

Étude sur le choix du seuil de troncature en analyse des séries de durées partielles : application au Canada

Selection of Truncation Level for Partial Duration Series Analysis: Applications in Canada

D. Caissie and N. El-Jabi

Volume 5, Number 2, 1992

URI: <https://id.erudit.org/iderudit/705133ar>

DOI: <https://doi.org/10.7202/705133ar>

[See table of contents](#)

Publisher(s)

Université du Québec - INRS-Eau, Terre et Environnement (INRS-ETE)

ISSN

0992-7158 (print)

1718-8598 (digital)

[Explore this journal](#)

Cite this article

Caissie, D. & El-Jabi, N. (1992). Étude sur le choix du seuil de troncature en analyse des séries de durées partielles : application au Canada. *Revue des sciences de l'eau / Journal of Water Science*, 5(2), 291–307. <https://doi.org/10.7202/705133ar>

Article abstract

Accurate forecasting of flood flows is required for the efficient design and construction of hydraulic structures in rivers as well as for the effective management of water resources. Underestimation of flood flows can result in tragic consequences while overdesigned structures are expensive.

One method for estimating flood flows is the partial duration series analysis. In this approach a truncation level defining the intermittent time series is chosen. All flows above this level (exceedances) are analyzed by assuming the time of occurrence of these floods to represent a Poisson distribution. In addition, exceedances are considered to be independent random variables identically distributed over a one-year time interval. The selection of the truncation level is somewhat problematic and not very well defined in the literature. This study presents a truncation level estimation technique based on a series of regional regression equations for 9 distinct regions in Canada, and documents their graphical derivation.

According to previous research, the truncation level can be obtained in two ways. First, it is selected in accordance with physical criteria such as the overflowing of a river, the critical flow for flooding of a crop, etc. The second method of truncation level selection is primarily mathematical such as satisfying the analytical fit or the assumptions of the model. In the present study, the truncation level was selected based on the mathematical approach using a graphical technique and using the Chi-square test. The graphical truncation level selection technique is based on the equality of the mean and variance of the Poisson distribution. Given this property of the Poisson distribution, one can study the mean-to-variance ratio as a function of truncation level.

Truncation levels were selected graphically from 238 gauging stations across Canada. The Chi-square at a level of significance of 5 % was used for their validation. It was observed, upon examination of the mean-to-variance ratio, that selection of the truncation level was easier in the eastern and western regions of Canada where the ratio varied very little around unity. In the prairies and northern regions, a small variation in the truncation level led to a large variation in the ratio, making the selection of a level more difficult.

Following the selection of a truncation level for each hydrometric station, homogeneous regions were selected based on previous countrywide hydrological studies. The regression analysis was then carried out to explain this truncation level using several variables (including both physiographic variables and streamflow characteristics). Physiographic variables included area of drainage basin (km^2); area of lakes and swamps (km^2); area of forests (km^2); mean elevation (m); slope of drainage basin (%); slope of principal watercourse (m/km, %); length of principal water course (km); area controlled by lakes and swamps (km^2); and drainage density (km/km^2). The Streamflow characteristics included the mean annual flow and mean annual flood. The mean annual flood was included because previous research has shown a strong correlation between the truncation level and the two-year flood estimated by a log-Pearson type III distribution function. For ease of application, the mean annual flood was used as a low return flood estimate as it represents the two-year flood calculated by a normal distribution function. Low return floods of any given distribution function are demonstrated to be similar in magnitude.

Multiple regression was carried out using original data and logarithmically transformed data. Equations were derived by selecting one variable, two variables, and so on until all explanatory variables were accounted for in a single equation. Among the selected parameters, the mean annual flood was the best parameter to explain the truncation level for all of the regions across Canada. The coefficient of determination, R^2 , of the equations calculated using mean annual flood varied between 0.847 and 0.987 from original data, and varied between 0.824 and 0.988 for logarithmically transformed data. In practice, the regional equation can eliminate the graphical estimation technique and a first truncation level can be obtained. This truncation level can be used in the partial duration series analysis with the different tests involved. If this first estimated level does not meet all the statistical tests, a second usually higher level is selected. It is important to note that when the selected level meets the statistical test requirements, the estimation of QT should be relatively insensitive to levels near the selected truncation level.

Étude sur le choix du seuil de troncature en analyse des séries de durées partielles : application au Canada

Selection of Truncation Level for Partial Duration Series Analysis : Applications in Canada

D. CAISSIE¹, N. EL-JABI²

Reçu le 18 janvier 1991, accepté pour publication le 3 avril 1992*.

SUMMARY

Accurate forecasting of flood flows is required for the efficient design and construction of hydraulic structures in rivers as well as for the effective management of water resources. Underestimation of flood flows can result in tragic consequences while oversized structures are expensive.

One method for estimating flood flows is the partial duration series analysis. In this approach a truncation level defining the intermittent time series is chosen. All flows above this level (exceedances) are analyzed by assuming the time of occurrence of these floods to represent a Poisson distribution. In addition, exceedances are considered to be independent random variables identically distributed over a one-year time interval. The selection of the truncation level is somewhat problematic and not very well defined in the literature. This study presents a truncation level estimation technique based on a series of regional regression equations for 9 distinct regions in Canada, and documents their graphical derivation.

According to previous research, the truncation level can be obtained in two ways. First, it is selected in accordance with physical criteria such as the overflowing of a river, the critical flow for flooding of a crop, etc. The second method of truncation level selection is primarily mathematical such as satisfying the analytical fit or the assumptions of the model. In the present study, the truncation level was selected based on the mathematical approach using a graphical technique and using the Chi-square test. The graphical truncation level selection technique is based on the equality of the mean and variance of the Poisson distribution. Given this property of the Poisson distribution, one can study the mean-to-variance ratio as a function of truncation level.

Truncation levels were selected graphically from 238 gauging stations across Canada. The Chi-square at a level of significance of 5 % was used for their

-
1. Direction des sciences, Pêches et Océans Canada, C.P. 5030, Moncton, Nouveau-Brunswick, E1C 9 B6, (506) 851-6287.
 2. École de génie, Université de Moncton, Moncton, Nouveau-Brunswick, E1A 3E9, (506) 858-4296.

* Les commentaires seront reçus jusqu'au 30 décembre 1992.

validation. It was observed, upon examination of the mean-to-variance ratio, that selection of the truncation level was easier in the eastern and western regions of Canada where the ratio varied very little around unity. In the prairies and northern regions, a small variation in the truncation level led to a large variation in the ratio, making the selection of a level more difficult.

Following the selection of a truncation level for each hydrometric station, homogeneous regions were selected based on previous countrywide hydrological studies. The regression analysis was then carried out to explain this truncation level using several variables (including both physiographic variables and streamflow characteristics). Physiographic variables included area of drainage basin (km^2); area of lakes and swamps (km^2); area of forests (km^2); mean elevation (m); slope of drainage basin (%); slope of principal watercourse (m/km , %); length of principal watercourse (km); area controlled by lakes and swamps (km^2); and drainage density (km/km^2). The Streamflow characteristics included the mean annual flow and mean annual flood. The mean annual flood was included because previous research has shown a strong correlation between the truncation level and the two-year flood estimated by a log-Pearson type III distribution function. For ease of application, the mean annual flood was used as a low return flood estimate as it represents the two-year flood calculated by a normal distribution function. Low return floods of any given distribution function are demonstrated to be similar in magnitude.

Multiple regression was carried out using original data and logarithmically transformed data. Equations were derived by selecting one variable, two variables, and so on until all explanatory variables were accounted for in a single equation. Among the selected parameters, the mean annual flood was the best parameter to explain the truncation level for all of the regions across Canada. The coefficient of determination, R^2 , of the equations calculated using mean annual flood varied between 0.847 and 0.987 from original data, and varied between 0.824 and 0.988 for logarithmically transformed data. In practice, the regional equation can eliminate the graphical estimation technique and a first truncation level can be obtained. This truncation level can be used in the partial duration series analysis with the different tests involved. If this first estimated level does not meet all the statistical tests, a second usually higher level is selected. It is important to note that when the selected level meets the statistical test requirements, the estimation of QT should be relatively insensitive to levels near the selected truncation level.

Key-words : *flood, partial duration series, truncation level.*

RÉSUMÉ

L'estimation des débits de crues est d'une importance majeure pour la conception des ouvrages d'art et la gestion des ressources hydriques. La mauvaise évaluation de ces débits entraîne un surdimensionnement ou sous dimensionnement des ouvrages hydrauliques, induisant ainsi un investissement excessif ou un risque démesuré d'inondations. Un modèle de données chronologiques de durée partielle est retenu pour l'étude des crues. Ce modèle nécessite d'abord l'estimation du seuil de troncature qui définit la série chronologique intermittente à analyser. Ce paramètre joue un rôle majeur dans la solution du problème malgré le peu d'importance que lui reconnaît la littérature. Le but de la présente étude est d'élaborer une méthode graphique comme guide dans le choix du seuil. Par la suite, cette méthode est appliquée à 238 stations hydrométriques au Canada. Une fois le seuil obtenu pour ces stations hydrométriques, une régionalisation utilisant l'analyse de régression a été réalisée. Ces équations régionales peuvent servir à estimer le seuil d'une nouvelle série hydrologique sans avoir à passer par la méthode graphique.

Mots clés : *crue, série à durée partielle, seuil de troncature.*

INTRODUCTION

Condition essentielle à la conception des ouvrages hydrauliques et la gestion des ressources hydriques, l'estimation des débits de crues correspondant à une période de récurrence quelconque est un des principaux objets des études hydrologiques. Ainsi, l'hydrologue est amené à modéliser le processus de crue, phénomène complexe, variable dans le temps et dans l'espace, et possédant des paramètres globaux et/ou discrétisés. Pour ce faire, l'hydrologue fait appel à des modèles divers, dont le modèle de dépassement ou le modèle des séries chronologiques de durées partielles, ce dernier étant particulièrement adapté à ce genre d'études. Ce modèle comporte toutefois une difficulté majeure lors de son application, qui est de déterminer le seuil de troncature. Ce dernier est défini comme le débit minimum de crue. D'un point de vue pratique, il faut avoir une méthodologie simple et efficace pour choisir ce débit. Le but principal de ce travail est de déterminer d'abord d'une telle méthode, puis d'effectuer une étude régionale, en utilisant l'analyse de régression, pour estimer le seuil.

Même si le choix du seuil de dépassement est très important dans l'analyse des séries partielles, il n'existe que très peu d'informations permettant de l'identifier dans la littérature. Plusieurs chercheurs présentent le seuil comme le débit correspondant à une certaine période de récurrence. DALRYMPLE (1960) suggère un seuil correspondant au débit d'une période de récurrence de 1,15 années environ. WAYLEN et WOO (1983a ; 1983b) choisit le seuil comme celui qui correspond à une période de récurrence de 1,4 ou 2,0 années selon de la rivière étudiée. D'autre part, IRVINE et WAYLEN (1986) utilisent des périodes de récurrence de 1,2 et 1,8 années pour chaque rivière. Une autre approche consiste à fixer ce débit afin d'avoir un certain nombre de crues en moyenne par année. TAESOMBUT et YEVJEVICH (1978) utilisent un nombre moyen de crues annuelles, situé entre un et quatre environ afin d'étudier différentes caractéristiques des crues en fonction du seuil. Cette approche fut également appliquée par CORREIA (1985 ; 1987) qui étudia plusieurs stations hydrométriques en choisissant le seuil afin d'avoir une moyenne de une, deux et quatre crues par année. En imposant trois seuils différents, le nombre de calculs est plus grand, mais cette approche a permis d'étudier la dépendance des crues en fonction du seuil. KONECNY et NACHTNEBEL (1985) mentionnent dans leur étude qu'il faut choisir le seuil afin d'avoir au moins une crue en moyenne par année. LAURENSEN (1987) présente le choix du seuil comme le débit minimum avec un intérêt pratique.

D'après ASHKAR et ROUSSELLE (1983), le seuil peut être obtenu selon deux méthodes. La première consiste à le choisir selon des critères physiques tels que le débordement d'une rivière ou le débit critique d'inondation. La deuxième méthode consiste à choisir le seuil selon des critères purement mathématiques pour satisfaire les hypothèses d'un modèle. ASHKAR et ROUSSELLE (1983) mentionnent aussi que le choix du seuil dépend souvent du jugement de l'hydrologue et que deux personnes peuvent obtenir des résultats complètement différents.

Afin d'établir une méthodologie efficace dans le choix du seuil, EL-JABI *et al.* (1986) ont effectué une étude détaillée de 34 stations hydrométriques du Québec. Ils suggèrent une estimation du seuil selon le critère du rapport entre la moyenne et la variance qui est égale à l'unité dans le cas des processus de Poisson. Afin d'établir des équations régionales expliquant le seuil, EL-JABI *et al.* (1986) ont utilisé une approche par régression. Parmi les variables explicatives étudiées, ils considèrent le débit moyen annuel, le débit de deux ans pour une loi de distribution répartie selon log-Pearson type III ajustée aux débits maximum annuels, et la superficie drainée. EL-JABI *et al.* (1986) remarquent que le débit de deux ans de la distribution log-Pearson type III était le plus corrélé avec le seuil (i.e. 0,93). La deuxième variable la plus corrélée était la superficie drainée, avec un coefficient de corrélation de 0,83.

L'objectif de cet article est double : i) définir une procédure claire de choix du seuil pour une série donnée et ii) rechercher, à partir d'un grand nombre de séries traitées, une relation entre des variables explicatives simples et le seuil choisi, de façon à faciliter le choix du seuil pour l'étude de nouvelles séries.

CONSIDÉRATIONS THÉORIQUES

Modèle des séries à durées partielles

Le modèle des séries à durées partielles fut introduit par LANGBEIN (1949) qui compara ce modèle avec le modèle des séries annuelles. Par la suite, BORGMAN (1963), SHANE et LYNN (1964) et BERNIER (1967), appliquèrent cette nouvelle approche en analyse de la fréquence des crues. Ce modèle constitua alors une référence théorique pour l'analyse des crues (TODOROVIC, 1970 ; TODOROVIC et ZELENHASIC, 1970 ; et ZELENHASIC, 1970) complémentaire aux travaux de GUMBEL (1958) sur ce thème.

Le modèle de dépassement est basé sur la théorie des valeurs extrêmes et celle sur les nombres aléatoires des variables aléatoires (TODOROVIC, 1970). Considérant l'hydrogramme représentant le débit d'une rivière pour un intervalle de temps (T_{k-1}, T_k) (fig. 1a), le dépassement ξ_v (débit dépassant un seuil de troncature Q_B) réalisé au temps $\tau(v)$ dans l'intervalle de temps est défini comme suit :

$$\xi = Q_v - Q_B \quad (1)$$

Alors le nombre des dépassements ξ_v dans une intervalle de temps (T_{k-1}, T_k) , le temps d'occurrence de ces dépassements ainsi que leur valeurs sont des variables aléatoires (fig. 1b).

Il a été prouvé (TODOROVIC, 1970) que les dépassements maximaux suivent un processus stochastique non décroissant. Si de plus, les dépassements de crues sont supposés indépendants et identiquement distribués dans

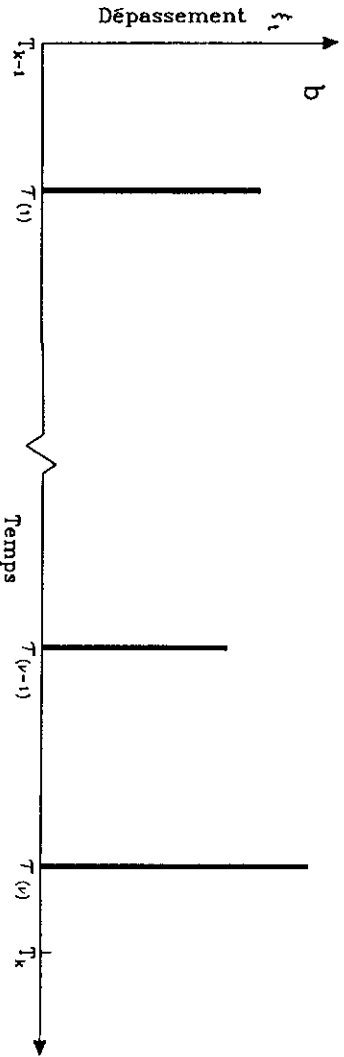
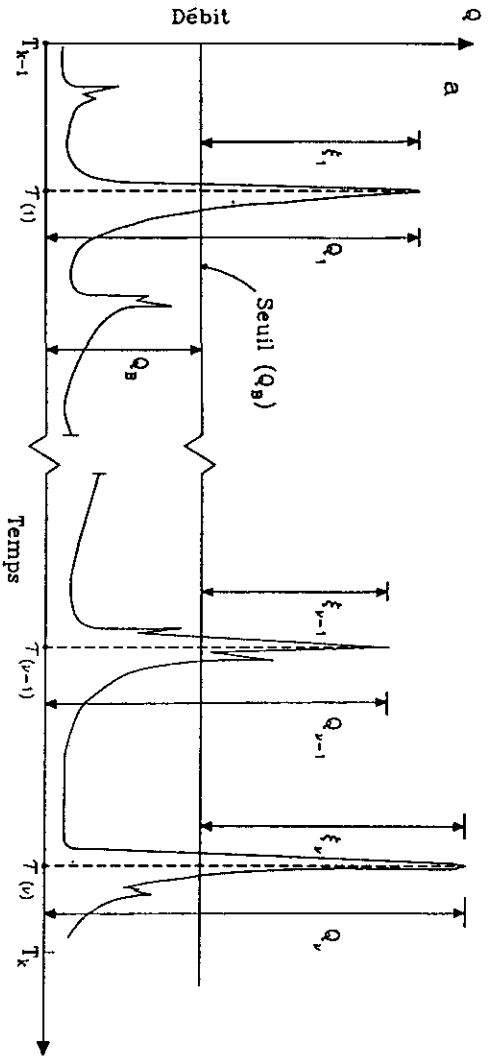


Figure 1

Hydrogramme du débit avec l'identification des dépassements de crue (a) et processus stochastique en résultant (b).
Streamflow hydrograph identifying flood exceedances (a) and resulting stochastic process (b).

un intervalle de temps (T_{k-1}, T_k) , alors la fonction de répartition du dépassement maximum $\chi(t)$ peut s'écrire :

$$F_t(x) = P[\chi(t) \leq x] \quad (2)$$

$$\text{où} \quad \chi(t) = \text{Sup}_{\tau(v) \leq t} \xi_v \quad (3)$$

Todorovic and Rousselle (1971) présentèrent la fonction de répartition du dépassement maximal comme,

$$F_t(X) = \exp \left\{ - \sum_{n=1}^{k-1} [\Lambda(T_n) - \Lambda(T_{n-1})] [1 - H_n(X)] - [\Lambda(t) - \Lambda(T_{k-1})] [1 - H_k(X)] \right\} \quad (4)$$

pour tout $k = 1, 2, \dots$ et $t \in (T_{k-1}, T_k)$, où $\Lambda(t)$ est le nombre moyen des dépassements $\xi \in (0, t)$, et $H(x)$ est la fonction de répartition de la valeur des dépassements. On peut donc écrire,

$$H(x) = P(\xi \leq x) \quad (5)$$

La fonction $H(x)$ peut être représentée par n'importe quelle loi statistique, selon les caractéristiques hydrométriques de la rivière. ROUSSELLE (1972) et EL-JABI *et al.* (1986) démontrèrent que la fonction de répartition exponentielle s'applique d'une façon satisfaisante pour décrire la distribution des dépassements de crues. Alors,

$$H(x) = 1 - e^{-\beta x} \quad (6)$$

où $\beta = \{E(\xi)\}^{-1} \geq 0$; et E représente la notation l'espérance mathématique.

Méthode graphique pour le choix du seuil

En analyse des séries de durées partielles, le nombre des dépassements de crues (i.e. la fréquence des événements dont le débit est supérieur au seuil) suit souvent un processus de Poisson (SHANE et LYNN, 1964). Une caractéristique importante de la loi de Poisson est l'égalité entre la moyenne et la variance. Afin de déceler la possibilité d'un processus de Poisson pour différents seuils, on effectue une étude graphique du rapport entre la moyenne et la variance (rapport MV) du nombre des dépassements de crues, et ceci pour différents seuils. Même si ce rapport est près de l'unité, on ne peut conclure que le processus est du type Poissonien, car l'égalité entre la moyenne et la variance est une condition nécessaire, mais non suffisante. Alors, on utilise ce rapport pour éliminer toute série de débit dont le rapport est éloigné de l'unité. De plus, on choisit le seuil le plus faible de cette série afin de maximiser les informations. Finalement, on teste ce seuil par un test statistique (khi-carré par exemple) à un seuil de signification choisi. Il est à remarquer que cette procédure graphique a été utilisée dans la présente étude comme outil supplémentaire pour faire le choix de notre série de seuil. Des tests statistiques ont été utilisés pour vérifier l'hypothèse de la loi de Poisson ainsi que les autres hypothèses du modèle (c'est-à-dire l'hypothèse sur les fonctions de répartition $H(x)$ et $F_t(x)$).

Analyse par régression multiple

L'analyse de régression multiple est utilisée afin d'obtenir une estimation du seuil sans passer nécessairement par la méthodologie décrite à la section précédente. Pour ce faire, des équations de régression multiple du seuil ont été obtenues à l'aide des programmes tirés de la bibliothèque de logiciel BMDP (1981). Ces équations de régression multiple permettront par la suite d'obtenir une estimation des seuils de façon économique et efficace. Les résultats seront présentés et discutés ultérieurement.

APPLICATION

L'étude des seuils a été effectuée sur 238 stations hydrométriques analysées par la méthode d'analyse des séries à durées partielles. Ces stations hydrométriques provenaient de la banque de données HYDROL (BOUCHER, 1986).

Description des données

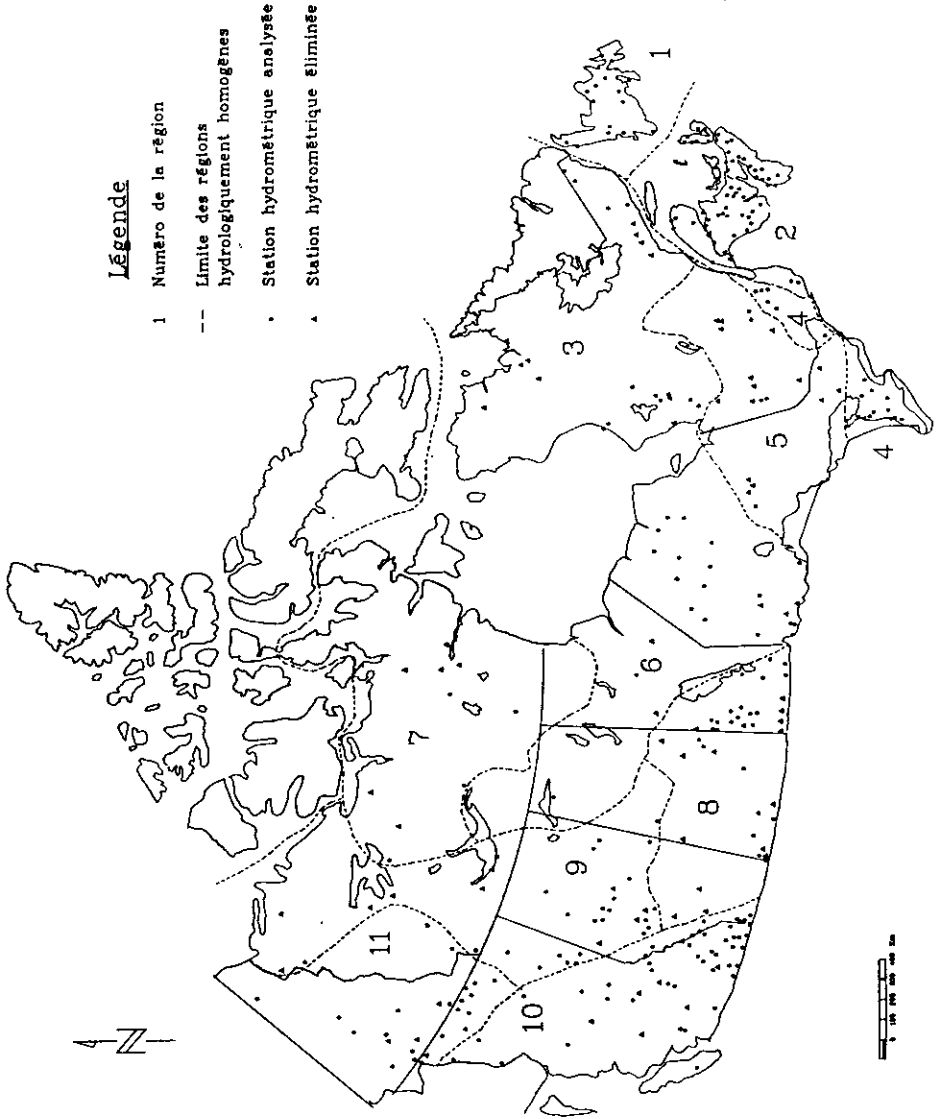
La banque de données HYDROL est composée de 495 stations hydrométriques au Canada ; chacune des stations est identifiée par la cote A, B ou C selon la durée des données de la station. Les cotes A, B et C sont définies comme suit :

- A : station ayant une période d'enregistrement supérieure à 30 ans.
- B : station ayant une période d'enregistrement allant de 20 à 29 ans.
- C : station ayant une période d'enregistrement allant de 11 à 19 ans.

Vu le nombre très grand des stations à analyser, une première élimination a été effectuée en se basant sur une distribution géographique plus équilibrée à l'intérieur de chaque province étudiée. Dans le processus d'élimination, une préférence a été attribuée aux stations de cote A, et après avoir fait une première élimination, le nombre de stations a été réduit à 325 (fig. 2). À l'aide de ces stations, le modèle de dépassement fut appliqué sur chaque station hydrométrique.

Choix du seuil

Le rapport MV du nombre moyen des dépassements de crues par année sur la variance de l'échantillon est considéré pour différents seuils. Ce rapport est alors tracé en fonction du nombre moyen des dépassements par année pour chaque station hydrométrique à partir de la banque de données HYDROL. La droite horizontale (c'est-à-dire la droite d'équation $y = 1$) représente la condition nécessaire (mais pas suffisante) à l'existence d'un processus de Poisson. La relation entre la courbe calculée et la droite horizontale permet de déterminer une région où les seuils peuvent engendrer un processus de Poisson. La figure 3 illustre l'étude du seuil pour quatre rivières au Canada.



L'étude du rapport MV nous permet de choisir une série de seuils qui semblent suivre un processus de Poisson. Parmi cette série, on choisit un débit (du plus faible au plus fort), pour maximiser les informations (ASHKAR et ROUELLE 1987). Lorsqu'un débit est choisi, ce dernier est utilisé pour effectuer une analyse stochastique des crues. Les tests statistiques impliqués dans l'analyse des séries à durées partielles sont le test du khi-carré pour la distribution de Poisson et le test de Kolmogorov-Smirnov pour la fonction de répartition des dépassements et du dépassement maximal. Lorsqu'un des tests ci-dessus n'est pas satisfait, on passe à un deuxième seuil supérieur au premier. On répète la procédure jusqu'à ce que l'on trouve un débit qui satisfasse tous les tests, ou jusqu'à ce que la série de débits soit épuisée. Certaines stations furent éliminées lorsque l'un des trois tests n'était pas satisfait et ceci porta le nombre de station de 325 à 238.

Il a été remarqué qu'il est plus facile de faire le choix du seuil pour les régions de l'est et de l'ouest canadien. Pour les régions des prairies et nordiques, le contraire fut remarqué, car une petite variation autour du seuil engendra une grande variation du rapport moyenne sur variance, ce qui indique que le processus n'est plus Poissonien. Certaines rivières ne semblaient pas répondre à la loi de Poisson à cause d'une régularisation artificielle. Cette régularisation a été observée lorsque la station hydrométrique était située à l'embouchure d'un lac, car les stations influencées par les opérations d'un barrage ne faisaient pas partie de la banque de donnée HYDROL.

Régions hydrologiques

Pour établir des équations de régression expliquant le seuil, les 238 stations hydrométriques ont été distribuées en plusieurs régions considérées comme hydrologiquement homogènes (fig. 2).

Le choix des régions homogènes est basé sur l'étude de JUDGE *et al.* (1988) dans laquelle le Canada était subdivisé en 12 régions. La région 12 de cette étude a été éliminée, car aucune information n'existe dans cette partie du Canada. A cause d'un manque de données, les régions 4 et 5 ont été combinées ensemble. De plus, les quelques stations de la région 7 ne font pas partie de l'analyse, car le nombre de quatre stations hydrométriques n'a pas été jugé suffisant à l'établissement d'équations régionales. Finalement, le nombre de régions à analyser a été fixé à 9 (i.e. 1, 2, 3, 4 + 5, 6, 8, 9, 10 et 11 ; fig. 2).

Variables explicatives disponibles

Afin de choisir une équation de régression qui explique bien le seuil (QB), plusieurs variables explicatives ont été étudiées. Ces variables peuvent être classées en deux catégories, soit les variables physiographiques et les variables hydrométriques. Les variables hydrométriques sont : le débit moyen (QM) et le débit maximal annuel moyen (QMAM). Ce dernier est obtenu par la moyenne des débits extrêmes annuels. Le débit moyen est couramment utilisé dans les études hydrologiques, mais le débit maximal annuel moyen présente un intérêt particulier dans cette étude.

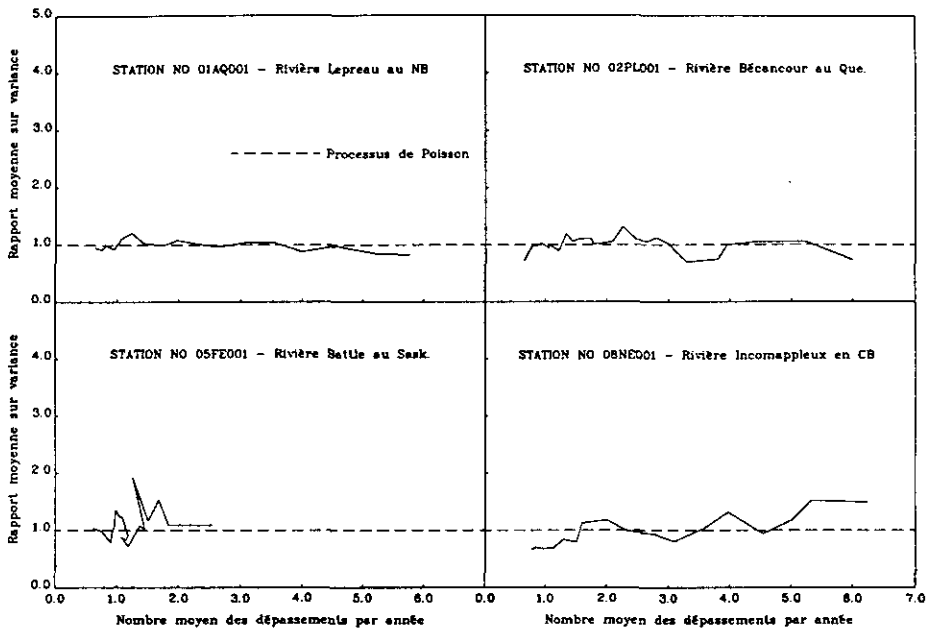


Figure 3 Étude du rapport moyenne sur variance pour choisir le seuil de troncature.

Mean to variance ratio analysis for selecting the truncation level.

Le choix du débit maximal annuel moyen comme paramètre explicatif du seuil n'est pas un choix arbitraire. En effet, EL-JABI *et al.* (1986) remarquèrent que le débit d'une récurrence de deux ans du modèle log-Pearson type III était corrélé avec le seuil aux environs de 0,93 à lui seul. A l'aide de la figure 4, on peut généralement observer que peu importe le modèle statistique, les débits de faible récurrence sont pratiquement égaux. Au lieu d'étudier le débit de deux ans du modèle log-Pearson type III, le débit de deux ans issu de la loi normale, bien plus simple, est utilisé. Le débit de deux ans de la distribution normale est égal à la moyenne des débits extrêmes annuels de la période d'observation, c'est-à-dire le débit maximal annuel moyen (QMAM) (KITE, 1978).

Les autres variables étudiées sont les caractéristiques physiographiques et ces dernières représentent l'aspect physique du bassin versant. Tous les paramètres physiographiques utilisés dans ce travail font parties de la banque de données HYDROL.

Les variables physiographiques sont :

A : superficie du bassin versant (km²)

SLM : superficie des lacs et marais (km²)

SF : superficie de forêts (km²)

ELM : élévation moyenne (m)

PB : pente du bassin versant (m/m, %)

PCP : pente du cours d'eau principal (m/km, %)

LCP : longueur du cours d'eau principal (km)
 SCLM : superficie contrôlée par les lacs et marais (km²)
 DD : densité de drainage (km/km²)

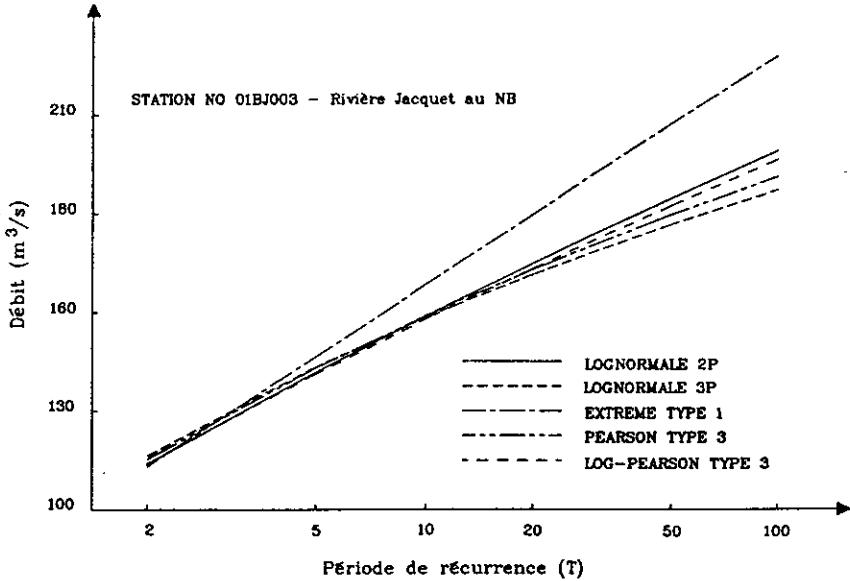


Figure 4 Différence entre les débits de récurrence 2 ans par différents modèles statistiques.
Differences in the 2-year flood estimates by different statistical distributions.

Régressions multiples

Parmi les modèles à plusieurs variables explicatives (modèles multivariés), le modèle linéaire multiple et le modèle linéaire multiple après transformation logarithmique des variables ont été étudiés en utilisant le seuil comme variable dépendante (BMDP, 1981). La matrice de corrélation, le coefficient d'explication multiple R² et l'erreur quadratique moyenne permettent par la suite de faire le choix de l'équation ou des équations représentant le mieux le seuil. Ce choix est fait en tenant compte des critères suivants :

- 1) la ou les variables choisies devraient être faciles à obtenir.
- 2) chaque équation devrait dépendre d'un minimum de variables indépendantes.

Par exemple, les résultats de dépendance entre les variables pour la région d'étude n° 2 sont présentés dans la matrice de corrélation de la figure 5. La région d'étude n° 2 comprend le Nouveau-Brunswick, la Nouvelle-Ecosse, et la Gaspésie (Québec). A partir de la figure 5, on remarque que le

	A	SLM	PCP	QM	QMAM	QB
A	1.000					
SLM	0.353	1.000				
PCP	0.031	-0.115	1.000			
QM	0.988	0.404	-0.015	1.000		
QMAM	0.978	0.280	0.087	0.976	1.000	
QB	0.904	0.354	0.058	0.923	0.944	1.000

Figure 5 Matrice de corrélation pour les variables utilisées dans la régression multiple de la région 2.

Variable correlation matrix used in multiple regression analysis of region 2.

Tableau 1 Équations de régression pour différentes régions au Canada.

Table 1 Regression equations for different regions of Canada

Région	Équation	R ²	Nombre de station	Erreur quadratique moyenne
1	QB = 0,535 QMAM ^{1,008}	0,966	15	20,1
	QB = 0,579 QMAM	0,962		19,7
2	QB = 0,617 QMAM ^{0,986}	0,981	54	14,9
	QB = 0,529 QMAM	0,968		13,5
3	QB = 0,630 QMAM ^{0,987}	0,949	12	145
	QB = 0,597 QMAM	0,868		142
4 + 5	QB = 0,356 QMAM ^{1,082}	0,961	20	17,6
	QB = 0,578 QMAM	0,955		18,1
6	QB = 0,422 QMAM ^{1,011}	0,836	17	81,5
	QB = 0,563 QMAM	0,964		52,1
8	QB = 0,176 QMAM ^{1,117}	0,824	25	6,04
	QB = 0,324 QMAM	0,847		6,03
9	QB = 0,169 QMAM ^{1,170}	0,974	22	83,8
	QB = 0,630 QMAM	0,971		99,7
10	QB = 0,542 QMAM ^{1,036}	0,988	42	67,0
	QB = 0,743 QMAM	0,987		61,4
11	QB = 0,722 QMAM ^{0,969}	0,966	23	300
	QB = 0,569 QMAM	0,867		299

coefficient de corrélation entre le paramètre QMAM et le seuil est de 0,94. La variable de seconde importance est le débit moyen QM dont le coefficient de corrélation est de 0,92. La superficie du bassin est également un paramètre important avec un coefficient de corrélation de 0,90. Pour toutes les régions au

Canada, le paramètre QMAM était parmi le paramètre le plus fortement corrélé. Le tableau 1 représente le résultat de cette étude pour le Canada. La figure 6 illustre les équations de la région hydrologique 2. On remarque que le tableau 1 présente les équations linéaires et non linéaires. Dans le cas des équations linéaires, nous avons fait le choix de l'équation qui passe par l'origine pour éviter la possibilité d'obtenir des seuils négatifs. Ce tableau présente les résultats des 231 stations hydrométriques seulement, quatre stations ayant été éliminées à cause de l'incompatibilité des paramètres physiographiques. Le choix d'un modèle à plusieurs variables explicatives aurait été possible, mais seulement le débit maximal annuel moyen (QMAM) fut choisi afin de conserver des équations simples et le choix d'un paramètre facile à calculer sans l'utilisation de cartes topographiques.

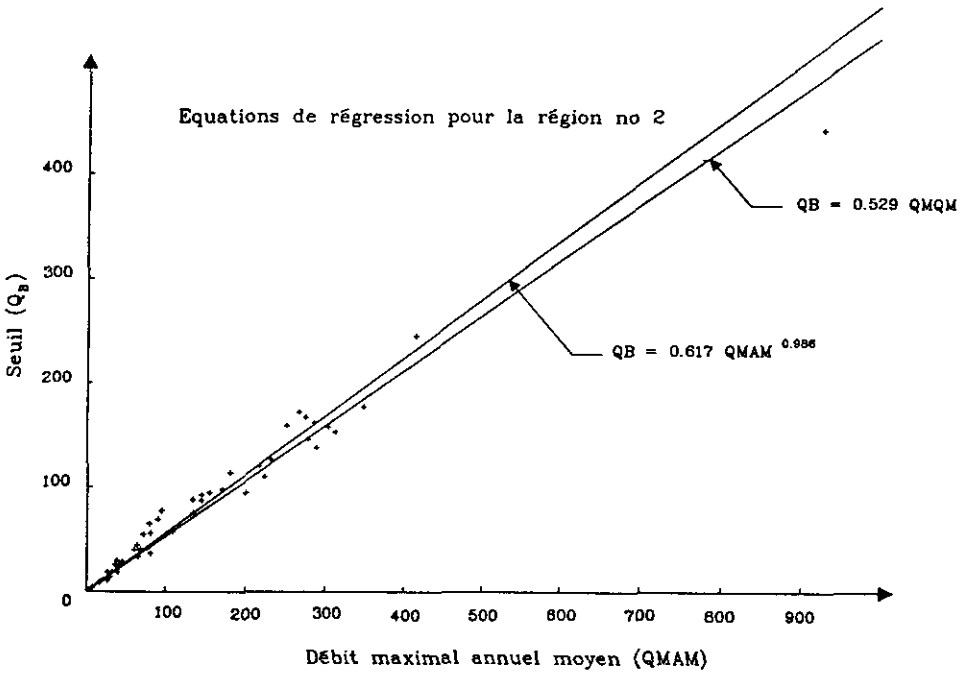


Figure 6 Equations de régression entre le seuil et le débit maximal annuel moyen (QMAM) de la région 2.

Regression equations between truncation level and mean annual flood of region 2.

Les figures 7a et 7b illustrent la relation entre le seuil utilisé pour l'analyse stochastique des crues (que l'on réfère comme QB exacte), et le seuil calculé à l'aide des équations de régression. La figure 7a présente les résultats selon les équations linéaires tandis que la figure 7b présente les résultats des équations obtenues après une transformation logarithmique.

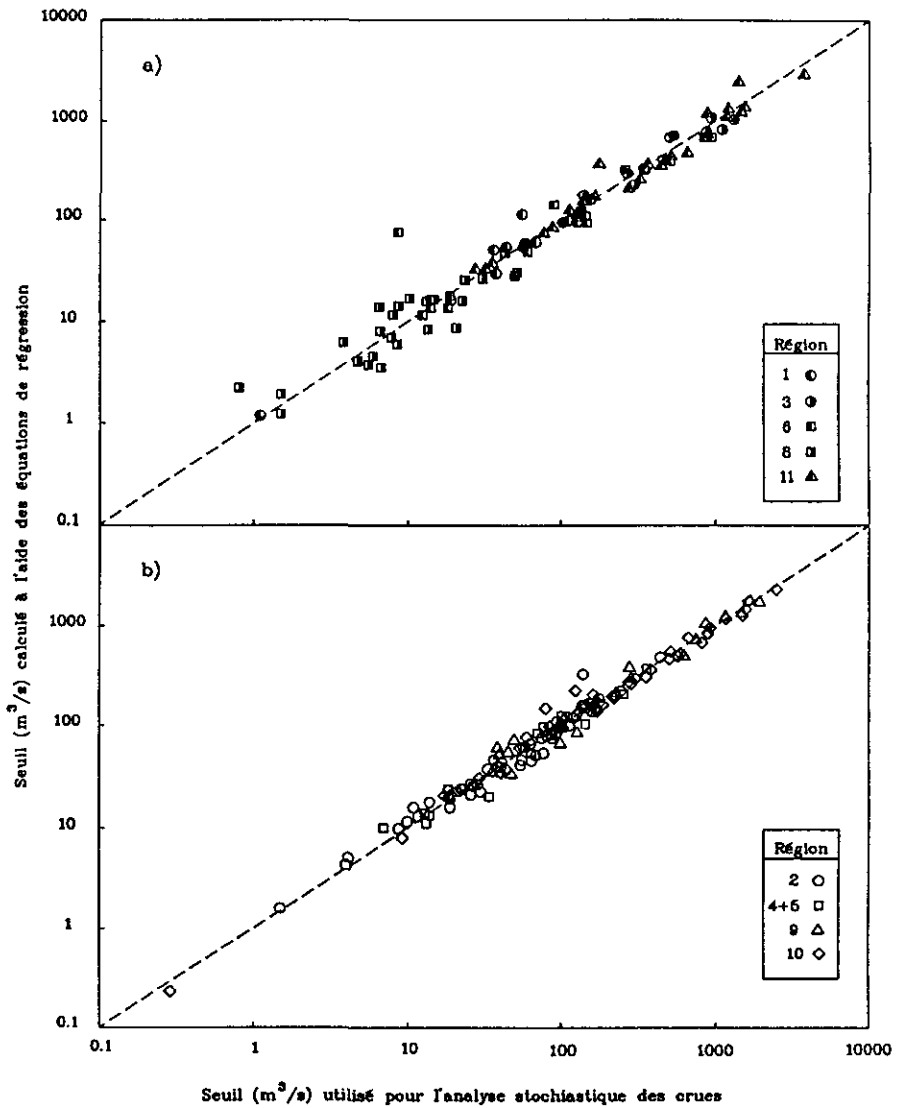


Figure 7 Seuil estimé graphiquement et utilisé pour l'analyse des séries à durée partielle en fonction du seuil calculé à l'aide des équations de régression régionales (la droite représente l'égalité des approches).
Graphically estimated truncation level used in the partial duration series analysis vs truncation level calculated using regional regression equations (line represents the equivalency of approaches).

CONCLUSIONS

L'établissement des équations régionales pour estimer le seuil fait partie des objectifs de cette étude. Mais, avant d'arriver à ces résultats, nous avons dû estimer le seuil de chaque station hydrométrique. Ceci a été possible grâce à la méthode développée par EL-JABI *et al.* (1986) qui consiste à étudier une propriété importante de la loi de Poisson (i.e. moyenne = variance). Après ce choix préliminaire d'un seuil assez faible pour maximiser les informations de dépassements de crues, le modèle stochastique des crues fut appliqué pour vérifier l'ajustement de la loi de Poisson et des autres fonctions de répartition impliquées (à un seuil de signification de 5 %). Lorsque le choix préliminaire ne satisfait pas tous les tests impliqués, un deuxième seuil, supérieur au premier, est choisi, jusqu'à ce que les tests soient satisfaits ou que la série de débits soit épuisée.

Il a été observé par l'étude du rapport MV que le choix du seuil était assez facile pour les régions est et ouest du Canada. Dans ce cas, le rapport MV variait très peu autour de la droite d'équation $y = 1$ (droite représentant le processus de Poisson). Par contre, dans les régions des prairies et nordiques, une petite variation du seuil engendra une grande variation du rapport, ce qui rendit le choix du seuil plus difficile. Sans même effectuer une analyse de sensibilité, on peut conclure que les équations de régression s'appliqueront avec une plus grande facilité dans les régions est et ouest du Canada que dans les régions des prairies et nordiques.

La dernière étape de ce travail était l'établissement d'équations afin de faciliter le calcul du seuil. Le paramètre retenu des analyses est le débit maximal annuel moyen (QMAM). De plus, le coefficient d'explication multiple (R^2) des équations linéaires varie entre 0,85 et 0,99 tandis que ce même coefficient varie entre 0,88 et 0,99 pour les équations transformées. Une étude sur l'erreur quadratique moyenne est également présentée au tableau 1 pour comparer les résultats des équations linéaire avec ceux des résultats des équations non linéaires. Ce paramètre varie d'environ 6 pour la région 8 jusqu'à 300 pour la région 11. En comparant les deux types d'équations utilisées, on remarque que les résultats sont très proches, sauf pour les régions 6 et 8, où la différence est plus grande.

En pratique, les équations de régression éliminent la procédure graphique de l'étude du rapport MV, et une première estimation du débit (seuil) est obtenue. Par la suite, l'hydrologue utilise ce seuil pour effectuer une analyse des crues par séries partielles en testant les différentes hypothèses du modèle. Parfois, cette première estimation du seuil ne répond pas aux hypothèses de départ ; on choisit alors un deuxième seuil près du premier (ordinairement supérieur) jusqu'à ce que les critères d'ajustement du modèle soient satisfaits. Il faut toutefois signaler que d'autres seuils peuvent convenir. En général, il existe même une plage de débits à l'intérieur de laquelle le seuil peut être choisi. De plus, il est important en pratique de ne pas faire reposer les estimations de Q_T sur un seul seuil, mais de faire les ajustements pour une série de seuils différents afin de s'assurer que les critères d'ajustement utilisés

sont satisfaits et que, dans une plage suffisante, les estimations sont quasiment indépendantes du seuil.

REMERCIEMENTS

Cette étude a reçu le soutien financier du Conseil de recherches en sciences naturelles et en génie du Canada. Les auteurs tiennent à remercier les réviseurs de ce texte pour leurs commentaires judicieux.

RÉFÉRENCES BIBLIOGRAPHIQUES

- ASHKAR F. et ROUSSELLE J., 1983. Some Remarks on the Truncation Used in Partial Flood Series Models, *Water Resources Research*, **19** (2) : pp.477-480.
- ASHKAR F. et ROUSSELLE J., 1987. Partial Duration Series Modelling Under the Assumption of a Poissonian Flood Count, *Journal of Hydrology*, **90** : 135-144.
- BERNIER J., 1967. Sur la théorie du renouvellement et son application en hydrologie, *Electricité de France*, *HYD67*, 10 p.
- BMDP 1981, Statistical Software, UCLA, University of California Press, 725 p.
- BORGMAN, L.E., 1963. Risk Criteria, *Journal of the Waterways and Harbors Division, ASCE*, **89** : 1-35.
- BOUCHER S., 1986. Analyse de données : description et utilisation des banques de données physiographiques et hydrométriques, *Projet thématique CRNSG-G1617, Modélisation statistique et régionalisation des caractéristiques de crues au Canada*, Volet C, 226 p. et références bibliographiques.
- CORREIA F.N., 1985. Engineering Risk in Flood Studies Using Multivariate Partial Duration Series, ASI, NATO, Tuscon, AZ, 32 p.
- CORREIA F.N., 1987. Multivariate Partial Duration Series in Flood Risk Analysis, *Proceedings of the International Symposium on Flood Frequency and Risk Analysis*, 14-17 May 1986, Louisiana State University, Baton Rouge, Hydrologic Frequency Modeling, V.P. Singh (ed.), Reidel Publishing Company, pp. 541-554.
- DALRYMPLE T., 1960. Flood Frequency Analysis, Manual of Hydrology : Part 3, Flood-Flow Tech., *Geol. Surv. Water-Supply Paper 1543-A*, 80 p.
- EL-JABI N., ASHKAR F., ROUSSELLE J. et HOANG V.D., 1986. Etude stochastique des crues au Québec, Ministère de l'Environnement, Québec, Direction des relevés aquatiques, *H.P.-59*, 91 p.
- GUMBEL E.J., 1958. Statistics of Extremes, Columbia University Press, N.Y., 371 p.
- IRVINE K.N. et WAYLEN P.R., 1986. Partial Series Analysis of High Flows in Canadian Rivers, *Canadian Water Resources Journal*, **11** (2) : pp. 83-91.
- JUDGE D.G., JAMES E.A., McCLENNAN B.I., et PARK E.T., 1988. Hydrologic Design Methodologies for Prefeasibility Studies of Small-Scale Hydro at Ungauged Sites, *Canadian Journal of Civil Engineering*, **15** (3) : pp.289-298.
- KITE G.W., 1978. Frequency and Risk Analysis in Hydrology, Water Resources Publications, Fort Collins, Colorado, 224 p.

- KONECNY F. et NACHTNEBEL H.P., 1985. Extreme Value Processes and the Evaluation of Risk in Flood Analysis, *Appl. Math. Modelling*, 9 : pp. 11-15.
- LANGBEIN W.B., 1949. Annual Flood and the Partial-Duration Flood Series, *Transactions American Geophysical Union*, 30 : 879-881.
- LAURENSEN E.M., 1987. Back to Basic on Flood Frequency Analysis, The Institution of Engineers, Australia, *CE29 (2)* : pp. 47-53.
- MIQUEL J., 1984. *Guide pratique d'estimation des probabilités de crues*, Collection de la Direction des Études et Recherches d'Électricité de France, n° 53, Édition Eyrolles, 160p.
- TAESOMBUT V. et YEVJEVICH V., 1978. Use of Partial Duration Series for Estimating Distribution of Maximum Annual Flood Peak, *Hydrology Paper n° 97*, Colorado State Univ., Fort Collins, Colorado, 71 p.
- ROUSSELLE J., 1972. On Some Problems of Flood Analysis, Ph.D. Thesis, Colorado State University, Fort Collins, Colorado, 226 p.
- SHANE R.M. et LYNN W.R., 1964. Mathematical Model for Flood Risk Evaluation, *Journal of the Hydraulics Division, ASCE*, 90 : pp. 1-20.
- TODOROVIC P., 1970. On Some Problems Involving Random Number of Random Variables, *Annals of Mathematical Statistics*, 41 : pp.1059-1063.
- TODOROVIC P. et ROUSSELLE J. 1971. Some problem of Flood Analysis, *Water Resources Research*, 7 : pp. 1144-1150.
- TODOROVIC P. et ZELENHASIC E., 1970. A Stochastic Model for Flood Analysis, *Water Resources Research*, 6 : 1641-1648.
- WAYLEN P.R. et WOO M.-K., 1983a. Stochastic Analysis of High Flows in Some Central British Columbia Rivers, *Canadian Journal of Civil Engineering*, 10 : pp. 205-213.
- WAYLEN P.R. et WOO M.-K., 1983b. Stochastic Analysis of High Flows Generated by Mixed Processes, *Canadian Journal of Civil Engineering*, 10 : pp. 639-648.
- ZELENHASIC E., 1970. Theoretical Probability Distributions for Flood Peaks, *Hydrology Paper n° 42*, Colorado State University, Fort Collins, Colorado, 35 p.