

Record Number: 18360
Author, Monographic: Haché, M.//Bouchard, A.//Cantin, J. F.//Bobée, B.
Author Role:
Title, Monographic: Pré-validation des données hydrométriques horaires en temps réel
Translated Title:
Reprint Status:
Edition:
Author, Subsidiary:
Author Role:
Place of Publication: Québec
Publisher Name: INRS-Eau
Date of Publication: 1998
Original Publication Date: Septembre 1998
Volume Identification:
Extent of Work: x, 38
Packaging Method: pages incluant un annexe
Series Editor:
Series Editor Role:
Series Title: INRS-Eau, rapport de recherche
Series Volume ID: 536
Location/URL:
ISBN: 2-89146-306-4
Notes: Rapport annuel 1998-1999
Abstract: Rapport rédidé pour Environnement Canada, Direction de l'environnement atmosphérique, Division Monitoring et Technologies, Services hydriques
Chaire en hydrologie statistique Hydro-Québec / CRSNG
Call Number: R000536
Keywords: rapport/ ok/ dl

**PRÉ-VALIDATION DES DONNÉES
HYDROMÉTRIQUES HORAIRES
EN TEMPS RÉEL**

**PRÉ-VALIDATION DES DONNÉES HYDROMÉTRIQUES
HORAIRES EN TEMPS RÉEL**

Rapport rédigé pour

**Environnement Canada
Direction de l'environnement atmosphérique
Division Monitoring et Technologies
Services hydriques**

par

**Mario Haché
André Bouchard
Jean-François Cantin
Bernard Bobée**

**Chaire en Hydrologie statistique Hydro-Québec / CRSNG
Institut national de la Recherche scientifique, INRS-Eau
2800, rue Einstein, CP 7500, Sainte-Foy, Québec, G1V 4C7**

Rapport de recherche No R-536

Septembre 1998

©

INRS-Eau, 1998
ISBN 2-89146-306-4

ÉQUIPE DE RECHERCHE

CHARGÉ DE PROJET	INRS-Eau	MARIO HACHÉ
CHARGÉS DE PROJET	Environnement Canada	ANDRÉ BOUCHARD JEAN-FRANÇOIS CANTIN
COLLABORATEUR	INRS-Eau	FABRICE DÉSILES
TITULAIRE DE LA CHAIRE	INRS-Eau	BERNARD BOBÉE

TABLE DES MATIÈRES

LISTE DES FIGURES	vii
LISTE DES TABLEAUX	ix
CHAPITRE 1 : Introduction	1
1.1 Problématique	1
1.2 Objectifs du projet	2
CHAPITRE 2 : Généralités sur la pré-validation	5
2.1 Contrôle du domaine de variation	8
2.2 Contrôle de la cohérence temporelle	9
2.2.1. Taux de variation	9
2.2.2. Valeurs constantes	10
CHAPITRE 3 : Détermination des seuils critiques	11
3.1 Contrôle du domaine de variation	11
3.2 Contrôle de la cohérence temporelle	14
3.2.1. Taux de variation	14
3.2.2. Valeurs constantes	17
CHAPITRE 4 : Recommandations sommaires	19
4.1 Estimation des débits par rapport de superficies de bassins versants	19
4.2 Reconstitution des données manquantes	23
4.3 Pré-validation des valeurs hivernales	25
CHAPITRE 5 : Conclusion	29
CHAPITRE 6 : Bibliographie	31
ANNEXE A : Programmes <i>Matlab</i>	33

LISTE DES FIGURES

Figure 2.1.	Illustration des ensembles de données de la procédure de validation.....	6
Figure 3.1.	Seuils du contrôle du domaine de variation obtenus pour la station 02OA003 (Mille-Iles Bois-des-Fillions) à partir des maximums et minimums journaliers (lignes minces) et des quantiles 1% et 99% d'une loi log-normale (lignes grasses).	12
Figure 3.2.	Histogramme du taux de variation des niveaux horaires pour la station 02OJ016 (Richelieu St-Jean) pour l'année 1997.....	15
Figure 3.3.	Seuils du contrôle du taux de variation des niveaux horaires (cohérence temporelle) obtenus pour la station 02OJ016 (Richelieu St-Jean) pour l'année 1997.	16
Figure 3.4.	Histogramme de la durée des épisodes de valeurs constantes de niveaux horaires pour la station 02OJ007 (Richelieu Fryers) pour les années 1994-1997. Le seuil critique ($k_{0,99}$) est 8 heures.....	18
Figure 4.1.	Débits aux stations 02PL005 et 02PL007 et débits estimés à la station 02PL007 pour l'année 1993.....	22
Figure 4.2.	Erreur relative des débits estimés à 02PL007 pour l'année 1993.	22

LISTE DES TABLEAUX

Tableau 2.1	Analogie entre le test statistique et la procédure de validation.....	8
--------------------	---	---

1 INTRODUCTION

Le groupe Services hydriques de la Direction de l'environnement atmosphérique (DEA) d'Environnement Canada a donné à la Chaire en Hydrologie statistique Hydro-Québec/CRSNG établie à l'INRS-Eau le mandat de réaliser une étude portant sur la pré-validation automatisée de données horaires récoltées en temps réel. Cette étude se divise en trois phases:

- A. Validation des critères de contrôle de la qualité envisagés par le groupe Services hydriques;
- B. Calcul des critères statistiques pour chacun des sites considérés dans l'étude; et
- C. Recommandations sommaires relatives aux points suivants :
 - la pré-validation des valeurs hivernales (effet de glace);
 - l'estimation des débits à l'embouchure de la rivière par rapport de superficies de bassins versants;
 - la reconstitution des données manquantes.

Ce document vise à présenter les travaux de recherche réalisés dans le cadre de cette étude.

1.1 Problématique

La connaissance des données hydrométriques horaires en temps réel, en particulier les niveaux d'eau et les débits des rivières, est importante pour le groupe Services hydriques d'Environnement Canada. En effet, ces données acquises en temps réel sont souvent utilisées à des fins opérationnelles pour alimenter les modèles hydrodynamiques 1-D, 2-D et 3-D du fleuve St-Laurent. Ces modèles, opérés par Environnement Canada et le ministère des Pêches et des Océans, sont utilisés pour la prévision des niveaux d'eau et la réponse aux urgences maritimes entre autres.

La fiabilité des systèmes basés sur la modélisation hydrodynamique du fleuve Saint-Laurent repose en grande partie sur la bonne qualité des données hydrométriques qui sont utilisées comme intrants aux modèles. Or, ces données peuvent être entachées d'erreurs ou encore manquantes (bris d'appareil, problèmes de transmission, effet de glace). Les erreurs associées à la mesure des variables hydrométriques en temps réel conduisent à des résultats moins fiables et même erronés. Il importe donc de valider et éventuellement de remplacer ces mesures en temps réel.

Contrairement à la validation des données en temps différé (données historiques) qui est bien documentée, il existe peu d'études concernant la validation en temps réel des mesures hydrométriques. Plusieurs raisons peuvent expliquer cette situation, entre autres :

- les contraintes pratiques du temps réel puisque, contrairement à la validation des données en temps différé, les données aux temps ultérieur $t+1$, $t+2$, ... ne sont pas disponibles lors de la validation ou de l'estimation au temps t ;
- le pas de temps est petit (horaire dans le cas du présent projet alors qu'habituellement il est journalier);
- la nature très variable des données hydrologiques (processus aléatoire influencé par de nombreux facteurs naturels);
- les caractéristiques d'occurrence des erreurs (causes, fréquence, amplitude) sont mal connues.

1.2 Objectifs du projet

Les objectifs du projet sont, d'une part, de valider les contrôles de la qualité des données envisagés par le groupe Services hydriques et, d'autre part, de procéder à la définition des critères et au calcul des seuils pour les différents contrôles. De plus, des recommandations sommaires concernant le traitement des valeurs hivernales, l'estimation du débit par rapport de superficies de bassins versants et la reconstitution des données manquantes sont fournies.

Les stations hydrométriques de suivi pour lesquelles Environnement Canada (région du Québec) a un intérêt sont celles :

- 1) du projet de modélisation hydrodynamique d'Environnement Canada (Suivi) (59 sites);
- 2) du projet de détection/suivi des changements climatiques (19 sites);
- 3) du projet CRRO (Conseil de régularisation de la rivière des Outaouais) (19 sites);
- 4) d'autres intérêts fédéraux (11 sites).

La pré-validation des données du projet de modélisation hydrodynamique d'Environnement Canada (groupe 1) est prioritaire pour le groupe Services hydriques. La présente étude vise donc à calculer les critères pour les différents contrôles de validation pour les sites appartenant à ce projet. Les groupes 3 et 4 seront évalués par la suite en regard du besoin de pré-validation qui dépend des clients et des usages des données. Quant au groupe 2, il n'est pas nécessaire d'effectuer une pré-validation car la détection et le suivi des changements climatiques ne se fait pas en temps réel.

Ce rapport décrit les méthodes de pré-validation qui seront utilisées par les Services hydriques. La section 2 présente les différents contrôles de la qualité des mesures. La section 3 donne ensuite l'approche adoptée pour définir les seuils nécessaires aux contrôles de la qualité. Enfin, la section 4 présente les diverses recommandations sommaires concernant les points soulevés par le groupe Services hydriques. Les programmes utilisés pour extraire les seuils sont fournis à l'annexe A.

2 GÉNÉRALITÉS SUR LA PRÉ-VALIDATION

La validation consiste à effectuer un contrôle de la qualité des observations. D'une manière générale, ce contrôle permet de déterminer si une donnée mesurée ou calculée est erronée ou non. En effet, plusieurs sources d'erreurs peuvent affecter la qualité des données, en particulier celles relatives aux niveaux d'eau et aux débits des rivières. Pour Environnement Canada, la validation est un processus qui conduit à l'archivage des données sur la banque de données *HYDAT* (HYDAT, 1997) alors qu'un système de pré-validation permet la détection des erreurs plus ou moins grossières en temps réel et la prise d'actions en cas d'anomalies. Les actions entreprises consistent à avertir les personnes appropriées d'un problème et à remplacer la valeur erronée ou manquante par la meilleure valeur qui peut être estimée. Une procédure de pré-validation en temps réel doit détecter les données erronées et prévenir leur utilisation, ou du moins, aviser l'utilisateur de leur existence. En général, une donnée singulière après détection est signalée par un indicateur de qualité ou un code ("flag") qui spécifie le type de contrôle de la qualité échoué. Ces données pré-validées sont ensuite archivées avec leur indicateur de qualité dans une banque de données temporaire jusqu'à leur validation définitive et leur publication dans *HYDAT*.

Les données singulières sont celles détectées par la procédure de pré-validation. Elles peuvent être erronées ou non mais elles sortent en général de l'ordinaire. Les données aberrantes sont celles qui sont réellement erronées. Normalement, une valeur singulière détectée par le système de pré-validation est examinée par un spécialiste au moment de la validation afin de déterminer s'il s'agit :

- d'une valeur extrême réelle qu'il est alors très important de conserver; ou
- d'une valeur aberrante que l'on doit éliminer.

La décision finale tient compte des conditions hydrométéorologiques du bassin versant concerné et des bassins versants voisins.

La figure 2.1 donne une illustration des différents ensembles de données qu'implique une procédure de pré-validation. La totalité des données est composée par les ensembles A, B, C et D. Les données détectées par la procédure de pré-validation (données singulières) sont la somme des ensembles B et D. Les ensembles A et B forment la partie des données fiables alors que les ensembles D et C forment la partie des données aberrantes.

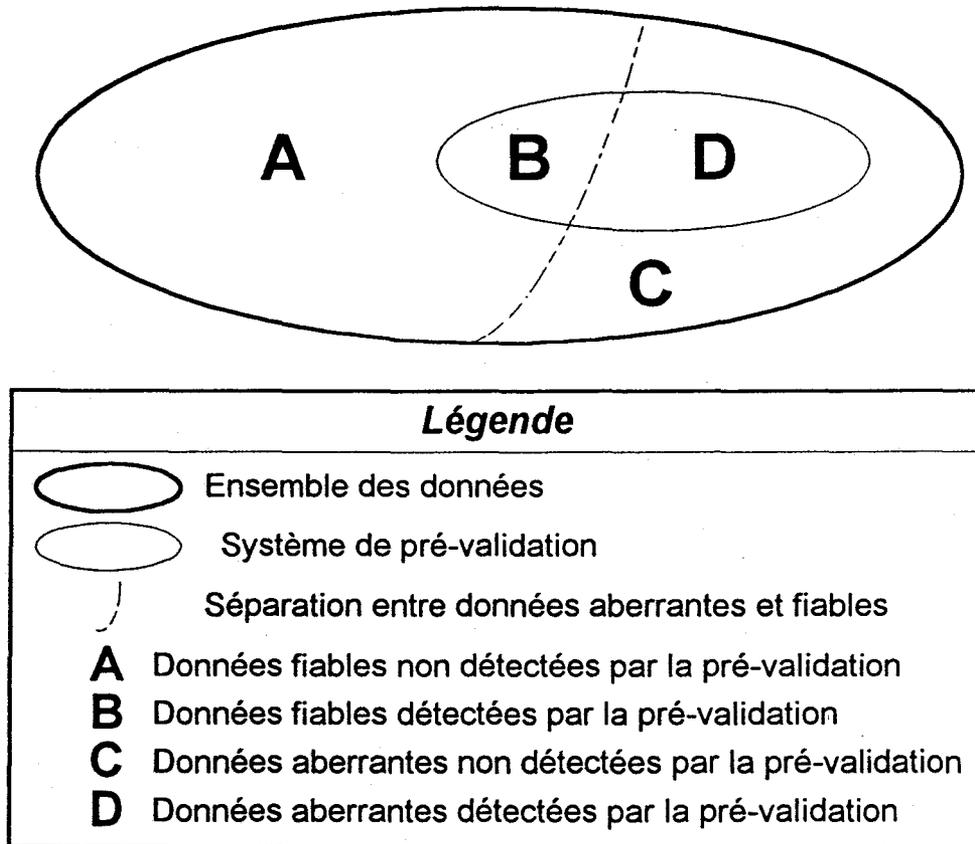


Figure 2.1 Illustration des ensembles de données de la procédure de pré-validation

L'efficacité d'un système de contrôle de la qualité se mesure par un faible taux de données aberrantes non détectées (données aberrantes considérées fiables par la validation : ensemble C) et un faible taux de fausses alarmes (données singulières détectées mais non aberrantes : ensemble B). Plus les critères de la validation sont sévères et plus l'ensemble C devient petit mais en contrepartie plus l'ensemble B devient grand. Inversement, plus les critères de la validation sont permissifs et plus l'ensemble B devient petit mais l'ensemble C devient grand. Le degré de sévérité des critères dépend donc du degré de tolérance de l'utilisateur face au

nombre de fausses alarmes (ensemble B) et au nombre de données aberrantes non détectées (ensemble C). Les critères devraient être tels qu'ils minimisent les deux ensembles B et C en même temps. Mais, une différente politique peut être appliquée dépendant de la gestion des données et de leur utilisation. Par exemple, un trop grand taux de fausses alarmes (ensemble B) peut engorger le système de validation, particulièrement si une présence humaine est nécessaire pour vérifier si les données singulières sont aberrantes, et rendre la gestion des données complètement inopérante. Il est dans ce cas plus important de minimiser l'ensemble B que l'ensemble C. Un autre exemple est l'utilisation d'un trop grand nombre de données aberrantes (ensemble C) comme intrants à des modèles qui peut rendre les résultats de ces derniers complètement erronés. Ici, il est plus approprié de minimiser l'ensemble C que l'ensemble B.

La procédure de validation s'apparente aux tests statistiques entre une hypothèse nulle (H_0) et une hypothèse alternative (H_A). Le tableau 2.1 illustre les deux types d'erreurs possibles lors d'un test et l'analogie avec la procédure de validation. Idéalement, les erreurs de type I et de type II doivent être minimisées. Mais tout comme dans le cas des ensembles B et C, si on minimise la probabilité de faire une erreur de type I alors on augmente celle du type II et vice-versa.

Les procédures de contrôle des données retenues par le groupe Services hydriques sont celles du domaine de variation et de la cohérence temporelle. Ces deux types de contrôle sont communément employés pour la validation des données, notamment pour le contrôle de la qualité de données météorologiques et hydrométriques. Ils sont suggérés par l'Organisation météorologique mondiale (OMM, 1985 et 1986) et utilisés par le Groupe Ressources hydriques (GRH) de la Société d'électrolyse et de chimie Alcan Ltée (SÉCAL) (Haché et Larouche, 1996), le Service de l'environnement atmosphérique (SEA) d'Environnement Canada (SEA, 1991), la Direction des réseaux atmosphériques (DRA) du Ministère de l'environnement et de la faune du Québec (MEFQ) et la Direction Planification de l'exploitation du parc d'équipement d'Hydro-Québec (HQ). Les deux sections suivantes décrivent de manière générale ces deux types de contrôle.

Tableau 2.1 Analogie entre le test statistique et la procédure de validation

Test statistique

Conclusion du test	Réalité	
	H ₀ est vraie	H _A est vraie
H ₀ (non rejet de H ₀)	Décision correcte	Décision incorrecte (erreur de type II)
H _A (rejet de H ₀)	Décision incorrecte (erreur de type I)	Décision correcte

Procédure de validation

Conclusion de la validation	Réalité	
	Données fiables	Données aberrantes
Données fiables	A	C
Données singulières	B	D

2.1 Contrôle du domaine de variation

La procédure de contrôle du domaine de variation consiste à vérifier si la valeur mesurée au temps t , par exemple le niveau d'eau dans une rivière N_t , est à l'intérieur d'un certain intervalle délimité par une borne supérieure et une borne inférieure:

$$[\text{limite inférieure}] \leq N_t \leq [\text{limite supérieure}] \quad (2.1)$$

Les limites utilisées lors du contrôle du domaine de variation sont:

- des limites absolues, c'est-à-dire des limites impossibles à dépasser pour des raisons physiques; et
- des limites relatives, c'est-à-dire des limites qui ont une faible probabilité d'être dépassées.

Les limites absolues peuvent être définies par des critères physiques de la variable mesurée ou par l'horizon d'enregistrement de l'instrument de mesure utilisé. Par exemple, il n'est pas

possible d'obtenir un débit négatif ou un niveau d'eau supérieur à la limite maximum de l'appareil d'enregistrement.

Les limites relatives communément employées sont :

- les extrêmes historiques (valeurs maximum et minimum);
- $\mu \pm k\sigma$ où μ est la moyenne, σ l'écart-type et k est une constante fixée en fonction du niveau de confiance désiré;
- un quantile (1%, 5% ou 10%).

Les limites absolues et relatives peuvent être fixes ou variables dans le temps. Les limites variables ont l'avantage de tenir compte de la saisonnalité des observations souvent non négligeable dans le cas des phénomènes hydrométéorologiques.

2.2 Contrôle de la cohérence temporelle

La procédure de contrôle de la cohérence temporelle permet de vérifier si la donnée mesurée n'est pas trop éloignée des valeurs précédentes (taux de variation). Elle rend également possible la détection des données qui ne présentent aucune variation temporelle depuis un certain temps (valeurs constantes). Ce type de contrôle est particulièrement efficace pour des variables ayant une forte dépendance temporelle.

2.2.1 Taux de variation

La première étape du contrôle de la cohérence temporelle s'exécute en comparant la valeur mesurée au temps t avec celles mesurées aux temps $t-1$, $t-2$, ... Cette comparaison peut s'appliquer en vérifiant si le taux de changement de la valeur enregistrée ne dépasse pas un certain seuil:

$$|\Delta N_t| = |N_t - N_{t-1}| \leq [\text{seuil critique}] \quad (2.2)$$

Le contrôle de la cohérence temporelle peut également être effectué en comparant la valeur mesurée avec une valeur prévue par un modèle de prévision temporelle, \hat{N}_t :

$$|N_t - \hat{N}_t| \leq [\text{seuil critique}] \quad (2.3)$$

La valeur prévue \hat{N}_t peut être déterminée à partir d'un modèle ARIMA ou par simple extrapolation linéaire ou parabolique. Le seuil critique est défini par des considérations physiques ou par des analyses statistiques.

2.2.2 Valeurs constantes

La deuxième étape du contrôle de la cohérence temporelle est de vérifier les épisodes de valeurs consécutives qui ne changent pas. Ces épisodes de données constantes sont souvent indicatifs d'un mauvais fonctionnement de l'équipement à la station de mesure ("gel" des équipements). Le test sur les épisodes constants se fait de la manière suivante:

$$\text{Soient } N_{t+1} = N_{t+2} = \dots = N_{t+k} \text{ alors } k \leq [\text{seuil critique}] \quad (2.4)$$

Ici aussi, le seuil critique est défini par des considérations physiques ou par des analyses statistiques.

3 DÉTERMINATION DES SEUILS CRITIQUES

Ce chapitre présente les procédures suivies pour déterminer les différents seuils nécessaires à chacun des contrôles pour la validation des données. Ces procédures ont été programmées dans le langage *MATLAB* Version 5.0. Les programmes utilisés sont fournis à l'annexe A.

Les contrôles peuvent être appliqués sur les débits ou sur les niveaux, tout dépendant de l'information qui est disponible. Toutefois, il ne faut pas oublier que ces deux variables sont reliées et que corriger l'une revient à corriger l'autre. Il faut donc porter un soin particulier pour conserver l'homogénéité des banques de données lorsque les deux variables sont disponibles.

3.1 Contrôle du domaine de variation

L'étape du contrôle sur le domaine de variation se fait en vérifiant si la valeur observée appartient à un certain domaine réaliste du phénomène mesuré (niveau d'eau et débit des rivières). Les limites envisagées par les Services hydriques sont des limites relatives variables dans le temps. Il s'agit des valeurs minimums et maximums calculées sur une base journalière. Ces limites journalières seront ensuite utilisées pour le contrôle de la qualité des données horaires. Les données employées pour l'extraction des statistiques proviennent de la banque de données nationale *HYDAT* (HYDAT, 1997). Un exemple graphique des seuils obtenus pour la station 02OA003 est donné à la figure 3.1 (lignes minces).

Quelques points faibles sont à souligner concernant cette approche :

- Les valeurs horaires peuvent atteindre des limites plus extrêmes que les valeurs journalières. En effet, les données journalières représentent une moyenne des données horaires. Ainsi, les seuils de contrôle calculés sur une base journalière sont moins sévères que ceux qui seraient calculés sur une base horaire.
- Les maximums et les minimums dépendent de la longueur de la série. Ainsi, la sévérité des seuils n'est pas homogène d'une station à l'autre et elle est fortement reliée à la

longueur de la série. Par exemple, les seuils calculés sur une période de 100 ans sont généralement moins sévères que ceux calculés sur une période de 10 ans.

- Les valeurs de maximums et de minimums apparaissent très variables d'un jour à l'autre (cf. fig. 3.1) alors qu'il n'y a pas de raison qu'un seuil critique change énormément d'un jour à l'autre.

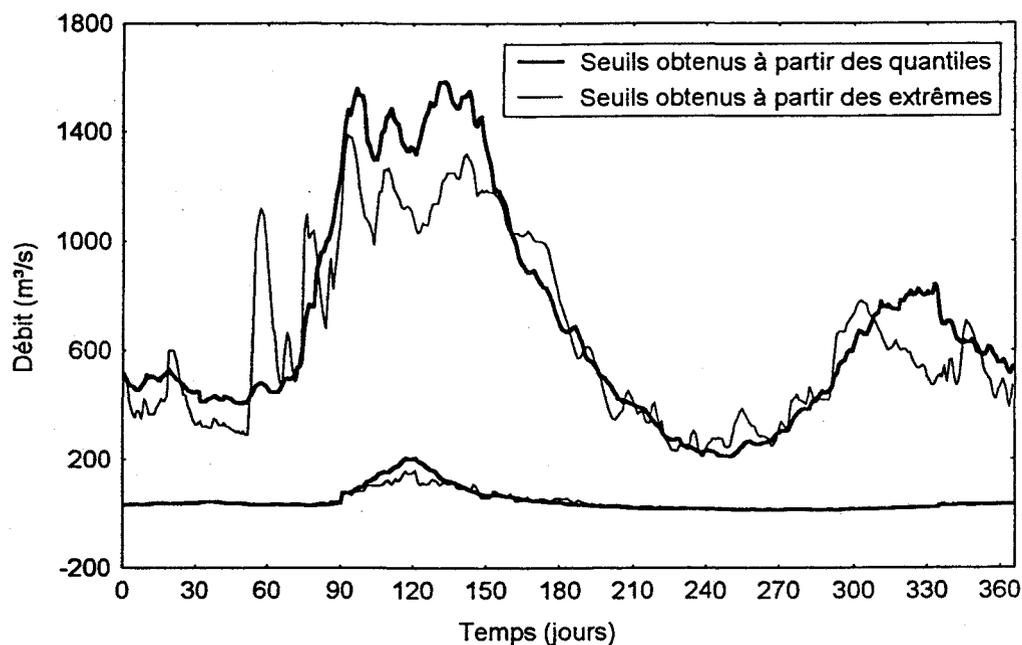


Figure 3.1 Seuils du contrôle du domaine de variation obtenus pour la station 02OA003 (Mille-Iles Bois-des-Fillions) à partir des maximums et minimums journaliers (lignes minces) et des quantiles 1% et 99% d'une loi log-normale (lignes grasses).

Les séries de données horaires ne sont pas assez longues (4 ou 5 ans au maximum) pour permettre de calculer des valeurs minimums et maximums qui rendent le contrôle du domaine de variation adéquat. Il n'est donc pas possible de calculer ces valeurs autrement qu'en utilisant les séries beaucoup plus longues de données journalières. Les données horaires seront éventuellement utilisées dans le futur pour raffiner les limites lorsque les séries seront suffisamment longues.

L'utilisation des quantiles plutôt que des valeurs extrêmes ferait en sorte que toutes les stations auraient des critères de même sévérité qui ne dépendraient pas de la longueur des séries. De plus, les quantiles n'ont pas tendance à changer énormément d'un jour à l'autre comme dans le cas des valeurs extrêmes. Dans la figure 3.1, on a également représenté les seuils que l'on aurait obtenus en supposant une distribution log-normale à deux paramètres pour les débits à chaque jour et en déterminant les quantiles de probabilité au dépassement 1% et 99% (lignes grasses). On remarque que la courbe des quantiles est plus lisse que celle des extrêmes sur ce graphique. Un autre avantage lié à l'utilisation des quantiles est de pouvoir changer la sévérité des critères en modifiant simplement la probabilité au dépassement.

Une procédure qui ajuste une distribution statistique aux observations d'un jour donné de l'année et qui calcule les quantiles désirés comme seuils critiques est tout à fait envisageable. Mais la tâche est grande et dépasse largement le cadre de la présente étude. En effet, cela représente 365 ajustements pour chacune des stations, ce qui est énorme. De plus, quelques questions essentielles se posent :

- Quelles distributions suivent les débits journaliers et les niveaux journaliers?
- Quels critères utiliser pour discriminer les distributions?
- Quelles méthodes automatiques employer pour choisir la meilleure distribution pour représenter les données?

L'utilisation des maximums et minimums comme seuils est adéquate si l'on considère l'objectif du contrôle du domaine de variation qui est d'éliminer les données les plus anormales. C'est un contrôle grossier, si on le compare au contrôle de la cohérence temporelle, qui sert à détecter les données singulières majeures. Dans cette optique, l'approche envisagée par le groupe Services hydriques est tout à fait valable. Bien sûr, il serait possible de raffiner la méthode en considérant les quantiles comme seuils dans une phase ultérieure du projet de validation.

3.2 Contrôle de la cohérence temporelle

L'examen de la cohérence temporelle proposé consiste à comparer la valeur observée au temps t avec la valeur précédente mesurée au temps $t-1$ (pas de temps horaire). Cette section présente le cheminement conduisant à la détermination des seuils critiques pour le taux de variation et la longueur des épisodes de valeurs constantes.

3.2.1 Taux de variation

La détermination du seuil critique du contrôle de la cohérence temporelle des niveaux d'eau et des débits est fonction des caractéristiques statistiques du taux de variation, des propriétés physiques du bassin versant et de la précision de l'instrument de mesure.

L'utilisation des données horaires est préconisée pour le calcul des critères de cohérence temporelle lorsque la période d'enregistrement est assez grande (au moins une année). Une donnée journalière de niveau ou de débit représente une moyenne sur 24 heures et ne permet pas d'avoir une idée précise des variations horaires. Le calcul d'un taux de variation journalier a pour effet de lisser le taux de variation horaire et il n'est donc pas possible de définir des seuils de contrôle fiables à partir des données journalières. En particulier, le taux de variation du débit des petits bassins peut changer substantiellement durant une journée car la réponse du bassin est rapide lorsqu'il est soumis à une précipitation importante.

Étant donnée la courte période d'enregistrement des données horaires, le calcul d'un seuil critique pour chaque jour de l'année comme dans le cas du contrôle du domaine de variation n'est pas réalisable. Les changements d'une valeur à l'autre, contrairement aux valeurs elles-mêmes, ne montrent pas nécessairement un patron saisonnier prononcé. Dans certains cas, un seuil unique peut-être acceptable et, dans d'autre cas, non. La division de l'année en trois saisons (hiver, printemps, été-automne) et le calcul d'un seuil pour chacune des saisons est un compromis acceptable. Une saison de données sur quelques années est suffisante (au moins une année) car elle représente un éventail assez large de taux de variation. Malgré une période d'enregistrement relativement courte et la présence de données aberrantes (les données horaires ne sont pas validées), une étude statistique des différences de niveaux ou

de débits horaires permet d'évaluer d'une manière réaliste les seuils critiques du contrôle de la cohérence temporelle du taux de variation.

Le taux de variation, ou première différence, est défini par l'équation:

$$\Delta N_t = N_t - N_{t-1} \quad (3.1)$$

où t est l'heure. La figure 3.2 donne un exemple d'un histogramme du taux de variation du niveau horaire à la station 02OJ016 (Richelieu St-Jean) pour l'année 1997. La distribution du taux de variation pour cette station est symétrique et elle a un mode (valeur correspondant à la fréquence plus élevée) nul. L'histogramme a l'allure d'une distribution normale.

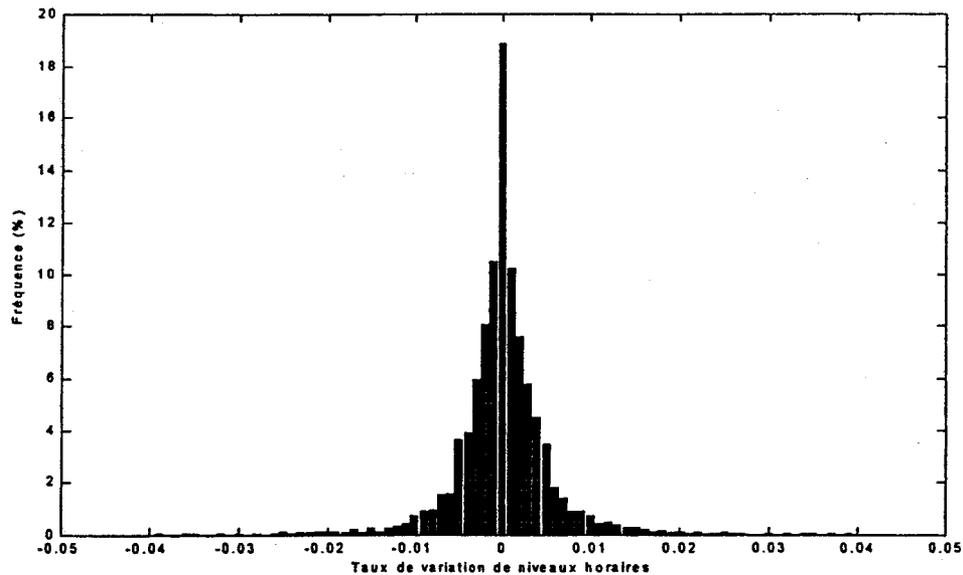


Figure 3.2 Histogramme du taux de variation des niveaux horaires pour la station 02OJ016 (Richelieu St-Jean) pour l'année 1997.

À partir d'une étude de fréquence sur les séries ΔN_t de chacune des stations, les deux quantiles suivants sont calculés empiriquement:

- $\Delta N_{0.005}$ tel que $\text{Prob}\{\Delta N_t \leq \Delta N_{0.005}\} = 0.005$ (Seuil inférieur)

- $\Delta N_{0.095}$ tel que $\text{Prob}\{\Delta N_t \leq \Delta N_{0.095}\} = 0.095$ (Seuil supérieur)

Les différences de niveaux horaires sont en dehors des deux limites $\Delta N_{0.005}$ et $\Delta N_{0.095}$ une fois sur 100, c'est-à-dire en moyenne environ une fois tous les 4 jours. Le choix de ces quantiles est arbitraire et peut être modifié. Par exemple, les quantiles $\Delta N_{0.05}$ et $\Delta N_{0.95}$ sont plus sévères alors que les quantiles $\Delta N_{0.0005}$ et $\Delta N_{0.0095}$ sont moins sévères.

La figure 3.3 donne un exemple des seuils établis pour le contrôle de la cohérence temporelle pour trois saisons. Dans cet exemple, les seuils ne fluctuent pas beaucoup d'une saison à l'autre. En considérant la précision des instruments de mesure et l'incertitude sur les quantiles calculés, on peut considérer qu'il n'y a pas de différence significative d'une saison à l'autre et il est acceptable de prendre un seuil unique pour toute l'année dans le cas des taux de variation des niveaux pour cette station. Les seuils calculés pour une autre station peuvent conduire à une toute autre conclusion et, en général, les seuils ne sont pas les mêmes d'une saison à l'autre en particulier pour le taux de variation du débit.

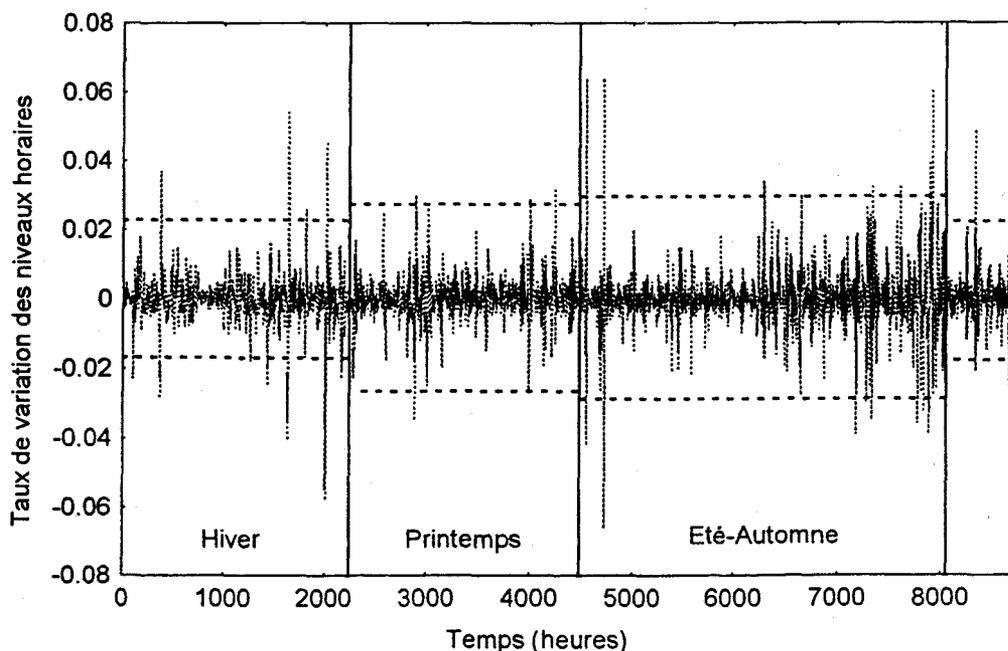


Figure 3.3 Seuils du contrôle du taux de variation des niveaux horaires (cohérence temporelle) obtenus pour la station 02OJ016 (Richelieu St-Jean) pour l'année 1997.

3.2.2 Valeurs constantes

Le contrôle de la cohérence temporelle est également utile pour détecter les épisodes de valeurs constantes dans les observations. Ces épisodes sont souvent l'indice d'un problème d'instrumentation. Il est alors important d'intervenir rapidement pour remédier au bris technique. Une étude de fréquence des épisodes de niveaux et de débits constants nous a permis de définir une durée limite supérieure révélatrice d'un problème.

La série de données est tout d'abord transformée en une série binaire (0 et 1) Y selon la règle suivante:

- $Y = 0$ si $|N_t - N_{t-1}| > 0$; et
- $Y = 1$ si $|N_t - N_{t-1}| = 0$.

Un épisode de données constantes de longueur k est défini par une série continue de $k-1$ observations horaires sans changement ($Y = 1$) bornée de deux observations horaires avec changement ($Y = 0$):

$$\{Y_t=0, Y_{t+1}=1, Y_{t+2}=1, \dots, Y_{t+k-2}=1, Y_{t+k-1}=1, Y_{t+k}=0\}$$

Les épisodes de durée $k=1$ correspondent aux $Y_t=0$. Enfin, la série de la durée des épisodes de données constantes est construite à partir de la série Y et de la définition d'un épisode. La figure 3.4 donne l'histogramme de la durée des épisodes constants pour la station 02OJ007 (Richelieu Fryers). La forme de l'histogramme est de type exponentiel.

Une analyse de fréquence empirique de la série des durées permet de calculer le quantile ayant une probabilité de dépassement donnée. Comme pour les seuils critiques afin de vérifier le taux de variation, le seuil critique pour les durées de valeurs constantes est défini par le quantile de probabilité au non-dépassement de 99% :

- $k_{0,99}$ tel que $\text{Prob}\{k < k_{0,99}\} = 0.99$ (seuil supérieur)

Naturellement, le test peut être moins sévère en augmentant la probabilité ou plus sévère en diminuant la probabilité.

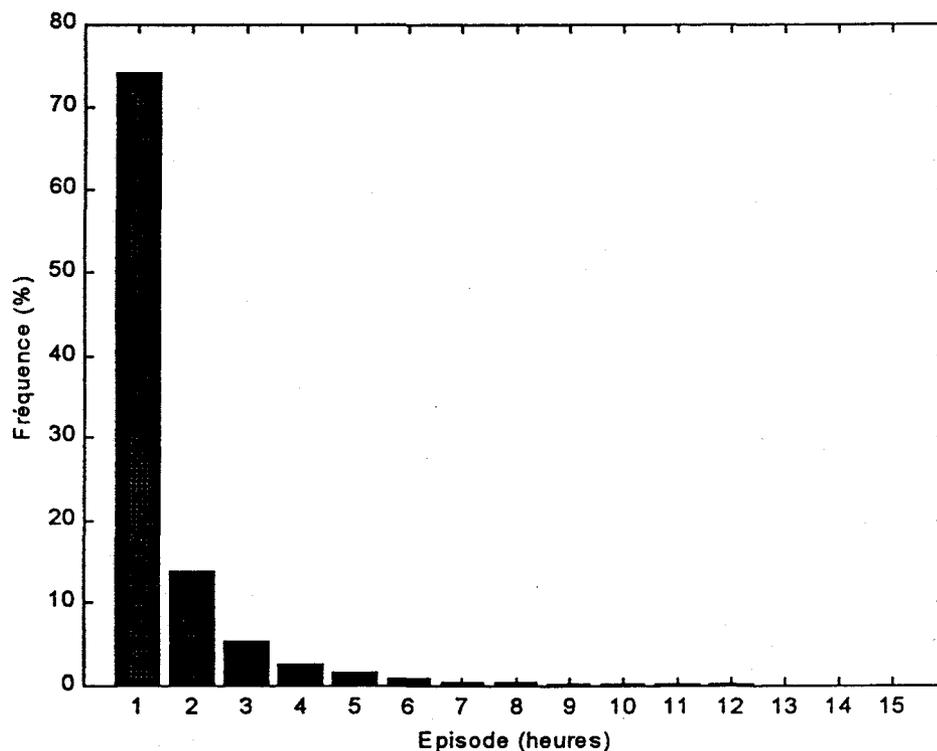


Figure 3.4 Histogramme de la durée des épisodes de valeurs constantes de niveaux horaires pour la station 02OJ007 (Richelieu Fryers) pour les années 1994-1997. Le seuil critique ($k_{0,99}$) est 8 heures.

4 RECOMMANDATIONS SOMMAIRES

Ce chapitre a pour objet de présenter quelques recommandations sommaires concernant trois points soulevés par l'équipe des Services hydriques d'Environnement Canada. Il s'agit des thèmes suivants :

- l'estimation des débits à l'embouchure de la rivière par rapport de superficies de bassins versants;
- la reconstitution des données manquantes;
- la pré-validation des valeurs hivernales (effet de glace).

4.1 Estimation des débits par rapport de superficies de bassins versants

La méthode de transfert de données par rapport de superficies de bassins versants s'apparente à celle de la régression multiple en ce sens qu'elle repose aussi sur une relation linéaire entre les apports de divers sites. Toutefois, cette relation ne s'appuie pas sur des techniques statistiques mais plutôt sur des prémisses hydrologiques et physiques : des sites voisins soumis à des conditions climatiques et physiographiques équivalentes devraient posséder le même débit spécifique. Le débit spécifique à un site est obtenu en divisant le débit par la superficie du bassin versant correspondant.

La procédure générale pour transférer le débit d'un site à l'embouchure de la rivière comme cela est préconisé dans le projet-fleuve (système d'aide à la décision sur le Saint-Laurent) peut s'effectuer par le biais du rapport de superficies de bassins versants. Pour un sous-bassin non-jaugé Y , on recherche un sous-bassin voisin X jaugé (situé préférablement en amont sur la même rivière). Habituellement, le sous-bassin Y englobe le sous-bassin X . Les débits du sous-bassin Y sont alors calculés à partir des débits spécifiques du sous-bassin X pour lequel on dispose de débits mesurés. Plus précisément, le débit \hat{Q}_Y estimé par cette méthode s'exprime de la façon suivante :

$$\hat{Q}_Y = \frac{A_Y}{A_X} Q_X \quad (4.1)$$

où A_X et A_Y désignent respectivement les superficies des sous-bassins versants X et Y et Q_X le débit du sous-bassin X (site de référence). Cette méthode peut être facilement généralisée de manière à utiliser plus d'un sous-bassin voisin. Le débit estimé devient alors une moyenne pondérée des apports aux sites voisins. Il est important de souligner que la relation (4.1) est applicable lorsque l'ensemble des sous-bassins utilisés possède des régimes hydrologiques et climatiques semblables.

Le paramètre de transfert $k=A_Y/A_X$ est constant et implique donc que la relation linéaire entre les données des deux sites est invariante dans le temps. Dans le cas d'un transfert de débit d'un site vers l'embouchure d'une rivière, la relation (4.1) est bien respectée si la différence de superficies entre les bassins versants n'est pas trop grande. Le fait que A_X est inclus dans A_Y implique que la climatologie et la physiographie des deux bassins sont, en général, assez semblables. Il y a cependant de fortes raisons de croire que cette approche appliquée à un pas de temps horaire engendre un biais à court terme lorsque la station de mesure est éloignée de l'embouchure de la rivière. En effet, on ne tient alors nullement compte du temps de cheminement de l'eau entre la station et l'embouchure qui peut être de plusieurs heures. Il est donc recommandé d'ajouter un délai pour tenir compte du temps de cheminement de l'eau dans la rivière en modifiant l'équation (4.1) de la manière suivante :

$$\hat{Q}_{Y,t} = \frac{A_Y}{A_X} Q_{X,t-d} \quad (4.2)$$

où t est l'heure et d est l'estimation du temps moyen (en heures) du cheminement de l'eau entre la station de mesure (X) et l'embouchure (Y).

L'évaluation du paramètre de transfert peut aussi être la cause d'une sur-estimation ou d'une sous-estimation systématique des débits au site d'intérêt. Les superficies de bassins versants sont calculées approximativement et peuvent donc être la source d'erreurs engendrant un biais à long terme. Enfin, la variance des données estimées \hat{Q}_Y est entièrement déterminée

et imposée par la méthode du transfert par superficies de bassins versants. Elle correspond à $k^2 \text{Var}\{Q_x\}$.

Malgré les lacunes de la méthode du transfert du débit par rapport de superficies de bassins versants, il n'existe pas d'autres méthodes aussi simples et aussi efficaces pour estimer approximativement le débit à l'embouchure d'une rivière où il n'y a jamais eu de débits mesurés. En effet, c'est dans l'absence de données à l'endroit où l'on désire estimer les débits que réside la difficulté de développer une relation entre les débits à un site jaugé et ceux à l'embouchure d'une rivière.

Il est préférable de procéder à une estimation du débit à l'embouchure de la rivière que d'utiliser directement les données de débits mesurés à une station en amont sur la rivière. Les données estimées par la méthode du transfert du débit sont certainement plus conformes à la réalité que celles mesurées à quelques ou plusieurs kilomètres en amont.

Dans le cadre de cette étude, nous n'avions pas à notre disposition de données horaires pour deux stations d'une même rivière afin d'évaluer sommairement l'efficacité de la méthode. Toutefois, les données de débits journaliers des stations 02PL007 et 02PL005 situées sur la rivière Bécancour étaient disponibles. La superficie du bassin versant de la première station située en amont de la rivière est de 919 km² et celle de la deuxième station située en aval est de 2330 km². Il s'agit donc d'une différence de superficie de près de 61% par rapport au plus grand bassin. En général, les différences de superficies de bassins versants entre une station et l'embouchure de la rivière sont moins grandes pour les affluents du Saint-Laurent.

Les données de la station 02PL005 ont été utilisées pour estimer les débits à la station 02PL007 en utilisant un paramètre de transfert $k = 2330 \text{ km}^2 / 919 \text{ km}^2 = 2.54$. La figure 4.1 donne le résultat de l'estimation du débit pour l'année 1993 à la station 02PL007. La courbe estimée est assez voisine de la courbe mesurée à la station 02PL007. Il est certainement plus raisonnable d'utiliser les données estimées que d'utiliser les données mesurées à la station 02PL005 pour représenter les débits de la station 02PL007. La figure 4.2 donne l'erreur relative du débit estimé à la station 02PL007 en 1993. L'erreur relative est définie de la manière suivante :

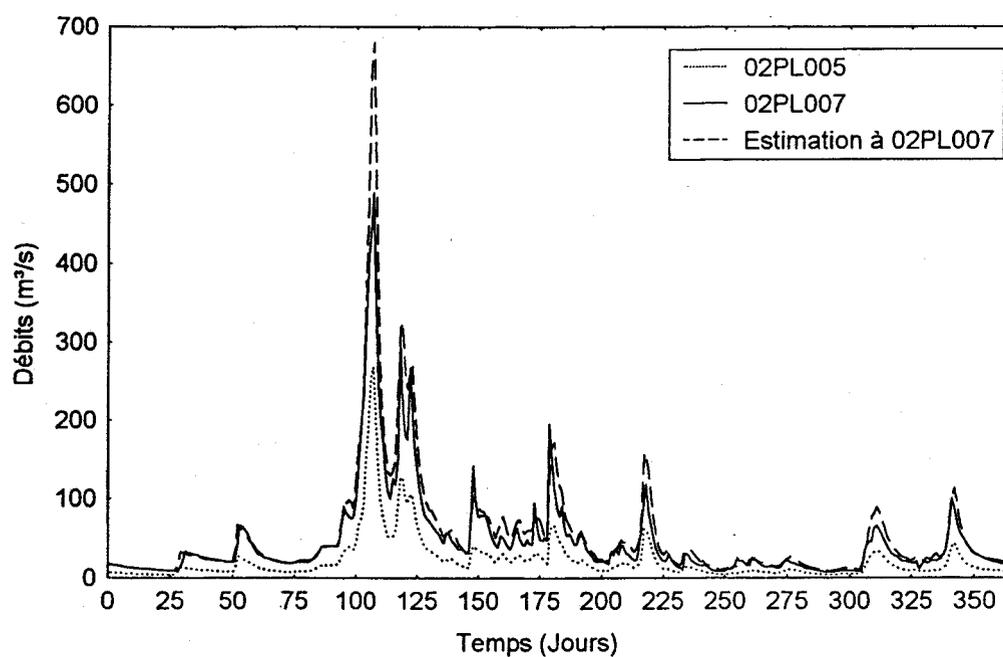


Figure 4.1 Débits aux stations 02PL005 et 02PL007 et débits estimés à la station 02PL007 pour l'année 1993.

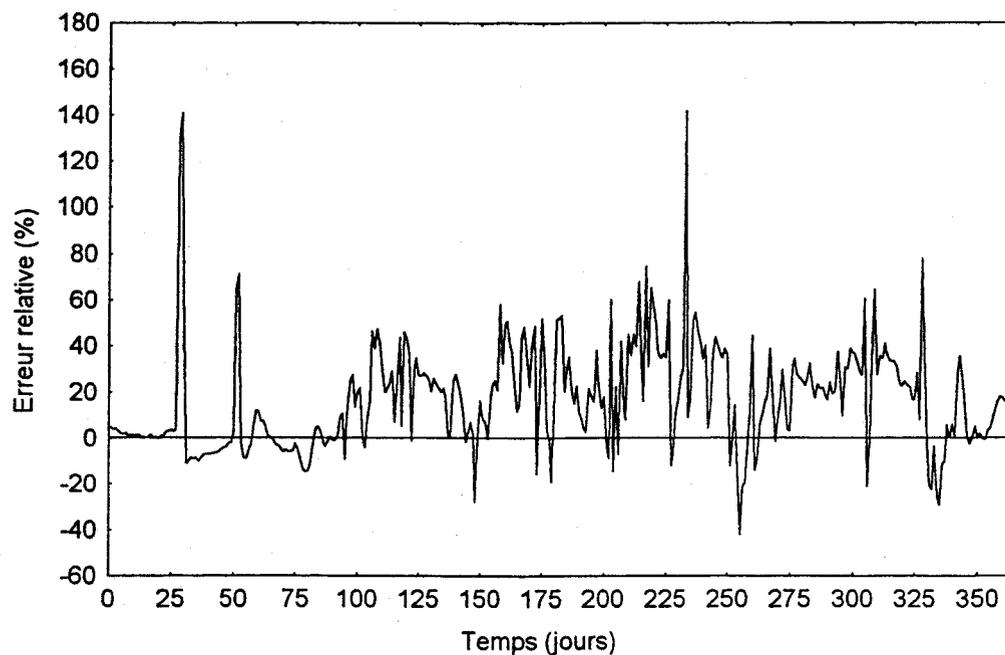


Figure 4.2 Erreur relative des débits estimés à 02PL007 pour l'année 1993.

$$ER = 100(\hat{Q}_Y - Q_Y) / Q_Y \quad (4.3)$$

L'erreur relative a tendance à être positive ce qui provient d'une sur-estimation du débit estimé pour cette station. Cela peut résulter d'une mauvaise estimation des superficies des bassins versants ou bien d'un régime pluviographique plus intense en amont qu'en aval de la rivière. L'erreur relative moyenne pour l'ensemble de la série historique (1970-1994) est de 9.3% et l'erreur relative en valeur absolue moyenne est de 21.3%.

4.2 Reconstitution des données manquantes

Les données de débits peuvent être manquantes en temps réel pour diverses raisons techniques ou pour des problèmes de transmission de l'information. De plus, les données rejetées par la procédure de pré-validation sont inutilisables. Pour un emploi pratique de l'information hydrométrique en temps réel, il importe d'estimer les données de débits lorsqu'elles sont manquantes et de les corriger lorsqu'elles sont rejetées par la pré-validation.

Il existe deux groupes principaux de méthodes pour la prévision des débits, soient les modèles stochastiques et conceptuels. Les modèles stochastiques font intervenir le concept que le phénomène mesuré peut être représenté par des variables aléatoires. De manière générale, l'estimation des données de débits en rivières par une méthode statistique peut s'effectuer en considérant deux approches :

- la modélisation spatiale qui prend en compte l'information régionale (données de débits des stations voisines), physiographique et météorologique à l'aide, par exemple, de méthodes régressives;
- la modélisation temporelle qui tient compte du caractère autorégressif des observations hydrologiques à partir, par exemple, des méthodes de type ARIMA ou d'une procédure de lissage.

Les modèles conceptuels sont basés sur des lois physiques ou des concepts empiriques qui permettent de prévoir une variable de sortie à partir de différentes variables d'entrée. Les

modèles physiques basés uniquement sur les lois physiques fondamentales (conservation de la masse et de l'énergie) et les machines hydrauliques font partis de cette catégorie.

Fortin *et al.* (1996) ont effectué une revue bibliographique des méthodes de prévisions des débits. Plusieurs modèles de prévision ont été évalués et comparés dans cette étude. Il s'agit :

- Modèles empiriques :
 - Modèles stochastiques
 - Modèles à base de réseau de neurones artificiels
 - Régression linéaire floue
 - Modèle de systèmes chaotiques
- Modèles conceptuels :
 - Modèles physiques
 - Machines hydrauliques
 - Modèles à base de règles floues

Un classement des méthodes pour différents pas de temps (prévisions à court, moyen et long terme) a été proposé. Ce sont la méthode des k plus proches voisins (Yakowitz, 1987) et les modèles de Box-Jenkins (Box et Jenkins, 1976) (ARIMA et ARMAX, selon le type de données disponibles) qui ont été suggérés comme modèles les plus appropriés et les plus simples à appliquer dans le cas de la prévision à court terme (horaire ou journalière). La forte corrélation entre les observations successives à cette échelle de mesure favorise ces deux types de modèles stochastiques.

Enfin, l'utilisation d'un filtre de Kalman (Kalman, 1960) en temps réel permet d'améliorer les prévisions et d'éliminer l'autocorrélation des erreurs d'estimation souvent présente à une échelle de temps horaire. Le filtre de Kalman modifie les paramètres du modèle de manière adaptative au fur et à mesure que de nouvelles données deviennent disponibles.

4.3 Pré-validation des valeurs hivernales

Les conditions hivernales conduisent à deux catégories de problèmes affectant la mesure des débits. Il s'agit de:

- la formation d'embâcles et les débâcles (problème surtout printanier);
- la présence d'une surface de glace sur la rivière (problème surtout hivernal).

Nous ne traiterons pas du premier cas dans ce présent travail qui sort du cadre de cette étude. Nous nous attarderons donc dans ce qui suit au deuxième cas.

La présence de glace à la surface ou au fond de la rivière influence grandement sa section et son gradient d'écoulement, et par conséquent, l'estimation du débit. L'augmentation de la pente de la ligne d'eau a pour effet de rehausser le niveau en un point donné sans pour autant s'accompagner d'une augmentation de débit. Dans de telles conditions, la courbe de tarage niveau-débit établie en condition libre de glace conduit à une sur-estimation du débit et ne devrait pas être utilisée.

L'évolution du couvert de glace et son effet sur les niveaux et les débits sont complexes. Il existe toutefois quelques points de repère permettant d'évaluer l'effet de glace sur le débit, en particulier :

- les jaugeages d'hiver (deux ou trois effectués par année);
- le débit des rivières voisines qui ne sont pas gelées;
- les apports naturels calculés par bilan hydrique en aval d'une station de mesure;
- les modèles conceptuels de l'effet de la glace sur le niveau;
- les modèles stochastiques d'estimation de débits;

L'effet de la glace sur le débit d'une rivière qui se jette dans un réservoir où l'apport naturel est calculé par bilan hydrique peut s'évaluer en comparant le rapport des débits (débit rivière

/ apport réservoir) et le rapport des superficies (bassin versant de la rivière / bassin versant du réservoir). Normalement, ces rapports devraient être égaux si les conditions climatiques et physiographiques sont similaires pour les deux bassins versants (section 4.1). Il est alors possible d'estimer grossièrement la sur-évaluation du débit dans la rivière. Cependant, les niveaux du réservoir nécessaires au bilan hydrique sont également influencés par la présence de la glace.

Il existe des modèles conceptuels basés sur la description physique simple de la formation et de la fonte du couvert de glace et qui permettent de déduire l'effet de la glace sur le niveau de l'eau dans la rivière (Huttenen *et al.*, 1997; Teal *et al.*, 1994; Hamilton et Moore, 1996). Ces modèles sont basés sur des équations d'écoulements qui peuvent tenir compte, par exemple, du type de sol, du coefficient de frottement avec la glace, de l'épaisseur de la glace, de la température, des précipitations et du rayonnement solaire.

Finalement, l'estimation des débits d'hiver peut être effectuée de la même manière que pour les données manquantes de débits (section 4.2). En particulier, les modèles stochastiques, les fonctions de transfert et les modèles à base de réseaux de neurones artificiels servent à estimer le débit lorsque la rivière est couverte de glace. Plus précisément, le débit corrigé d'hiver Q_c s'écrit selon l'équation suivante:

$$Q_c = f(Q_m, \Theta) \quad (4.4)$$

où Q_m est le débit d'hiver mesuré;

Θ est un vecteur de données physiographiques et météorologiques.

Lorsque Θ est connu, l'équation (4.4) correspond à une fonction de transfert alors que s'il est inconnu, elle correspond à un modèle à base de réseau de neurones artificiels.

Le problème essentiel pour ces modèles est de disposer de données hivernales validées de bonne qualité pour effectuer la calibration d'une manière efficace. Des travaux ont déjà débuté dans ce domaine. Le U.S. Geological Survey (USGS) et le Water Survey of Canada (WSC) ont constitué une banque de données pour la mesure des débits sous couvert de glace dans la perspective de recherche future dans ce domaine (Walker et Wang, 1997). De

plus, de nouveaux équipements pour mesurer le débit tels les profileurs acoustiques à effet Doppler et les débitmètres acoustiques de type AFFRA (Acoustic Flowmeter For Remote Area) (Environnement Canada, 1988) donnent des valeurs de débits justes même sous couvert de glace. Les données obtenues avec ce genre de matériel, si elles sont en quantité suffisante, peuvent servir pour calibrer le modèle (4.4).

Il serait important de développer une procédure rationnelle et efficace pour l'estimation des débits de rivières dans les conditions hivernales et l'équipe de la Chaire à l'intention d'effectuer des travaux dans cette direction avec des partenaires des milieux gouvernementaux provincial et fédéral. La procédure permettrait d'estimer la valeur du débit réel sous couvert de glace, étant donné la valeur enregistrée du débit, un ensemble de séries temporelles décrivant l'évolution d'un nombre de variables explicatives (tels que la température, le rayonnement, le couvert de neige, l'emmagasinement à l'amont, etc.), et un vecteur des caractéristiques physiographiques de la rivière.

5 CONCLUSION

Pour le groupe Services hydriques, il est indispensable de connaître avec un maximum de précision les données hydrométriques horaires (niveaux d'eau et débits des rivières) en temps réel. Ainsi, la détermination de la qualité d'une donnée lors de son acquisition permet de l'utiliser telle quelle ou de la reconstituer pour alimenter des applications nécessitant des données en temps réel. Cette pré-validation peut également faciliter le travail de validation puisque des données suspectes sont déjà identifiées.

Les contrôles du domaine de variation et de la cohérence temporelle (taux de variation et valeurs constantes) servent à effectuer une pré-validation efficace des données. Les seuils critiques pour ces deux contrôles ont été calculés de manière optimale compte tenu des données disponibles. Pour le contrôle du domaine de variation, les valeurs maximales et minimales de la série historique des données journalières servent de seuils critiques. Pour le contrôle de la cohérence temporelle, ce sont les quantiles empiriques ayant une probabilité au dépassement donnée (la valeur 1% est suggérée) qui sont utilisés comme seuils critiques. Dans ce dernier cas, la sévérité du seuil peut être modifiée en changeant simplement la probabilité au dépassement.

L'estimation des débits à l'embouchure des rivières par la méthode du rapport de superficies de bassins versants conduit à des résultats acceptables. Cette méthode est simple et valable. Les alternatives pour estimer le débit sont rares puisqu'il n'y a pas de données mesurées à l'embouchure des rivières pour calibrer des modèles plus complexes. Il est fortement suggéré d'inclure le temps de cheminement de l'eau dans la méthode.

La reconstitution des données manquantes ou éliminées par la procédure de pré-validation peut être effectuée par la méthode des voisins les plus proches (Yakowitz, 1987) ou par les modèles de Box-Jenkins (Box et Jenkins, 1976). Ces méthodes sont particulièrement bien adaptées pour des données très autocorrélées comme les débits horaires et les niveaux horaires.

Les problèmes de la validation et de l'estimation des débits sous couvert de glace ont été peu étudiés et ne trouvent pas de solution simple. L'équipe de la chaire en Hydrologie statistique en collaboration avec différents partenaires envisage d'effectuer des travaux de recherche dans ce domaine.

6 BIBLIOGRAPHIE

- Box, G.E.P et G.M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*. 2^e éd., Holden-Day, San-Francisco.
- Environnement Canada (1988). *Operational Specification for Acoustic Flow Meters*. Water Resources Branch, Inland Waters Directorate, Environment Canada, Ottawa, Ontario.
- Fortin, V., Ouarda, T.B.M.J., Rasmussen, P.F. et B. Bobée (1996). *Revue bibliographique des méthodes de prévisions des débits*. INRS-Eau, Rapport de recherche No R-477, 26 p.
- Haché M. et B. Larouche (1996). Validation des niveaux en temps réel. 49^e Conférence annuelle de l'Association canadienne des Ressources hydriques, Québec, 26-28 juin. Delisle, C.E. et M.A. Bouchard (Éds), Collection Environnement de l'Université de Montréal. Vol I (6 hors série): 71-80.
- Hamilton, A.S. et R.D. Moore (1996). Winter Streamflow variability in two groundwater-fed sub-Artic rivers, Yukon Territory, Canada. *Can. J. Civ. Eng.*, 23 : 1249-1259.
- Huttunen, M. et B. Vehviläinen (1997). Neural Networks in the Ice-Correction of Discharge Observations. *Nordic Hydrology*, 28(4): 283-296.
- HYDAT (1997). *Données sur les eaux de surface et les sédiments*. Version 4.95, Service de l'environnement atmosphérique, Environnement Canada.
- Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* (Transactions of the ASME, Series D), 82: 34-45.
- OMM (1985). *Guidelines for computerized data processing in operational hydrology and land and water management*. Organisation Météorologique Mondiale, Document technique WMO No. 364.
- OMM (1986). *Guidelines on the quality control of surface climatological data*. Préparé par P.F. Abbott (K.F.) comme rapporteur de la Commission Climatologique, Organisation Météorologique Mondiale, Document technique WCP-85.
- SEA (1991). *Regional quality control of climatological data - User's reference - Version 2.0*. Service de l'Environnement Atmosphérique, Document technique, 210 p.
- Teal, M.J., Ettama, R. et J.F. Walker (1994). Estimation of Mean Flow Velocity in Ice-Covered Channels. *Journal of Hydraulic Engineering ASCE*, 120(12): 1385-1400.
- Walker J.F. et D. Wang (1997). Measurement of Flows under Ice Covers in North America. *Journal of Hydraulic Engineering ASCE*, 123(11): 1037-1040.

Yakowitz, S. (1987). Nearest neighbor methods for time series analysis, *J. Time Ser. Anal.*, 8(2): 235-247.

ANNEXE A : Programmes *Matlab*

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Nom du fichier:      MINMAXJ.M
%
% Action:   Calcul du seuil critique pour la durée des épisodes de valeurs constantes
%           temporelle (taux de variation)
%
%           [min_j,max_j]=minmaxj(x)
%
%           où min_j,max_j Seuils pour le contrôle du domaine de variation (les valeurs
%           maximales (max_j) minimales (min_j) journalières;
%           x           Matrice contenant les débits journaliers (ou les niveaux journaliers) où
%           les lignes sont les années et les colonnes les jours (les données
%           manquantes sont spécifiées par la valeur -9999)
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [min_j,max_j]=minmaxj(x)

max_j = zeros(365,1);
min_j = zeros(365,1);

for i=1:59
    x_j = x(find( x(:,i)~-9999 ),i);
    max_j(i) = max(x_j);
    min_j(i) = min(x_j);
end;
for i=61:366
    x_j = x(find( x(:,i)~-9999 ),i);
    max_j(i-1) = max(x_j);
    min_j(i-1) = min(x_j);
end;
```

```

#####
%
% Nom du fichier:      TAUX_VARIATION.M
%
% Action:   Calcul du seuil inférieur et du seuil supérieur pour le controle de la cohérence
%           temporelle (taux de variation)
%
%           seuil=taux_variation(x, saison, pourc)
%
%           où  seuil   Seuils pour le controle de la cohérence temporelle (les quantiles
%                   correspondant au probabilité au non-dépassement spécifiée dans le vecteur
%                   pourc et aux saisons données dans la matrice saison) où les lignes
%                   correspondent au probabilité et les colonnes aux saisons (la première
%                   colonne correspond au seuils pour l'année complète);
%           x       Matrice contenant les debits horaires (ou les niveaux horaires) où
%                   les lignes sont les heures et les colonnes les années (les données
%                   manquantes sont spécifiées par la valeur -9999);
%           saison   Matrice contenant les heures définissant les saisons ou la première
%                   colonne marque le debut de la saison et la deuxième la fin de la saison;
%           pourc    Vecteur contenant la probabilité au non-dépassement exprimée en
%                   pourcentage (%).
%
#####

function seuil=taux_variation(x, saison, pourc)

% dimension de la matrice x;

[n_1 n_c]=size(x);

% calcul de la première difference

dx = x(2:n_1,:)-x(1:n_1-1,:);

% identification des données manquantes dans la matrice des différences dx

for i=1:n_c
    ix1 = find(x(:,i)==-9999);
    ix2 = ix1-1;
    ix1 = ix1(find(ix1==n_1));
    ix2 = ix2(find(ix2==0));
    dx(ix1,i) = -9999;
    dx(ix2,i) = -9999;
end;

% détermination du nombre de saisons et du nombre de seuils

n_pourc = length(pourc);
n_saison = size(saison,1);
seuil = zeros(n_pourc,n_saison+1);

% Calcul des quantiles pour l'année

dx_an = reshape(dx,1,(n_1-1)*n_c);
dx_an = dx_an(find(dx_an==-9999));
seuil(:,1) = prctile(dx_an,pourc)';

% Calcul des quantiles pour les saisons

for i=1:n_saison
    if saison(i,1)>saison(i,2)
        dx_s = [dx(saison(i,1):n_1-1,:);-9999*ones(1,n_c);dx(1:saison(i,2),:)];
        n_ls = size(dx_s,1);
        dx_ss = reshape(dx_s,1,n_ls*n_c);
        dx_ss = dx_ss(find(dx_ss==-9999));
        seuil(:,i+1) = prctile(dx_ss,pourc)';
    else
        dx_s = dx(saison(i,1):saison(i,2),:);
        n_ls = size(dx_s,1);
        dx_ss = reshape(dx_s,1,n_ls*n_c);
        dx_ss = dx_ss(find(dx_ss==-9999));
        seuil(:,i+1) = prctile(dx_ss,pourc)';
    end;
end;
end;

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Nom du fichier:      VALEUR_CONSTANTE.M
%
% Action:   Calcul du seuil inférieur et du seuil supérieur pour le contrôle de la cohérence
%           temporelle (taux de variation)
%
%           seuil = valeur_constante(x,prob);
%
%           où  seuil      Seuils pour le contrôle de la cohérence temporelle (les quantiles
%                       correspondant au probabilité au non-dépassement spécifiée dans le vecteur
%                       prob;
%           x          Matrice contenant les débits horaires (ou les niveaux horaires) où
%                       les lignes sont les heures et les colonnes les années (les données
%                       manquantes sont spécifiées par la valeur -9999);
%           prob       Vecteur contenant la probabilité au non-dépassement exprimée en
%                       pourcentage (%).
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function seuil = valeur_constante(x,prob);

% dimensions de la matrice x

[nl nc]=size(x);

% calcul des épisodes

x=[reshape(x,nl*nc,1); (x(nl,nc)+1)];
episode=zeros(100,1);
j=2;
while j<=nl*nc
    duree=1;
    while x(j)==x(j-1)
        duree=duree+1;
        j=j+1;
    end;
    if x(j-1) ~= -9999
        episode(duree)=episode(duree)+1;
    end;
    j=j+1;
end;

% Calcul des quantiles

frequence=100*cumsum(episode./sum(episode));

n_prob=length(prob);
seuil=zeros(n_prob,1);
for i=1:n_prob
    seuil(i)=max(find(frequence<prob(i)))+1;
end;

```