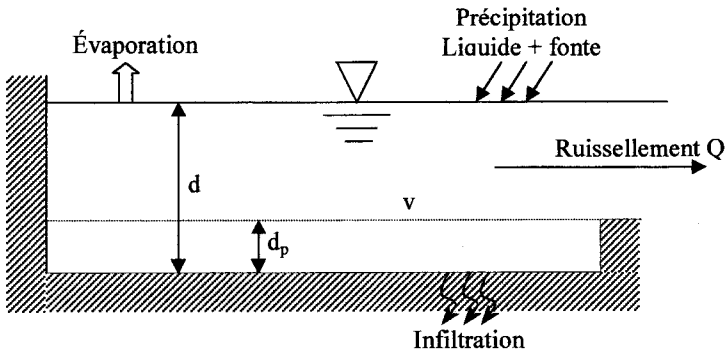


Figure 2 Schématisation du modèle du réservoir non linéaire.
Non linear reservoir model.

Pour étudier l'évolution d'une onde de crue on utilise généralement les équations de continuité et de la dynamique de Barré de SAINT-VENANT. Dans le logiciel XP_SWMM qui a été utilisé dans ce projet, ces deux équations sont résolues par des différences finies explicites ou implicites (HUBER et DICKINSON, 1988).

4 – SITE DE L'ÉTUDE ET RÉSULTATS

Le réseau d'égout de la ville de Verdun, située au Québec, se divise en 5 sous-bassins hydrographiques desservis par un réseau unitaire.

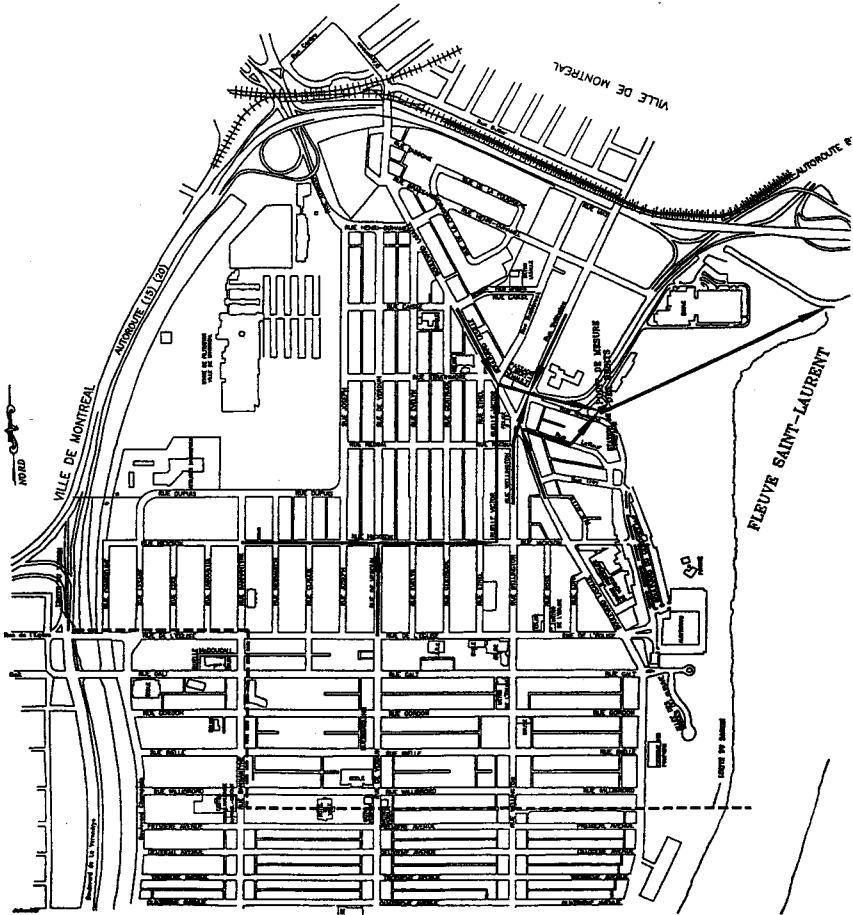


Figure 3 Schéma du secteur I de la Ville de Verdun (Québec).
Schematic of the Verdun bassin (Quebec).

Le site d'étude est le secteur I représenté à la *figure 3*. Le bassin est limité à l'ouest par le canal de l'aqueduc et l'usine de filtration de la ville de Montréal. Au Nord, le secteur est limité par le territoire de la ville de Montréal. Au sud, il est limité par la première avenue et finalement à l'est par le fleuve Saint-Laurent.

Le réseau d'égout unitaire est constitué en grande partie de conduites de forme ovoïdale (600×900) et installées selon de faibles pentes. Dans ces conditions, il est difficile d'acheminer un débit de récurrence 1/5 ans sans débordement et le réseau unitaire du secteur I de la ville de Verdun a causé plusieurs inondations. En vue de le restructurer, la municipalité a entrepris la construction d'un nouveau réseau souterrain pour drainer le surplus d'eau pluviale et l'acheminer directement au milieu récepteur soulageant ainsi le réseau existant et évitant les débordements fréquents d'eau unitaire dans le fleuve Saint-Laurent.

Le débit a été mesuré à l'entrée de la station de pompage Rhéaume (*figure 3*). Ce point a été choisi car actuellement le débit y représente la résul-

tante de toutes les eaux pluviales, sanitaires, industrielles et d'infiltration du secteur I. Les lames de pluie sont enregistrées, en continu, à des intervalles de temps de 5 minutes grâce à deux pluviographes à augets basculeurs. Dans notre application, le pluviographe utilisé est situé sur le toit de cette même station de pompage, c'est-à-dire au même endroit où s'effectuent les mesures des débits (figure 3). Ces derniers sont aussi mesurés, en continu, à des intervalles de temps de 5 minutes grâce à un appareil de type hauteur vitesse avec sonde, installé à l'entrée de la station Rhéaume depuis le 21 septembre 1999. Cet appareil permet de mesurer le niveau et la vitesse pour obtenir le débit pour la géométrie concernée. 4 événements météorologiques ont été documentés. Pour des raisons pratiques de représentation, un seul événement a été retenu pour illustrer la méthodologie de validation proposée.

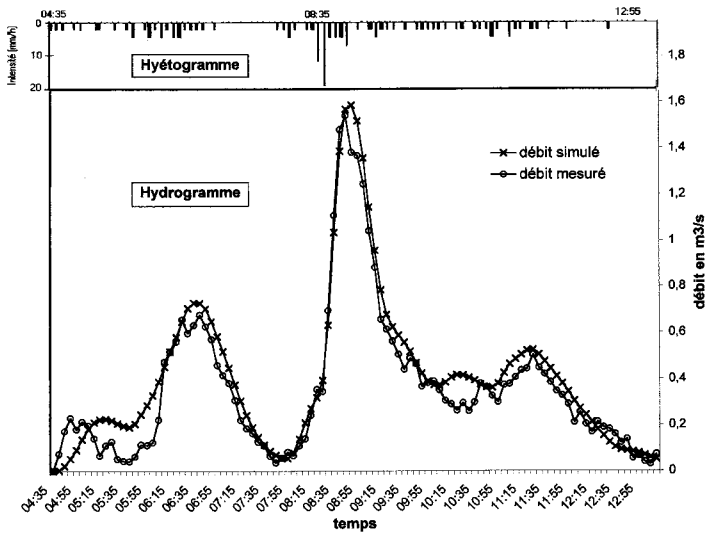


Figure 4 Comparaison entre débit simulé et débit mesuré.
Comparison between measured and simulated flows.

La figure 4 montre la comparaison entre les débits simulés à l'aide du logiciel XP_SWMM et les débits mesurés à l'entrée de la station de pompage Rhéaume pour l'événement du 30 septembre 1999. Comme les paramètres utilisés dans le logiciel XP_SWMM ont été prédéterminés à l'aide des caractéristiques morphologiques et physiographiques de chaque sous-bassin, la calibration des modèles hydraulique et hydrologique a été grandement facilitée.

Les paramètres qui permettent d'apprécier la concordance entre les débits, mesuré et simulé, sont :

- Le coefficient de NASH :
$$NASH = 1 - \frac{\sum_{i=1}^N (Q^{Mesuré}(i) - Q^{Simulé}(i))^2}{\sum_{i=1}^N (Q^{Mesuré}(i) - Q^{Moyen}(i))^2} \quad (10)$$

Le rapprochement de ce coefficient de l'unité est une indication de la bonne performance du modèle de simulation.

- Le rapport des volumes : $RV = \frac{\text{Volume total de ruissellement simulé}}{\text{Volume de ruissellement mesuré}}$

Le rapprochement de ce rapport de l'unité est une indication de la bonne performance du modèle de simulation.

- Le rapport entre les débits de pointe simulé et mesuré : $RD = \frac{Q_{Max}^{Simulé}}{Q_{Max}^{Mesuré}}$ (11)

Le rapprochement de ce rapport de l'unité est une indication de la bonne performance du modèle.

- Synchronisme des débits de pointe Δt = écart (\pm) entre l'occurrence du débit de pointe mesuré et calculé.

$\Delta t = 0$ indique un synchronisme parfait.

Le *tableau 1* fournit ces quatre paramètres pour tous les événements étudiés. Le rapport moyen des volumes est 1,025. Le rapport moyen des débits de pointe est 1,005. Le coefficient de NASH moyen est 0,727. Le décalage moyen entre l'occurrence du débit de pointe simulé et mesuré est de 5 minutes. Dans ce contexte, on peut affirmer qu'on a réussi à créer une redondance d'information provenant d'une part de la mesure sur le terrain et d'autre part du modèle de simulation hydrologique. La *figure 5* montre la comparaison entre les débits mesurés et les débits validés par la méthodologie proposée. On constate que peu de valeurs mesurées ont été modifiées après le processus de validation et que les modifications apportées sont relativement légères. Ce résultat qui confirme la validité des valeurs mesurées et simulées était prévisible compte tenu de la concordance quasi parfaite entre les deux sources d'information.

Tableau 1 Résultat récapitulatif de performance de la simulation des débits.

Table 1 Performance criteria for flow simulation model.

Événements	Rapport des volumes	Rapport des débits de pointe	Coefficient de NASH	Décalage dans le temps entre les débits de pointe, mesuré et simulé (min.) simulé (min.)
24-09-1999	0,99	1,06	0,60	0
30-09-1999	1,12	1,03	0,93	5
06-10-1999	0,99	0,99	0,58	10
13-10-1999	1,00	0,94	0,80	5

Afin de tester véritablement la méthodologie proposée, nous avons rajouté volontairement un bruit, de distribution normale, de moyenne nulle et d'écart type constant, égal à 10 % de la moyenne des débits mesurés. Dix perturbations de grandes amplitudes ont ensuite été introduites. Ces dernières, de forme géométrique quelconque : en Z, en U et en W (*figure 6*), sont réparties sur la période totale pour simuler plusieurs situations : une, deux, trois... erreurs isolées, placées volontairement, en périodes, de montée et de récession de l'hydrogramme. Pour tester la capacité de la validation proposée et son comportement vis-à-vis des débits de pointe, ces derniers ont fait aussi l'objet de grandes perturbations.

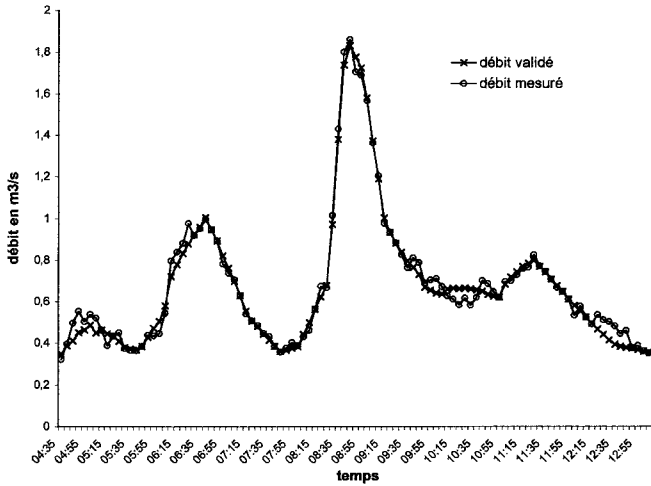


Figure 5 Validation des données mesurées le 30 septembre 1999.
Validation of measured data, September 30, 1999.

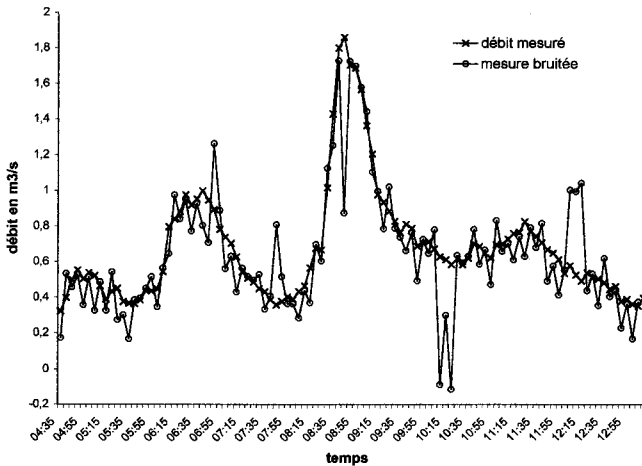


Figure 6 Hydrogrammes mesuré et bruité.
Measured and disturbed hydrographs.

Pour évaluer l'efficacité de cette méthodologie, on propose l'indice suivant de performance :

$$\eta = 1 - \frac{\varepsilon_Q^2 \text{Validée}}{\varepsilon_Q^2 \text{Bruitée}} \quad (12)$$

où : $\varepsilon_Q^2 \text{Validée}$ est la moyenne des écarts quadratiques entre les données validées et les données initiales bruitées.

et : $\varepsilon_Q^2 \text{Bruitée}$ est la moyenne des écarts quadratiques entre les données bruitées et les données initiales mesurées.

Mathématiquement, ces deux quantités s'expriment comme suit :

$$\varepsilon_{Q^{Validé}}^2 = \frac{1}{N} \sum_{j=1}^N [Q^{Validé}(j) - Q^{Mesuré}(j)]^2 \tag{13}$$

$$\varepsilon_{Q^{Bruité}}^2 = \frac{1}{N} \sum_{j=1}^N [Q^{Validé}(j) - Q^{Mesuré}(j)]^2 \tag{14}$$

Par conséquent, si η est négatif ou nul, la technique de validation est à rejeter. Sinon cette technique est efficace. Elle devient très performante si cet indice avoisine l'unité (1).

La figure 7 montre la comparaison entre les données mesurées modifiées et les données validées. Il est clair sur cette figure que toutes les perturbations introduites ont été éliminées après l'exercice de validation. L'hydrogramme des données bruitées et validées a été superposé aux valeurs initiales non bruitées afin de mettre en évidence la performance de la méthodologie (figure 8). Le processus de validation a permis, après l'introduction volontaire des perturbations, de retrouver pratiquement les mesures initiales. Le coefficient η , défini plus haut, vaut 0,898. Ce coefficient ne reflète pas réellement la performance de la méthodologie de validation puisque les mesures initiales non bruitées ont été légèrement modifiées lors du premier exercice de validation (figure 5). On a dû alors recalculer le coefficient η en utilisant la valeur suivante de $\varepsilon_{Q^{Validée}}^2$:

$$\varepsilon_{Q^{Validée}}^2 = \frac{1}{N} \sum_{j=1}^N [Q^{Validé1}(j) - Q^{Validé2}(j)]^2 \tag{15}$$

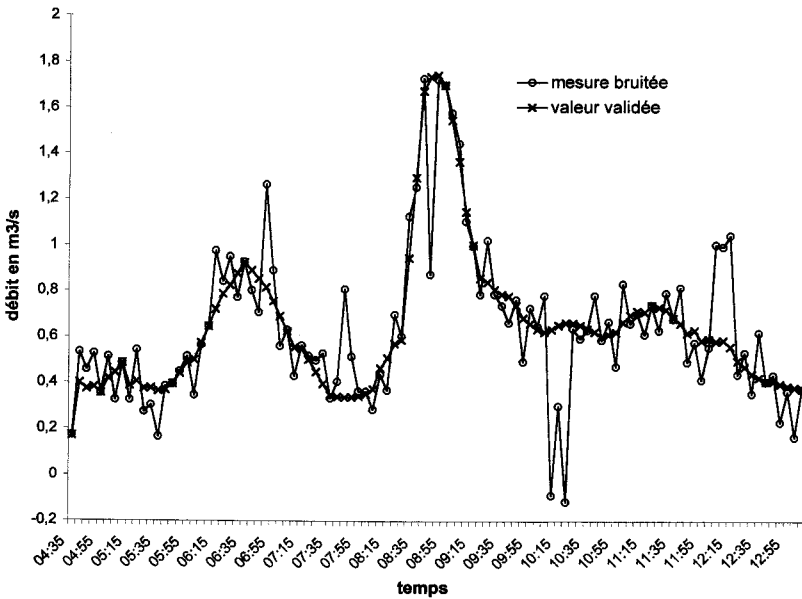


Figure 7 Validation des données bruitées.
Disturbed data validation.

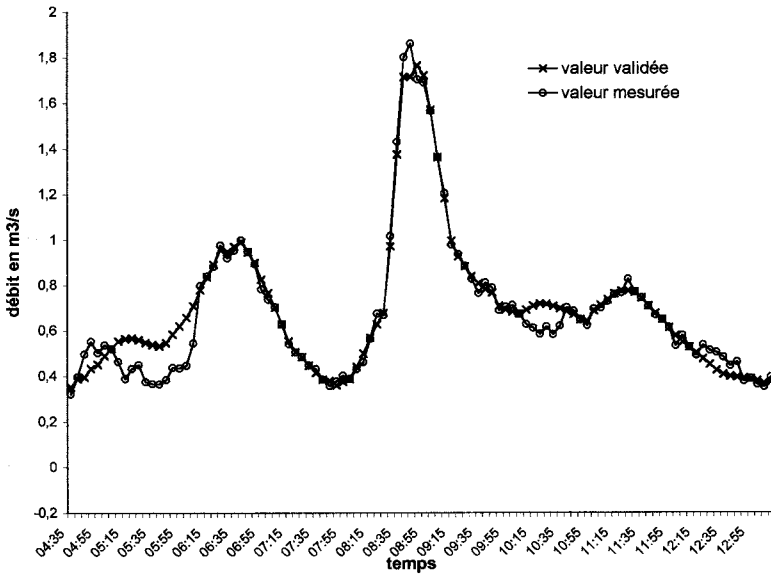


Figure 8 Comparaison entre les données, mesurées initiales et modifiées validées.
Comparison between initial measured data and validated disturbed data.

Les valeurs de référence utilisées pour qualifier la méthodologie de validation sont les débits mesurés et validés avant l'introduction volontaire du bruit. Le coefficient η calculé de cette manière vaut 0,97, valeur qui avoisine l'unité et qui est supérieure à 0,898 obtenue précédemment. Sur la *figure 8*, on constate qu'en début de la première crue les valeurs validées correspondent aux données simulées et s'écartent des valeurs mesurées sur une assez longue période. Ce phénomène qui a l'apparence d'un biais dans la chaîne de mesure peut s'expliquer par le processus de validation adopté. En effet, les paramètres du modèle autorégressif, qui est utilisé comme indicateur de la tendance de variation des hydrogrammes, sont réajustés à chaque pas de temps « t » à l'aide de la valeur validée. Lorsque à un instant « t » cette valeur validée correspond à la donnée simulée, la valeur prévue par le modèle autorégressif pour l'instant « t » s'approche davantage de la valeur simulée que de la valeur mesurée. En l'absence de perturbations dans les valeurs simulées, la valeur validée pour l'instant « $t + 1$ » correspondra encore à la valeur simulée. Ceci ne constitue pas réellement un inconvénient de la méthodologie proposée car, dans beaucoup d'applications, il est préférable d'utiliser les valeurs simulées lissées lorsqu'elles sont voisines des données mesurées. Au même titre, si des instabilités numériques avaient perturbé l'hydrogramme simulé, le processus de validation aurait privilégié systématiquement les données mesurées.

Les résultats de la méthodologie ont été comparés à ceux d'un filtre à régression linéaire, sur fenêtre mobile (*figure 9*), fréquemment utilisé compte tenu de sa simplicité. Pour une fenêtre de largeur m , choisie a priori, ce filtre (GAGNON, 1994) s'écrit comme suit :

$$Q_t^{\text{Validé}} = \frac{1}{2m+1} (Q_{t-m} + Q_t + Q_{t+1} + \dots + Q_{t+m}) \quad (16)$$

Le coefficient η obtenu pour ce filtre ($m = 5$) est de 0,602. Il reste inférieur à 0,97 obtenu pour la méthodologie proposée. Cette performance est d'autant plus remarquable que le filtre à régression linéaire utilisé comme référence fait intervenir à l'instant t , les valeurs futures mesurées $Q_{\text{mesuré } t+1}$ et $Q_{\text{mesuré } t+2}$ qui ne sont pas disponibles en temps réel.

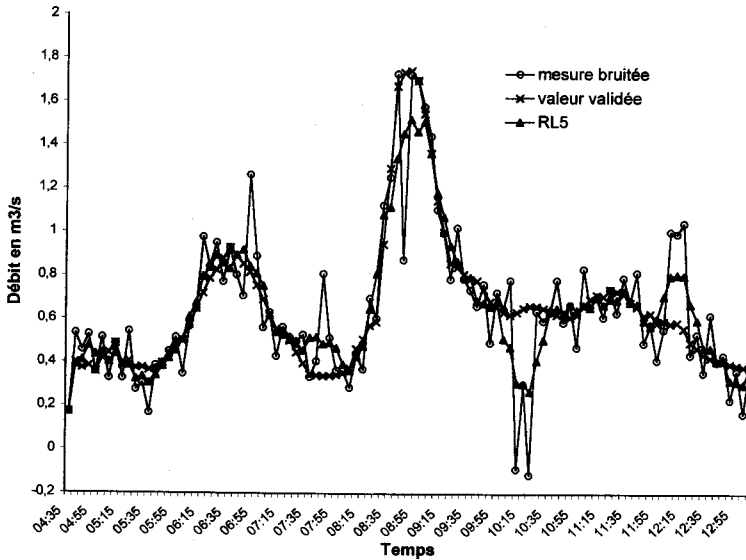


Figure 9 Comparaison entre la validation et le filtrage linéaire à fenêtre mobile.
Comparison between validated data and linear filtering with moving window.

CONCLUSION

L'objectif du travail présenté ici a été de développer une méthodologie qui permet la validation automatisée, en temps réel, des données hydrométriques dans un réseau d'assainissement. Pour cela, il fallait créer une redondance d'information et établir une procédure de diagnostic permettant la détection d'erreurs. La redondance d'information a été créée à l'aide de deux modèles : le premier conceptuel, hydraulique/hydrologique, préalablement calibrés et l'autre, stochastique autorégressif, d'ordre deux. Pour ce dernier modèle, les paramètres autorégressifs sont déterminés par utilisation de la technique adaptative du filtre de KALMAN.

La méthodologie proposée a été testée avec succès sur un bassin urbain de la municipalité de Verdun au Québec. L'hydrogramme mesuré a été bruité

artificiellement à la fois par un bruit blanc et par un certain nombre de perturbations de grandes amplitudes et de différentes formes. Le processus de validation a permis de retrouver pratiquement les mesures initiales non bruitées. Les critères de performance introduits sont largement concluants.

REMERCIEMENTS

Les auteurs tiennent à remercier le Conseil de recherche en sciences naturelles et en génie du Canada, ainsi que le Centre d'Expertise et de Recherche en Infrastructures urbaines, Hydro-Québec, la CUM, et ville de Verdun pour l'appui financier fourni à ce projet. Ils remercient également les réviseurs, qui ont contribué, avec leurs commentaires, à améliorer le contenu de l'article.

RÉFÉRENCES BIBLIOGRAPHIQUES

- AKAIKE H., 1974. A new look at the statistical model identification. *IEEE. Transactions on Automatic Control*, Ac. 19, 716-723.
- BENNIS S., 1987. Préviation des crues par l'analyse des séries chronologiques. Th. Ph. D. Université de Sherbrooke. Canada, 5, 55.
- BENNIS S., KANG N., 2000. Multivariate Technique for Validating Historical Hydrometric Data with Redundant Measurements. *Journal of Nordic Hydrology*, 31(2), 2000.
- BENNIS S., BRUNEAU P., 1993a. Comparaison de méthodes d'estimation des débits journaliers. *Revue canadienne de génie civil*, 20, 480-489.
- BENNIS S., BRUNEAU P., 1993b. Amélioration de méthodes d'estimation des débits journaliers. *Revue canadienne de génie civil*, 20, 490-499.
- BERRADA F., 1998. Contribution à l'analyse des séries chronologiques en hydrologie et élaboration de procédés de validation et de reconstitution des données hydrométriques. Doctorat d'état. Université Hassan II. Casablanca, Maroc. 153 p.
- BERRADA F., GAGNON L., BENNIS S., 1996. Validation des données hydrométriques par des techniques univariées de filtrage. *Revue Canadienne de Génie Civi.*, 23, 872-892.
- BLANCHET F., BREUIL B. VIOLA A., 1998. Aquaval : Un système d'acquisition et de validation automatique des mesure en réseau d'assainissement. Exemple d'application sur le département de la Seine Saint-Denis. Novatech 1998. 3^e conférence internationale sur les nouvelles technologies en assainissement pluvial. Lyon 4-6 mai 1998. 2, 155-162.
- BOX G.E.P., JENKINS G.M., 1976. *Time Series Analysis: forecasting and control*. Édition révisée, Holden-Day. Oakland, California. 46-82.
- CHOCAT B., 1997. Encyclopédie de l'hydrologie urbaine et de l'assainissement. Paris, Tec & Doc Lavoisier. ISBN : 2-7430-0126-7. 183 p.
- ENEL F., 1987. La métrologie et la gestion des réseaux d'assainissement. Rapport du SCORENA, Échirolles, France.
- GAGNON L., 1994. Filtrage des séries hydrométéorologiques. Mémoire de maîtrise en Technologie des Systèmes. ETS, Montréal.

- HUBER W.C., DICKINSON R.E., 1988. Storm Water Management Model: User's manual. Version 4. University of Florida, Gainesville, Florida.
- KALMAN R.E., 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* (Transactions of the ASME, série D), 82, 34-45.
- KALMAN R.E., BUCY R.S., 1961. New results in linear filtering and prediction theory. *Journal of Basic Engineering* (Transactions of the ASME, série D), 82, 95-107.
- MCCUEN R.H., SNYDER W.M., 1986. Hydrologic modelling: statistical methods and applications. Prentice-Hall, Englewood Cliffs.
- NGUYEN V.T.V., BISSON J.L., 1998. Validation en temps réel des données des apports naturels journaliers pour la gestion des réservoirs. *Journal Canadien du Génie Civi.*, 25, 1096-1102.
- PERRAULT L., ROY R., MATHIER L., BOBÉE B., 1995. La combinaison de modèles appliquée à la validation en temps réel des apports naturels aux réservoirs hydriques. *Revue Canadienne de Génie Civil*, 22(5), 934-944.
- PLEAU M., PELLETIER G., LAVALLÉE P., 2000. Un nouveau logiciel de gestion et de conception des réseaux d'assainissement : MED. 2^e Symposium International Québec-Paris sur « La réhabilitation et l'aménagement des cours d'eau en milieu urbain. » Paris. 18-20 octobre 2000. 775-790.
- RADIX C., 1984. Filtrage et lissage statistiques optimaux linéaires, Cepadues-Éditions. Toulouse, France, 46-73 et 163-165.
- ROY R., GAUTHIER R., HÉTU M., 1992. Hydrometeorological Data Quality Improvement by Implementing an Autonomous Acquisition Network in Barren Areas. Hydraulic Engineering Software IV, Computer Techniques and Applications. Southampton, Boston. 351-359.
- TODINI E., 1978. Mutually Interactive State-Parameters Estimation (MISP) estimation in application of Kalman filter to hydrology, hydrolics and water resources. Édité par C.L. Chiu. Proceeding of the American Geophysics Union (AGU). Chapman Conference, University of Pittsburgh, Pittsburgh, 135-151.