

UNIVERSITÉ DE MONTRÉAL

PROCESSUS DE POISSON FILTRÉ UTILISÉ POUR LA
MODÉLISATION, L'ESTIMATION ET LA PRÉVISION
DES DÉBITS D'UN FLEUVE

FATIMA BENSALMA

DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(MATHÉMATIQUES APPLIQUÉES)

DÉCEMBRE 2011

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

PROCESSUS DE POISSON FILTRÉ UTILISÉ POUR LA
MODÉLISATION, L'ESTIMATION ET LA PRÉVISION
DES DÉBITS D'UN FLEUVE

présenté par : BENSALMA Fatima

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. LE DIGABEL, Sébastien, Ph.D., président

M. LEFEBVRE, Mario, Ph.D., membre et directeur de recherche

M. FUAMBA, Musandji, Ph.D., membre

REMERCIEMENTS

J'exprime ma sincère reconnaissance et mon plus profond respect à mon directeur de recherche, le professeur Mario Lefebvre, pour sa réelle disponibilité et sa grande contribution à l'accomplissement de ce travail.

RÉSUMÉ

Cette étude examine le processus de Poisson filtré $X(t)$ proposé par Lefebvre et Guilbault (2008) comme modèle pour représenter les débits journaliers d'un fleuve. Soit

$$X(t) = \sum_{n=1}^{N(t)} Y_n (t - \tau_n)^k e^{-(t-\tau_n)/c}, \text{ pour } t \geq \tau_n,$$

où $\{N(t), t \geq 0\}$ est un processus de Poisson homogène de taux $\lambda > 0$, Y_n est une série de variables aléatoires indépendantes et identiquement distribuées (i.i.d) d'une distribution exponentielle de paramètre μ et indépendantes du processus $\{N(t), t \geq 0\}$ et τ_n sont les instants d'arrivée des événements ou signaux du processus de Poisson. Les paramètres du modèle λ et μ sont estimés par la méthode des moments à l'état asymptotique du processus une fois que les paramètres c et k seront trouvés par une approche statistique basée sur les coefficients de corrélation théoriques du modèle. Finalement, la qualité et la performance du modèle sont évaluées par le biais des coefficients de corrélation théoriques comparativement au modèle classique (lorsque $k = 0$) couramment utilisé en hydrologie et par la capacité prévisionnelle du modèle par rapport au modèle classique et à un modèle autorégressif. Une application sur les fleuves Delaware et Hudson situés aux États-Unis est présentée. Les résultats favorisent en général le modèle proposé par Lefebvre et Guilbault (2008).

ABSTRACT

This study examines the filtered Poisson process $X(t)$ proposed by Lefebvre and Guilbault (2008) as a model to represent the daily river flows of rivers. Let

$$X(t) = \sum_{n=1}^{N(t)} Y_n (t - \tau_n)^k e^{-(t-\tau_n)/c}, \text{ for } t \geq \tau_n,$$

where $\{N(t), t \geq 0\}$ is a homogeneous Poisson process with rate $\lambda > 0$, Y_n is a series of random variables independent and identically distributed (i.i.d) having an exponential distribution with parameter μ and independent of the process $\{N(t), t \geq 0\}$, and τ_n are the arrival times of the events or signals of the Poisson process. The model parameters λ and μ are estimated by the method of moments in the asymptotic state of the process once the parameters c and k are found by a statistical approach based on the theoretical correlation coefficients of the model. Finally, the quality and the performance of the model are evaluated through the theoretical correlation coefficient compared to the conventional model (with $k = 0$) commonly used in hydrology, and through the predictive power of the model compared to both the traditional model and an autoregressive model. An application on the Hudson and Delaware Rivers located in the United States is presented. The results generally favor the model proposed by Lefebvre and Guilbault (2008).

TABLE DES MATIÈRES

REMERCIEMENTS.....	iii
RÉSUMÉ	iv
ABSTRACT.....	v
TABLE DES MATIÈRES	vi
LISTE DES TABLEAUX	viii
LISTE DES FIGURES	ix
LISTE DES ANNEXES	xi
INTRODUCTION	1
CHAPITRE 1 : CONTEXTE ET SURVOL DE LA LITTÉRATURE	5
CHAPITRE 2 : MODÉLISATION, ESTIMATION ET PRÉVISION	17
2.1 Modélisation, estimation.....	17
2.2 Approche d'estimation des paramètres k et c	24
2.3 Prévision	26
2.3.1 Modèle avec $k = 1$	26
2.3.2 Modèle avec $k = 0$	30
CHAPITRE 3 : APPLICATION ET RÉSULTATS	31
3.1 Modèle avec la valeur du paramètre k estimée	32
3.2 Comparaison avec les modèles lorsque $k = 0$ et $k = 1$	34
3.3 Stabilité du paramètre k	37
3.4 Capacité prévisionnelle du modèle	39
3.4.1 Cas lorsque $k = 1$	40
3.4.2 Cas lorsque $k = 0$	42

CONCLUSION.....51

BIBLIOGRAPHIE.....54

ANNEXES :57

LISTE DES TABLEAUX

Tableau 2.1 : Caractéristiques du débit en (m ³ /s).....	31
Tableau 2.2 : Écart entre les coefficients de corrélation théoriques et empiriques des débits $X(t)$ et $X(t + \delta)$ pour $\delta = 1, 2, \dots, 7$, $\hat{k} = 0,63$ et $\hat{c} \simeq 3,4798$ (Delaware).....	34
Tableau 2.3 : Écart entre les coefficients de corrélation théoriques et empiriques des débits $X(t)$ et $X(t + \delta)$ pour $\delta = 1, 2, \dots, 7$, $k = 0$ et $\hat{c} \simeq 9,613$ (Delaware).....	36
Tableau 2.4 : Écart entre les coefficients de corrélation théoriques et empiriques des débits $X(t)$ et $X(t + \delta)$ pour $\delta = 1, 2, \dots, 7$, $k = 1$ et $\hat{c} \simeq 2,1923$ (Delaware).....	36
Tableau 2.5 : Variabilité du paramètre k	38
Tableau 2.6 : Critères de comparaison de prévisions entre modèles pour le fleuve Delaware.....	45
Tableau 2.7 : Critères de comparaison de prévisions entre modèles pour le fleuve Hudson.....	46

LISTE DES FIGURES

Figure 2.1 : La fonction de réponse unitaire de l'hydrogramme

$$h(t) = (t - \tau)^k e^{-(t-\tau)/c} \dots\dots\dots 25$$

Figure 2.2 : Hydrogramme du débit en (m³/s) des fleuves Delaware et Hudson (01/10/2002 à 30/09/2003)..... 31

Figure 2.3 : Les valeurs de $\sum_{\delta=1}^7 |\rho(\delta) - r(\delta)|$ en fonction des valeurs du paramètre k : prises de 0,01 à 3..... 33

Figure 2.4 : Les valeurs de $\sum_{\delta>20} |\rho(\delta) - r(\delta)|$ en fonction des valeurs du paramètre k : prises de 0,01 à 3..... 33

Figure 2.5 : Hydrogramme du débit en (m³/s) du fleuve Delaware de 2002 à 2009..... 39

Figure 2.6 : Hydrogramme du débit en (m³/s) du fleuve Hudson de 2002 à 2009..... 39

Figure 2.7 : Débits (m³/s) prévus et observés du fleuve Delaware (21/01/2004 au 30/04/2004)..... 47

Figure 2.8 : Débits (m³/s) prévus et observés du fleuve Delaware (01/10/2003 au 09/01/2004)..... 48

Figure 2.9 : Débits (m³/s) prévus et observés du fleuve Hudson (21/01/2004 au

30/04/2004).....	49
Figure 2.10 : Débits (m ³ /s) prévus et observés du fleuve Hudson (01/10/2003 au 09/01/2004).....	50
Figure A.1 : Exemple de la trajectoire d'un processus de Poisson	57
Figure A.2 : Exemple de la trajectoire de la fonction de réponse unitaire de l'hydrogramme; $h(t) = (t - \tau) e^{-(t-\tau)/c}$	57
Figure A.3 : Exemple de la trajectoire du processus de Poisson filtré (<i>Shot noise process</i>).....	58
Figure A.4 : Forme de l'HUI par le modèle de Nash (fonction gamma $\Gamma(n, k)$)...	58

LISTE DES ANNEXES

ANNEXE A : Illustration des trajectoires de processus.....	57
--	----

INTRODUCTION

Le cadre de la recherche de ce travail est la modélisation des débits journaliers d'un cours d'eau à partir du processus de Poisson filtré. Soit $\{N(t), t \geq 0\}$ un processus de Poisson homogène de taux $\lambda > 0$; le processus stochastique $\{X(t), t \geq 0\}$ est le processus de Poisson filtré défini par

$$X(t) = \sum_{n=1}^{N(t)} w(t, \tau_n, Y_n) \quad (X(t) = 0 \text{ si } N(t) = 0), \quad (1)$$

où Y_n est une série de variables aléatoires indépendantes et identiquement distribuées (i.i.d) et indépendantes du processus $\{N(t), t \geq 0\}$, τ_n sont les instants d'arrivée des événements ou signaux du processus de Poisson. La fonction w est appelée fonction de réponse.

L'hydrologie est un domaine, parmi différents autres, d'applications de ce processus. Le modèle défini en (1) représente le débit du fleuve ou de rivière à l'instant t . Ce modèle est couramment utilisé dans la littérature hydrologique pour générer les données de débits à de courtes échelles de temps (journalières à hebdomadaires). En particulier, ce modèle est capable de reproduire la présence des pics et de récessions dans la trajectoire des débits journaliers. Ce modèle s'exprime avec une fonction de réponse définie par

$$w(t, \tau_n, Y_n) = Y_n e^{-(t-\tau_n)/c}, \quad (2)$$

où Y_n sont des variables aléatoires i.i.d d'une distribution exponentielle de paramètre μ , le paramètre c prend une valeur positive qui dépend de chaque fleuve et doit être estimé, τ_n sont les instants d'arrivée des précipitations, par exemple, et Y_n sont les quantités d'eau observées à ces instants. On suppose donc ici, que les événements arrivent selon un processus de Poisson $\{N(t), t \geq 0\}$ de taux $\lambda > 0$ indépendant de la séquence Y_n , où n prend des valeurs entières. En pratique, la variable du temps est discrète, puisque les débits ne sont pas mesurés de façon continue (en général, le débit est mesuré quotidiennement). La fonction de réponse donne la valeur à l'instant t d'un signal (les précipitations) qui se réalise à l'instant τ_n et pour lequel la quantité Y_n a été ajoutée au

processus. La variable aléatoire $X(t)$ traduit alors la somme à l'instant t de l'effet de tous les signaux qui se sont produits depuis l'instant initial (Lefebvre, 2005b). Les figures 1 et 2 (en annexe A) illustrent un exemple de la trajectoire d'un processus de Poisson et un schéma de la fonction de réponse unitaire de l'hydrogramme respectivement. La variable Y_n est considérée comme la variable entrante dans le système, le terme $e^{-(t-\tau_n)/c}$ dans l'équation (2) permet de prendre en compte la décroissance graduelle du débit après une pointe.

Le modèle classique défini en (1), avec la fonction de réponse en (2), a conduit à des résultats relativement bons dans plusieurs études antérieures, voir par exemple (Weiss, 1977; Kelman, 1980; Koch, 1985; Konecny, 1992; Seidou *et al.*, 2002; Lefebvre *et al.*, 2003). La forme de la fonction de réponse est choisie principalement pour des raisons de simplicité, et l'hypothèse que les variables aléatoires Y_n sont exponentiellement distribuées est supposée tenir sans être testée (Lefebvre, 2005a; Claps *et al.*, 2005). Ce modèle a été utilisé par Lefebvre *et al.* (2003) pour prévoir les crues du fleuve Delaware. Étant donné que la variable dénotant le temps qui s'écoule entre l'arrivée des événements ne suit pas exactement une distribution exponentielle, Lefebvre (2005a) a utilisé une transformation (la racine carrée) de la variable temps de sorte que la nouvelle variable suivra une distribution exponentielle et, de ce fait, l'auteur a modélisé les débits par un processus de Poisson filtré transformé.

En outre, ce modèle suppose que l'effet de l'événement arrivé à l'instant τ_n est maximal à cet instant et commence aussitôt à décroître (Lawrance et Kottegoda, 1977; Konecny, 1992; Murrone *et al.*, 1997). La figure A.3 (en annexe A) montre un exemple de la trajectoire d'un processus de Poisson filtré; voir aussi Bado et Unny (1990) et Konecny (1992).

Cependant, l'observation d'un hydrogramme montre qu'il y a très souvent des périodes durant lesquelles le débit augmente (non pas d'une façon instantanée) puis tend à décroître plus au moins rapidement.

Plusieurs variantes du modèle ont été proposées dans la littérature hydrologique, la plupart des efforts ont été consacrés à la définition de la fonction de réponse et à l'estimation des paramètres du processus; voir Konecny (1992), Murrone et al. (1997) et Lefebvre et Guilbault (2008).

La fonction de réponse proposée par Lefebvre et Guilbault (2008) est l'expression suivante :

$$w(t, \tau_n, Y_n) = Y_n(t - \tau_n)^k e^{-(t - \tau_n)/c}. \quad (3)$$

En fait, cette fonction de réponse présente une forme qui ressemble à une partie de l'hydrogramme (comparable à la forme d'une distribution gamma), à l'opposé de la forme classique en (2) qui suppose une augmentation instantanée du débit suivie par une longue période de décroissance (voir figures 2 et 4 en annexe A). En théorie, le paramètre k peut prendre sa valeur dans l'intervalle $[0, \infty)$ et doit être estimé. Nous montrons dans ce travail que ce paramètre prendrait des valeurs plus élevées (soit $1 \leq k \leq 2$ pour un échantillon de débits d'une année) lorsque la dynamique du fleuve se distingue par l'arrivée de quelques crues exceptionnelles. Il y a lieu de noter que l'une des caractéristiques du modèle, qui stipule que les temps entre les arrivées des événements sont d'une distribution exponentielle, est dans la plupart des cas non satisfaite en pratique, et que de ce fait le processus $N(t)$ ne serait pas considéré comme un processus de Poisson.

Cette conceptualisation mathématique de la fonction de réponse devrait améliorer la qualité du modèle; c'est l'objectif d'analyse dans cette étude. Une approche basée sur les propriétés statistiques du processus est proposée, suivant les résultats de Lefebvre et Guilbault (2008), pour estimer les paramètres du modèle et tester sa qualité d'ajustement aux observations. En outre, on vérifie la performance du modèle dans les prévisions, en se servant de certains indicateurs d'écart, comparativement aux modèles classique et autorégressif. Enfin, l'apport de ce travail se résume en deux points : 1) nous avons trouvé une méthode pour estimer significativement le paramètre k du processus. 2) Nous avons développé une formule qui permet de prévoir le débit futur lorsque $k = 1$.

Trois chapitres seront consacrés à répondre à cet objectif; dans le premier chapitre, nous dressons le contexte général de l'application du processus de Poisson filtré en survolant la littérature hydrologique sur la modélisation des débits des cours d'eau. Le second chapitre porte sur les résultats théoriques de la modélisation, l'estimation et la prévision, en utilisant le processus de Poisson filtré. L'application de ces résultats sur les débits des fleuves Delaware et Hudson, situés aux États-Unis, fera l'objet du dernier chapitre. Enfin, nous concluons par une discussion et interprétation des constats de l'étude.

CHAPITRE 1 :

CONTEXTE ET SURVOL DE LA LITTÉRATURE

La modélisation théorique d'un système hydrologique a été au cœur de la recherche depuis de longues années; le but de cette modélisation est de représenter le système par une fonction reliant les intrants et les extrants et qui décrit, jusqu'à un certain niveau de détails, les interactions et les mécanismes des éléments physiques du système ainsi que les phases du mouvement de l'eau. Présentement, les recherches sont orientées sur le développement de modèles stochastiques qui reposent soit sur une approche physique des processus, soit sur une approche conceptuelle (Bado et Unny, 1990). La formulation d'un modèle stochastique spécifique pour générer les débits d'un cours d'eau, à court terme, constitue l'objet de cette étude. Nous présenterons, dans ce qui suit, une brève revue de littérature sur certains travaux importants dans la modélisation des débits journaliers, en commençant par la théorie linéaire de l'hydrogramme unitaire instantané (HUI) formalisé par Nash (1957) et par le processus de Poisson filtré proposé par Bernier *et al.* (1970) et Weiss (1977) par la suite. En effet, ces travaux reposent sur la notion de l'HUI sous l'hypothèse que la seule variable entrante dans le système est la précipitation. Le point commun est que la distribution marginale du processus de Poisson filtré et la fonction de réponse de l'hydrogramme est une fonction gamma à deux paramètres. On passe en revue d'autres variantes de la fonction de réponse du processus de Poisson filtré qui ont été mises au point dans les recherches actuelles sur la physique des processus. Par ailleurs, la formulation conceptuelle du processus générant les débits journaliers constitue le cadre de cette étude; cet aspect conceptuel prend compte, même de façon simple, l'aspect physique du système. En effet, le modèle proposé par Lefebvre et Guilbault (2008) est une forme conceptuelle qui repose sur la notion de l'HUI. Autrement dit, la forme de la fonction de réponse du modèle traduit les brèves périodes de montée suivies par des récessions lentes qui caractérisent l'hydrogramme des débits journaliers et qui tend à suivre théoriquement une distribution gamma.

Dans le but de modéliser la dynamique et la réponse du bassin versant, l'approche du réservoir linéaire de stockage est la méthode sur laquelle Nash (1957) s'est basé pour proposer un modèle de l'HUI. Il suppose que la trajectoire du débit dans un bassin versant avec une chute de pluie instantanée est équivalente à l'écoulement à travers une succession linéaire de réservoirs, le débit sortant d'un réservoir devenant l'entrée pour le suivant. Ainsi, le débit du n^{e} réservoir est donné par :

$$Q_n = \frac{1}{k\Gamma(n)} e^{-t/k} (t/k)^{n-1}. \quad (4)$$

Le système de réservoirs linéaire est caractérisé par la relation linéaire entre le stockage $s(t)$ et le débit sortant $Q(t)$ pour chaque réservoir à l'instant t , telle que $s(t) = k \times Q(t)$, où k est un coefficient correspondant à la capacité moyenne de stockage du réservoir dans le temps; on l'appelle en général "coefficient de stockage ou du bassin versant".

Le débit Q_1 du premier réservoir obéit à la fonction de réponse impulsive (FRI) du réservoir linéaire, qui représente la réponse du système à une impulsion unitaire instantanée à l'instant $t = 0$:

$$Q_1 = \frac{1}{k} e^{-t/k}.$$

Q_1 devient le débit entrant dans le second réservoir pour obtenir le débit sortant Q_2 du second réservoir. Ainsi de suite, les débits des autres réservoirs en série linéaire se calculent comme suit :

$$\begin{aligned} Q_2 &= \frac{1}{k} e^{-t/k} \int_0^t e^{t/k} Q_1 dt \\ &= \frac{1}{k} e^{-t/k} \int_0^t e^{t/k} \frac{1}{k} e^{-t/k} dt \\ &= \frac{1}{k^2} e^{-t/k} t, \end{aligned}$$

$$\begin{aligned}
 Q_3 &= \frac{1}{k} e^{-t/k} \int_0^t e^{t/k} \frac{1}{k^2} e^{-t/k} t \, dt \\
 &= \frac{1}{k^3} e^{-t/k} \frac{t^2}{2},
 \end{aligned}$$

$$Q_4 = \frac{1}{k^4} e^{-t/k} \frac{t^3}{3!},$$

⋮

$$Q_n = \frac{1}{k^n} e^{-t/k} (t/k)^{n-1} / (n-1)!,$$

Le débit du n^{e} réservoir (la dernière équation) peut se simplifier en remplaçant la factorielle par la fonction gamma, et obtenir l'équation (4).

L'hydrogramme unitaire instantané correspondant au modèle de Nash (1957) peut être obtenu par la fonction de distribution gamma à deux paramètres $\Gamma(n, k)$. La forme de la courbe de l'HUI est représentée par la figure A.4 (annexe A) pour différentes valeurs de n (paramètre de forme) et k (paramètre d'échelle). Les paramètres n et k peuvent être estimés par la méthode des moments.

Bernier *et al.* (1970) furent les premiers qui ont proposé le processus de Poisson filtré, appelé aussi "*shot noise*". Ils ont considéré explicitement le processus intermittent et le caractère "poissonnien" des précipitations et la fonction déterministe reliant les précipitations et les débits journaliers. Le débit peut évoluer de façon déterministe lorsqu'il n'y a plus d'incitations (précipitations) sur le bassin et suit ainsi une trajectoire obéissant à une loi particulière; il s'agit de la période de tarissement. Les précipitations sont considérées comme l'élément important entrant dans le processus aléatoire de la formation de débits. Les périodes de décrues des débits suivent en alternance les périodes de montées des débits causées par l'alimentation du bassin par les précipitations.

Selon Bernier *et al.* (1970), le débit à l'instant t peut être donné sous forme d'un processus continu, exprimé par

$$Q(t) = \int_0^t w(t, \tau) Y(\tau) dN(\tau), \quad (5)$$

où $w(t, \tau)$ est la fonction de transfert déterministe de l'hydrogramme; la forme la plus simple de cette dernière dans le cas d'un processus stationnaire est une exponentielle :

$$w(t, \tau) = \theta e^{-\theta(t-\tau)}, \quad \theta > 0.$$

$Y(\tau)$ est une série de variables aléatoires qui mesurent l'intensité des précipitations et sont supposées indépendantes et identiquement distribuées selon une loi exponentielle, $N(\tau)$ est un processus ponctuel de comptage, c'est-à-dire que $N(\tau)$ est le nombre d'averses produites dans l'intervalle $[0, \tau]$.

La structure de ce modèle n'est pas valable pour des régimes hydrologiques complexes, là où d'autres mécanismes aléatoires se combinent à la pluie, comme la fonte des neiges liée à la température. Cependant, la forme de l'hydrogramme peut être réaliste en période de tarissement du débit.

Weiss (1977) a étudié plus en détail le processus "*shot noise*" ou processus de Poisson filtré en le développant comme un modèle simple et physiquement réaliste qui reproduit les récessions des débits journaliers. Ce modèle développé équivaut à un processus autorégressif de premier ordre, mais dans lequel les innovations ne sont pas gaussiennes. La structure du processus de Poisson filtré consiste en un processus ponctuel reproduisant l'occurrence des événements de précipitations effectives (impulsions ou signaux) considérées comme l'entrée du système; la propagation de ces entrées est ensuite filtrée par la fonction de réponse du système pour obtenir l'écoulement (Murrone *et al.*, 1997).

Le débit est une variable aléatoire $X(t)$ représentée en temps continu par le processus de Poisson filtré (Weiss, 1977) :

$$\begin{aligned}
X(t) &= \sum_{n=1}^{N(t)} w(t - \tau_n, Y_n) \\
&= \sum_{n=1}^{N(t)} Y_n h(t - \tau_n) \\
&= \sum_{n=1}^{N(t)} Y_n e^{-b(t-\tau_n)}, \quad t \geq \tau.
\end{aligned} \tag{6}$$

où $N(t)$ est un processus de Poisson de taux λ , Y_n est l'entrée du processus (quantité des averses) qui est une variable aléatoire de distribution exponentielle de paramètre μ , w est la fonction de réponse du système qui décrit l'évolution et la propagation de l'impulsion (l'effet du signal) dans le temps, et $h(t - \tau_n) := e^{-b(t-\tau_n)}$ est la fonction de transfert déterministe de l'hydrogramme. Ce processus a trois paramètres à estimer : la moyenne des sauts $\theta (= 1/\mu)$, le taux d'occurrence des événements λ et le taux de décroissance $b (= 1/c)$. Weiss (1977) a utilisé la méthode des moments pour l'estimation des paramètres.

Le modèle présenté par Weiss (1977) se compose alors d'un processus couplé Poisson-exponentiel, reproduisant les occurrences et intensités des précipitations effectives. L'écoulement du bassin versant a, par hypothèse, une réponse exponentielle à la précipitation (Murrone *et al.*, 1997).

Pour modéliser la récession du débit, sous l'hypothèse du système linéaire, Weiss (1977) propose la forme suivante du processus $X(t)$:

$$X(t) = \int_0^t h(\tau) dY(t - \tau), \tag{7}$$

où $dY(t)$ est un processus indépendant et non corrélé décrivant tout l'aléa dans $X(t)$, et $h(\tau)$ est la fonction de réponse du système.

Cette dernière équation est définie de la même façon que l'équation (5); la somme en (6) sur l'ensemble des effets d'impulsions produits dans l'intervalle $(0, t)$ remplace

l'intégrale en (7). Weiss (1977) affirme que les événements aléatoires entrant dans la formation de $X(t)$ se traduisent par l'occurrence des crues à des instants aléatoires suivant un processus de Poisson.

Pour appliquer ce processus continu à la série de débits qui est de nature discrète, il est possible de représenter le processus sous une forme discrète $X(n)$ en l'intégrant sur un intervalle de temps de longueur T (une journée par exemple) :

$$X(n) = \int_{(n-1)T}^{nT} X(t)dt. \quad (8)$$

D'après Weiss (1977), ce processus $X(t)$ suit une distribution gamma (Pearson type 2) de paramètres $(\lambda c, \mu)$, de densité

$$f(X) = \frac{\mu^{\lambda c} X^{\lambda c - 1}}{\Gamma(\lambda c)} e^{-\mu X}, \quad (9)$$

ce qui permet d'obtenir les moments suivants :

$$E[X(t)] = \frac{\lambda c}{\mu}, \quad (10)$$

$$V[X(t)] = \frac{\lambda c}{\mu^2}, \quad (11)$$

$$\rho[X(t), X(t+s)] = e^{-s/c}, \quad s > 0. \quad (12)$$

Le choix d'une distribution exponentielle pour la fonction de transfert $h(\tau)$ rend le processus $X(t)$ un processus autorégressif d'ordre 1, $AR(1)$, et correspond à un réservoir linéaire unique. En plus, $h(\tau)$ détermine la pente de récession de $X(t)$ (Weiss, 1977).

Le débit à l'instant $t + s$ est donné par

$$\begin{aligned}
X(t+s) &= \sum_{i=1}^{N(t+s)} Y_i e^{-b(t+s-\tau_i)} \\
&= e^{-bs} \sum_{i=1}^{N(t)} Y_i e^{-b(t-\tau_i)} + \sum_{i=N(t)}^{N(t+s)} Y_i e^{-b(t+s-\tau_i)} \\
&= e^{-bs} X(t) + \epsilon_s(t+s),
\end{aligned}$$

$$\text{où } \epsilon_s(t+s) = \sum_{i=N(t)}^{N(t+s)} Y_i e^{-b(t+s-\tau_i)}.$$

Le processus de Poisson filtré, selon Weiss (1977), n'est que le processus $AR(1)$ en temps continu, mais dont les innovations $\epsilon_s(t+s)$ ont une distribution asymétrique (voir distribution gamma) et non pas gaussienne. La probabilité qu'aucun événement ne se produise dans l'intervalle $(t, t+s)$ est la probabilité que l'innovation soit nulle; cette probabilité est strictement positive :

$$P[\epsilon_s(t+s) = 0] = P[N(t+s) - N(t) = 0] > 0.$$

Étant donné le résultat de Weiss sur le réalisme des processus de Poisson filtrés pour modéliser les débits journaliers de rivières, par rapport aux autres modèles caractérisés par des entrées gaussiennes, le processus de Poisson filtré est défini par des discontinuités aux instants d'arrivée des événements (signaux), suivies par des décroissances exponentielles (Bado et Unny, 1990). Autrement dit, la trajectoire du processus recommence au sommet de chaque crue qui est atteinte au début de chaque signal. La figure A.3 (annexe A) montre une réalisation typique d'un processus de Poisson filtré; voir aussi Bado et Unny (1990) et Konecny (1992). Dans ce cas, la trajectoire entre deux impulsions successives peut être représentée par (Bado et Unny, 1990)

$$X(t) = X(t_0) e^{-b(t-t_0)}.$$

Le processus de Poisson filtré est bien connu pour modéliser les récessions de l'hydrogramme, mais il reste que la période de l'augmentation du débit n'est pas considérée comme un élément composant de ce modèle.

Weiss (1977) a développé, par ailleurs, un modèle composé de deux processus de Poisson filtrés représentant respectivement le débit de surface et le débit de base. Par ce modèle, les statistiques mensuelles peuvent être préservées en plus des statistiques quotidiennes.

La modélisation stochastique pour la simulation des débits journaliers, en particulier, est très vaste. Bernier *et al.* (1970) ont dressé un inventaire de ces modèles; voir aussi Bado et Unny (1990). Selon Lawrence et Kottegoda (1977), il y a deux défis importants dans la génération des débits journaliers synthétiques : le premier est la modélisation d'une série de variables dépendantes et non gaussiennes avec une grande variabilité; le deuxième est que la série des débits journaliers est caractérisée par des augmentations rapides suivies par des diminutions lentes. Étant donné que les modèles autorégressifs gaussiens ne peuvent pas reproduire fidèlement les caractéristiques de débits journaliers, l'approche théorique prometteuse est de pouvoir reproduire le mouvement ascension-récession de l'hydrogramme via un processus autorégressif avec un terme d'innovation asymétrique (Lawrance et Kottegoda, 1977).

Treiber et Plate (1977) ont développé un modèle plus général que celui de Weiss (1977). Le débit est généré d'un système linéaire, et à chaque jour pluvieux correspond une entrée (signal) du système. La séquence des jours secs et pluvieux est supposée constituer une chaîne de Markov. Par conséquent, les amplitudes des entrées sont reproduites par un modèle autorégressif de premier ordre. Les entrées sont transformées en débits par la fonction de réponse déterministe du système, et suivent une distribution exponentielle transformée. Le modèle a fourni de bon résultats, mais il n'est pas parcimonieux, vu le grand nombre de paramètres qu'il nécessite à déterminer.

Kelman (1980) s'est intéressé à la modélisation séparée de la montée et de la baisse des débits, la baisse étant soumise à la règle de vidange du bassin et la montée, causée par l'arrivée d'un événement, est modélisée de la même façon que les précipitations. En considérant deux processus intermittents alternatifs (voir aussi Évora, 1997, Koch, 1985, et Bado et Unny, 1990), l'auteur propose un modèle conceptuel basé sur un système de deux réservoirs linaires.

Koch (1985) a proposé une version à deux variables du modèle de Weiss (1977), basée sur des arguments du système physique. Le bassin est représenté comme un système de deux composants : le premier décrit la surface et l'autre le ruissellement aquifère. La combinaison linéaire des réponses de ces deux composants traduit la fonction de réponse impulsive du bassin. L'écoulement de surface a les précipitations excédentaires comme intrant, alors que le ruissellement aquifère est dû à la recharge de la zone durant l'infiltration et le drainage. Le processus du débit est représenté en temps continu comme la somme de ces deux composants. Les entrées de ce système, les précipitations excédentaires et la recharge, sont aléatoires et deviennent décroissantes par le processus d'acheminement. Le processus stochastique du débit $Q(t)$ à l'instant t est représenté par le processus de Poisson filtré donné par

$$\begin{aligned} Q(t) &= Q_0(t) + \sum_{i=1}^{N(t)} Q_{s_i}(t, \tau_i) + Q_{b_i}(t, \tau_i) \\ &= Q_0(t) + \sum_{i=1}^{N(t)} V_i c_1 e^{-c_1(t-\tau_i)} + Q_{D_i} c_2 e^{-c_2(t-\tau_i)}, \end{aligned}$$

où τ_i l'instant d'occurrence du i° orage;

$N(t)$ processus de Poisson, comptant le nombre d'orages dans l'intervalle $[0, t]$;

$Q_0(t)$ la valeur initiale du débit à l'instant t ;

Q_{s_i} l'écoulement de surface du i° orage;

Q_{b_i} l'écoulement aquifère du i° orage;

V_i le volume des précipitations excédentaires du i° orage;

Q_{D_i} le volume de recharge du i° orage;

c_1 et c_2 sont les coefficients de stockage de l'écoulement de surface et aquifère respectivement.

V , Q_D , $N(t)$ et τ sont des variables aléatoires.

Cette approche permet une concordance entre les paramètres du modèle et les paramètres du système physique (Murrone *et al.*, 1997), mais l'application du modèle s'avère irréalisable étant donné que le problème des estimations des paramètres n'est pas traité directement (Allamano *et al.*, 2007).

Le modèle continu peut être discrétisé sur une période de longueur T comme donné précédemment à l'équation (8). Comme l'hydrogramme est caractérisé par des périodes de montées et de baisses, d'après Koch (1985): "*A further simplification of the continuous model is proposed based on the assumption that the discretization period is "long" relative to storm durations. If this is the case, the rising limb occurs over a short time period and is represented as being instantaneous. This leads to a shot noise representation of the continuous model which is similar to the second order shot noise model of Weiss (1977)*".

Plusieurs autres modèles stochastiques décrivant la série temporelle des débits ont été proposés pour modéliser les débits de cours d'eau, par exemple Salas, 1980, et Évora, 1997. D'autres travaux ont été consacrés à l'analyse des propriétés statistiques des débits de rivières en les modélisant par le processus de Poisson filtré; voir Kavvas et Delleur, 1984, et Yue *et al.*, 1999.

Un autre modèle stochastique dérivé du modèle *shot noise* a été développé par Murrone *et al.* (1997) pour reproduire les données à court terme des débits (journaliers et hebdomadaires). Le système est composé de trois composants en parallèle : l'écoulement de base (souterrain), l'écoulement aquifère et l'écoulement de surface (les deux derniers composants représentent les précipitations directes). Les travaux de Allamano *et al.* (2007) et Claps *et al.* (2005) ont été réalisés sur le même modèle développé par Murrone *et al.* (1997).

La fonction du bassin $h(t)$ est supposée être la combinaison linéaire de la réponse des trois réservoirs en parallèle :

$$h(t) = u_0 + \frac{c_1}{k_1} e^{-t/k_1} + \frac{c_2}{k_2} e^{-t/k_2} + \frac{c_3}{k_3} e^{-t/k_3}, \quad t \geq 0,$$

où c_i est un paramètre de recharge des éléments du système respectant la condition de continuité $\sum_i c_i = 1$, et k_i est un coefficient de stockage. L'hydrogramme unitaire instantané est représenté par la fonction de réponse à une impulsion du réservoir linéaire $u_i = \frac{c_i}{k_i} e^{-t/k_i}$. La fonction de réponse du bassin est une combinaison linéaire de fonctions delta de Dirac. Le débit $X(t)$ en temps continu est représenté par le processus de Poisson filtré :

$$X(t) = \sum_{i=1}^{N(t)} Y_i h(t - \tau_i),$$

où Y_i est l'entrée du processus, supposée suivre une distribution exponentielle, et la fonction $h(t)$ est représentée par l'équation précédente.

Claps *et al.* (2005) ont, par ailleurs, considéré d'autres points importants qui, selon eux, ont été négligés par la littérature sur les processus de Poisson filtrés. Pour vérifier les hypothèses de base derrière le modèle de Poisson, à savoir l'indépendance des signaux consécutifs (précipitations effectives) et leur assignation à chaque crue, et la distribution de Poisson de leurs occurrences, les auteurs suggèrent des approches empruntées à l'hydrologie statistique pour estimer la séquence des précipitations effectives du modèle. Ils montrent aussi que la dépendance temporelle des débits journaliers augmente la pertinence d'avoir des modèles parcimonieux au niveau du nombre de paramètres.

La majorité des efforts dans l'étude des processus de Poisson filtré ont porté sur la définition de la fonction de réponse du système et sur les méthodes pour estimer les paramètres; voir les travaux précédents : Bernier *et al.*, 1970, Lawrance et Kottegoda, 1977, Weiss, 1977, Koch, 1985 et Murrone *et al.*, 1997.

Plus récemment, Lefebvre et Guilbault (2008) (voir aussi Lefebvre et Guilbault, 2009) ont proposé une fonction de réponse qui traduit la forme "ascension-récession" du débit journalier, soit

$$w(t, \tau_i, Y_i) = Y_i (t - \tau_i)^k e^{-(t-\tau_i)/c}.$$

Le débit $X(t)$ est alors exprimé par

$$X(t) = \sum_{i=1}^{N(t)} Y_i (t - \tau_i)^k e^{-(t-\tau_i)/c}, \quad t \geq \tau,$$

où k est une constante non négative à estimer.

Soit la fonction

$$h(t, \tau) = (t - \tau)^k e^{-(t-\tau)/c}.$$

Cette fonction permet d'avoir une forme semblable à celle d'une distribution gamma (voir figure A.2, en annexe A). En plus, si $k = 0$, on retrouve la forme classique du modèle de processus de Poisson filtré qui implique des récessions lentes, exponentielles et des augmentations instantanées du débit. Lorsque $k > 0$, le modèle proposé peut être meilleur avec le choix des bons paramètres. Le paramètre k joue le rôle du paramètre de forme de la distribution gamma. L'avantage de ce modèle est qu'il comprend la forme classique d'une part et, d'autre part, permet plus de flexibilité ou de variabilité du débit.

Cette conceptualisation mathématique pour modéliser les débits journaliers est retenue dans notre travail. Nous apporterons de nouvelles contributions à l'estimation des paramètres du modèle proposé et au développement d'un estimateur de la valeur du débit prévu, tout en testant la qualité d'ajustement du modèle et sa capacité prévisionnelle par rapport à d'autres modèles candidats.

CHAPITRE 2 :

MODÉLISATION, ESTIMATION ET PRÉVISION

Le contexte de ce chapitre se fonde sur les résultats théoriques d'estimation des paramètres du processus de Poisson filtré comme modèle du débit de cours d'eau, et d'estimation des valeurs prévues du débit, en parcourant et développant de nouvelles approches suivant les résultats de l'étude menée par Lefebvre et Guilbault (2008) (voir aussi Guilbault, 2008).

2.1 Modélisation, estimation

Le processus de Poisson filtré utilisé pour modéliser le débit $X(t)$ à court terme est donné par

$$\begin{aligned} X(t) &= \sum_{n=1}^{N(t)} w(t, \tau_n, Y_n) \\ &= \sum_{n=1}^{N(t)} Y_n (t - \tau_n)^k e^{-(t-\tau_n)/c}, \text{ pour } t \geq \tau_n \geq 0, \end{aligned} \quad (13)$$

Tel que $k \geq 0$ et $c > 0$. Ce modèle devrait reproduire vraisemblablement l'hydrogramme des débits d'un fleuve. Pour vérifier l'ajustement de ce modèle aux données observées des débits journaliers, il est tout d'abord nécessaire d'étudier les propriétés du processus de Poisson filtré.

- Distribution du nombre d'occurrences $N(t)$ des événements journaliers :

Étant donné que le processus $\{N(t), t \geq 0\}$ d'occurrence de précipitations est supposé suivre une distribution de probabilité de Poisson de paramètre λt , alors

$$P[N(t) = k] = \frac{(\lambda t)^k}{k!} e^{-(\lambda t)},$$

$$E[N(t)] = V[N(t)] = \lambda t.$$

Konecny (1992) a proposé, au lieu du processus de Poisson homogène de taux $\lambda (> 0)$ précédent, un processus non homogène (non stationnaire) avec une fonction d'intensité $\lambda(t) \geq 0$, pour $t \geq 0$. Cette fonction d'intensité décrit certaines tendances dans l'occurrence des événements du processus.

À partir de la définition du processus de comptage $N(t)$, il est évident que la séquence des instants d'arrivée des événements τ_i , pour $i = 0, 1, 2, \dots$, est équivalente au processus $\{N(t), t \geq 0\}$ (voir figure A.1, en annexe A).

Un processus de renouvellement est un processus de comptage où les temps entre les arrivées des événements ou le temps d'inter-occurrence (le temps que le processus passe dans un état $i \in \{0, 1, \dots\}$) $(\tau_i - \tau_{i-1})$ sont des variables i.i.d. Le processus de Poisson homogène est un cas particulier du processus de renouvellement lorsque la variable $(\tau_i - \tau_{i-1})$ est d'une distribution exponentielle.

Par ailleurs, le processus de Poisson est un cas particulier d'une chaîne de Markov en temps continu. Les variables aléatoire τ_i , pour $i = 0, 1, 2, \dots$, sont indépendantes par la propriété de Markov, et sont identiquement distribuées par le fait que le processus de Poisson se caractérise par des accroissements indépendants et stationnaires. En fait, du point de vue probabiliste, le processus recommence à zéro, à partir de n'importe quel instant τ_i . Les variables τ_i ont une distribution exponentielle de paramètre λ (Lefebvre, 2005b).

- **Distribution de l'amplitude Y des événements journaliers :**

La variable Y est en général d'une distribution exponentielle ou gamma, le choix de cette distribution devrait fournir un bon ajustement aux observations (Yue, Hashino *et al.*, 1999). Si Y suit une distribution exponentielle de paramètre μ , alors

$$f_Y(y) = \frac{1}{\mu} e^{-y/\mu},$$

$$E[Y] = \mu,$$

$$V[Y] = \mu^2.$$

- **Fonction caractéristique du processus de Poisson filtré :**

On peut obtenir la fonction caractéristique $C_X(\theta) := E[e^{i\theta X(t)}]$ du débit $X(t)$ en conditionnant sur le nombre d'événements se produisant à l'intérieur de l'intervalle $(0, t]$, comme suit :

$$\begin{aligned} C_{X(t)}(\theta) &= \sum_{k=0}^{\infty} \left(P(N(t) = k) \cdot E \left\{ \exp \left[i\theta \cdot \sum_{j=1}^k w(t, \tau_j, Y_j) \right] \middle| N(t) = k \right\} \right) \\ &= P(N(t) = 0) + \sum_{k=1}^{\infty} \left(P(N(t) = k) \cdot E \left\{ \exp \left[i\theta \cdot \sum_{j=1}^k w(t, \tau_j, Y_j) \right] \middle| N(t) = k \right\} \right). \end{aligned}$$

En utilisant le fait que l'instant de la réalisation d'un événement dans l'intervalle $(0, t]$ est une variable aléatoire U_j d'une distribution uniforme $U(0, t]$, par indépendance et pour $0 < \tau_1 < \tau_2 < \dots < \tau_k \leq t$, la fonction de densité conjointe des variables U_j est donnée par

$$f_{U_1, \dots, U_k}(\tau_1, \dots, \tau_k) = \prod_{j=1}^k f_{U_j}(\tau_j) = \prod_{j=1}^k \frac{1}{t} = \frac{1}{t^k}.$$

Par conséquent,

$$\begin{aligned} &E \left\{ \exp \left[i\theta \cdot \sum_{j=1}^k w(t, \tau_j, Y_j) \right] \middle| N(t) = k \right\} \\ &= \prod_{j=1}^k E \left\{ \exp \left[i\theta \cdot w(t, \tau_j, Y_j) \right] \right\} \\ &= \left(E \left\{ \exp \left[i\theta \cdot w(t, \tau, Y) \right] \right\} \right)^k \\ &= \left(\frac{1}{t} \int_0^t E \left\{ \exp \left[i\theta \cdot w(t, \tau, Y) \right] d\tau \right\} \right)^k. \end{aligned}$$

Il s'ensuit que

$$\begin{aligned}
C_{X(t)}(\theta) &= \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \left(\frac{1}{t} \int_0^t E \left\{ \exp[i\theta \cdot w(t, \tau, Y)] d\tau \right\} \right)^k \\
&= e^{-\lambda t} \sum_{k=0}^{\infty} \frac{1}{k!} \left(\lambda \int_0^t E \left\{ \exp[i\theta \cdot w(t, \tau, Y)] d\tau \right\} \right)^k \\
&= e^{-\lambda t} \exp \left\{ \lambda \int_0^t E \left\{ \exp[i\theta \cdot w(t, \tau, Y)] d\tau \right\} \right\} \\
&= \exp \left\{ \lambda \int_0^t E \left\{ \exp[i\theta \cdot w(t, \tau, Y)] - 1 \right\} d\tau \right\}.
\end{aligned}$$

- **Les statistiques de base du débit $X(t)$: la moyenne, la variance et le coefficient de corrélation :**

À partir de la fonction caractéristique, le calcul des moments par rapport à l'origine $E[X^n(t)]$ et de la covariance $Cov[X(t_1), X(t_2)]$ se fait comme suit :

$$E[X^n(t)] = (-i)^n \frac{d^n}{d\theta^n} C_{X(t)}(\theta) \Big|_{\theta=0},$$

$$Cov[X(t_1), X(t_2)] = (-i)^2 \frac{\partial^2}{\partial \theta_1 \partial \theta_2} C_{X(t_1), X(t_2)}(\theta_1, \theta_2) \Big|_{\theta_1=0, \theta_2=0}.$$

L'espérance et la variance de $X(t)$ ainsi que la covariance de $X(t_1)$ et $X(t_2)$ peuvent être obtenues à partir de ces formules, découlant des calculs précédents (Lefebvre, 2005b):

$$E[X(t)] = \lambda \int_0^t E[w(t, \tau_n, Y_n)] d\tau, \quad (14)$$

$$Var[X(t)] = \lambda \int_0^t E[w^2(t, \tau_n, Y_n)] d\tau, \quad (15)$$

$$Cov[X(t_1), X(t_2)] = \lambda \int_0^{\min(t_1, t_2)} E[w(t_1, \tau_n, Y_n) w(t_2, \tau_n, Y_n)] d\tau. \quad (16)$$

Étant donné que la variable Y_n est supposée suivre une distribution exponentielle de paramètre μ , alors les formules (14) et (15) précédentes peuvent être développées comme suit :

$$\begin{aligned}
 E[X(t)] &= \lambda E[Y_n] \int_0^t (t - \tau_n)^k e^{-(t-\tau_n)/c} d\tau \\
 &= \frac{\lambda}{\mu} \int_0^t (t - \tau_n)^k e^{-(t-\tau_n)/c} d\tau \\
 &\stackrel{k > -1}{=} \frac{\lambda}{\mu} c^{k+1} \gamma\left(k + 1, \frac{t}{c}\right), \tag{17}
 \end{aligned}$$

$$\begin{aligned}
 Var[X(t)] &= \lambda E[Y_n^2] \int_0^t (t - \tau_n)^{2k} e^{-2(t-\tau_n)/c} d\tau \\
 &= \frac{2\lambda}{\mu^2} \int_0^t (t - \tau_n)^{2k} e^{-2(t-\tau_n)/c} d\tau \\
 &\stackrel{k > -\frac{1}{2}}{=} \frac{2\lambda}{\mu^2} \left(\frac{c}{2}\right)^{2k+1} \gamma\left(2k + 1, \frac{2t}{c}\right), \tag{18}
 \end{aligned}$$

où $\gamma(.,.)$ est la fonction gamma incomplète définie par

$$\gamma(\alpha, x) = \int_0^x e^{-t} t^{\alpha-1} dt \quad (\Re(\alpha) > 0).$$

Les équations (17) et (18) s'obtiennent si $k \in (-1, \infty)$ et $k \in (-\frac{1}{2}, \infty)$ respectivement. La preuve est donnée par la formule 3.381(1) (Gradshteyn et Ryzhik, 2000, p. 342) qui suit :

$$\int_0^u x^{v-1} e^{-\mu x} dx = \mu^{-v} \gamma(v, \mu u) \quad \text{si } v > 0.$$

Par ailleurs, en appliquant la formule 8.352(1) (Gradshteyn et Ryzhik, 2000, p. 890) suivante :

$$\gamma(n + 1, x) = n! \left[1 - e^{-x} \left(\sum_{m=0}^n \frac{x^m}{m!} \right) \right] \quad \text{si } n = 0, 1, \dots,$$

les équations (16) et (17) s'expriment alors, pour des valeurs entières de k et $2k$ respectivement, comme suit :

$$E[X(t)] \stackrel{k \in \{0,1,\dots\}}{=} \frac{\lambda}{\mu} c^{k+1} k! \left[1 - e^{-t/c} \sum_{m=0}^k \frac{1}{m!} \left(\frac{t}{c}\right)^m \right], \quad (19)$$

$$Var[X(t)] \stackrel{2k \in \{0,1,\dots\}}{=} \frac{2\lambda}{\mu^2} \left(\frac{c}{2}\right)^{2k+1} (2k)! \left[1 - e^{-2t/c} \sum_{m=0}^{2k} \frac{1}{m!} \left(\frac{2t}{c}\right)^m \right]. \quad (20)$$

La formule (16) de la covariance de $X(t_1)$ et $X(t_2)$, lorsque $k \geq 0$, est donnée par (voir la Proposition 2.3 de Lefebvre et Guilbault, 2008) :

$$\begin{aligned} & Cov[X(t_1), X(t_2)] \\ &= \frac{2\lambda}{\mu^2} e^{(t_1+t_2)/c} (t_1 t_2)^k \sum_{v=0}^{\infty} \theta_v (-1)^{v+1} \left(\frac{c}{2}\right)^{v+1} v! \left[1 - e^{2t_1/c} \sum_{m=0}^v \frac{(-1)^m}{m!} \left(2t_1 c\right)^m \right]. \end{aligned} \quad (21)$$

Pour $k \in \{0,1,\dots\}$, la valeur de la covariance s'arrête à $2k$.

- Estimation des paramètres du processus de Poisson filtré :

L'estimation des paramètres du modèle, c , k , λ et μ , se fait en considérant que le processus a été en opération pour une durée assez longue dans le temps. Pour ce faire, des expressions des équations précédentes (17), (18) et (21) devraient être calculées à l'état asymptotique du processus, c'est-à-dire quand t tend vers l'infini.

Par le fait que

$$\lim_{t \rightarrow \infty} \gamma(\alpha, x) = \Gamma(\alpha),$$

l'espérance et la variance du processus sont alors données par

$$\lim_{t \rightarrow \infty} E[X(t)] = \frac{\lambda}{\mu} c^{k+1} \Gamma(k+1) \quad \text{si } k \in (-1, \infty), \quad (22)$$

$$\lim_{t \rightarrow \infty} Var[X(t)] = \frac{2\lambda}{\mu^2} \left(\frac{c}{2}\right)^{k+1} \Gamma(2k+1) \quad \text{si } k > -1/2. \quad (23)$$

Le calcul de la covariance de $X(t)$ et $X(t+\delta)$, avec $\delta > 0$, se fait comme suit :

$$\begin{aligned}
Cov[X(t), X(t + \delta)] &= \lambda \int_0^t E[Y_n^2[(t - \tau_n)(t + \delta - \tau_n)]^k e^{-(2t+\delta-2\tau)/c}] d\tau \\
&= \frac{2\lambda}{\mu^2} e^{-\delta/c} \int_0^t [s(s + \delta)]^k e^{-2s/c} ds,
\end{aligned}$$

lorsque $k > -1$. On a :

$$\lim_{t \rightarrow \infty} \int_0^t [s(s + \delta)]^k e^{-2s/c} ds = \frac{1}{\sqrt{\pi}} \left(\frac{\delta c}{2}\right)^{k+\frac{1}{2}} e^{\frac{\delta}{c}} \Gamma(k+1) K_{k+\frac{1}{2}}\left(\frac{\delta}{c}\right).$$

Il suit que :

$$\lim_{t \rightarrow \infty} Cov[X(t), X(t + \delta)] = \frac{2\lambda}{\sqrt{\pi} \mu^2} \left(\frac{\delta c}{2}\right)^{k+\frac{1}{2}} \Gamma(k+1) K_{k+\frac{1}{2}}\left(\frac{\delta}{c}\right), \quad (24)$$

où $K_{k+\frac{1}{2}}(\cdot)$ est la fonction de Bessel modifiée.

Vu son utilité pratique pour les besoins de modélisation, le coefficient de corrélation de $X(t)$ et $X(t + \delta)$ est aussi calculé, et il est donnée, lorsque $k > -\frac{1}{2}$, par

$$\begin{aligned}
\lim_{t \rightarrow \infty} \rho_{X(t), X(t+\delta)} &= \lim_{t \rightarrow \infty} \frac{Cov[X(t), X(t + \delta)]}{\sqrt{Var[X(t)]} \sqrt{Var[X(t + \delta)]}} \\
&= \frac{\frac{2\lambda}{\sqrt{\pi} \mu^2} \left(\frac{\delta c}{2}\right)^{k+\frac{1}{2}} \Gamma(k+1) K_{k+\frac{1}{2}}\left(\frac{\delta}{c}\right)}{\frac{2\lambda}{\mu^2} \left(\frac{c}{2}\right)^{k+1} \Gamma(2k+1)} \\
&= \frac{1}{\sqrt{\pi}} \left(\frac{2\delta}{c}\right)^{k+\frac{1}{2}} \frac{\Gamma(k+1)}{\Gamma(2k+1)} K_{k+\frac{1}{2}}\left(\frac{\delta}{c}\right). \quad (25)
\end{aligned}$$

Les formules asymptotiques précédentes (22), (23) et (25) dépendent des paramètres k et c ; il est impossible de les utiliser pour obtenir la valeur du paramètre k si la valeur du paramètre c est inconnue. En outre, si k est inconnu, il est pratiquement impossible aussi de se servir directement de ces formules asymptotiques. Pour cette raison, une méthode alternative d'estimation du paramètre c a été envisagée (se référer à Lefebvre et Guilbault, 2008).

2.2 Approche d'estimation des paramètres k et c

Une approche pour estimer le paramètre c consiste à trouver une relation linéaire entre le temps que cela prend pour qu'un débit atteigne son maximum après l'arrivée d'un signal et la valeur que peut prendre ce paramètre c . En effet, pour maximiser la fonction de réponse unitaire de l'hydrogramme proposée par Lefebvre et Guilbault (2008) (voir l'équation (3) et la figure A.2)

$$h(t) = (t - \tau)^k e^{-(t-\tau)/c}, \text{ pour } t \geq \tau \geq 0,$$

on pose

$$\frac{d}{dt}h(t) = 0 \Leftrightarrow t - \tau = kc (\geq 0).$$

Si l'on suppose que l'instant initial du signal est $\tau = 0$. Alors, pour $t \geq \tau$ la fonction h définie sur l'intervalle du temps $[0, +\infty)$ admet un maximum en $(t = kc) \in [0, \infty)$.

Intuitivement, le maximum serait atteint lorsque la durée de l'augmentation du débit, du minimum au maximum, équivaut à kc unités du temps. Autrement dit, le maximum doit être observé kc unités de temps après l'arrivée d'un événement à l'instant τ , et ce, si le temps écoulé entre les arrivées des différents événements successifs est assez long afin d'assurer une durée suffisante de la montée du débit suivie par une longue décroissance (Lefebvre et Guilbault, 2008).

À partir de la figure 2.1, en prenant, par exemple, un ensemble à valeurs discrètes dans $[0, +\infty)$ de 20 jours ($t \in \{0, 1, 2, \dots, 20\}$), $k = 1$ et $c = 2$. Le signal étant arrivé à l'instant $\tau = t = 0$ (aujourd'hui, par exemple). Le point maximum de la fonction h serait atteint deux jours après; $kc = 2$ et $h(kc + \tau) = (kc)^k e^{-k} = 2 e^{-1} = 0,7358$. On parle donc de maximum (ou maximum global) de la fonction $h(t)$ parce que, pour tout $T \in \{0, 1, 2, \dots, 20\}$, on a : $h(T) \leq h(2)$, et de maximum local pour le processus $X(t)$.

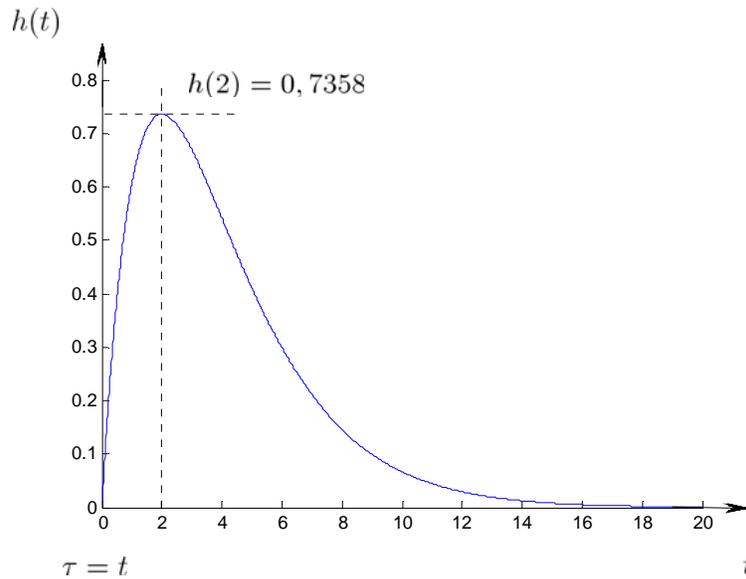


Figure 2.1. La fonction de réponse unitaire de l'hydrogramme

$$h(t) = (t - \tau)^k e^{-(t-\tau)/c}$$

Cependant, l'expression $t - \tau = kc$ ne permet pas d'obtenir une valeur précise du paramètre c sans fixer le paramètre k qui peut prendre n'importe quelle valeur dans l'intervalle $[0, \infty)$. Maintenant, si l'on suppose que le temps moyen nécessaire pour que le débit augmente à son maximum vaut kc unités du temps, c'est-à-dire que $\overline{(t - \tau)} = kc$, il devient facile de calculer les coefficients de corrélation théoriques (équation 25) en remplaçant le paramètre c par $\overline{(t - \tau)}/k$; les coefficients dépendront donc seulement des valeurs de δ et k , du moment que l'on peut calculer en pratique la moyenne arithmétique des temps d'augmentation du débit.

En donnant des valeurs réelles positives au paramètre k , nous obtenons des valeurs du coefficient de corrélation théorique ρ pour $\delta = 1, 2, \dots$. Le critère utilisé pour estimer le paramètre k est de calculer la somme des différences absolues entre les coefficients de corrélation théoriques et empiriques sur un horizon $\delta > 0$, et le point qui minimise la valeur de cette somme serait l'estimation du paramètre k .

2.3 Prévision

Cette section illustre l'utilisation du processus de Poisson filtré aux fins de prévision, en se basant sur l'espérance conditionnelle du débit prévu à l'instant $t + 1$ étant donné ses valeurs un ou deux jours d'avance. L'évaluation des prévisions est considérée dans le cas du modèle lorsque $k = 1$ et $k = 0$.

2.3.1 Modèle avec $k = 1$

Lorsque $k = 1$, il est possible de calculer l'espérance conditionnelle du débit d'un cours d'eau à l'instant $t + 1$, étant donnée l'historique du processus aux instants $t - 1$ et t . En effet, l'estimateur du débit à l'instant $t + 1$; $\widehat{X}(t + 1)$ est donné par

$$\begin{aligned} E [X(t + 1) | X(t), X(t - 1)] &= E \left[\sum_{n:0 \leq \tau_n \leq t+1} Y_n (t + 1 - \tau_n)^k e^{-(t+1-\tau_n)/c} \middle| X(t), X(t - 1) \right] \\ &= e^{-\frac{1}{c}} E \left[\sum_{n:0 \leq \tau_n \leq t} Y_n (t + 1 - \tau_n)^k e^{-(t-\tau_n)/c} \middle| X(t), X(t - 1) \right] \\ &\quad + E \left[\sum_{n:t \leq \tau_n \leq t+1} Y_n (t + 1 - \tau_n)^k e^{-(t+1-\tau_n)/c} \middle| X(t), X(t - 1) \right]. \end{aligned}$$

Lorsque $k > 0$, l'estimateur est équivalent à

$$\begin{aligned} E [X(t + 1) | X(t), X(t - 1)] &= e^{-\frac{1}{c}} E \left[\sum_{n:0 \leq \tau_n \leq t} Y_n (t + 1 - \tau_n)^k e^{-(t-\tau_n)/c} \middle| X(t), X(t - 1) \right] \\ &\quad + E[X(1)] \\ &= e^{-\frac{1}{c}} x_k + E[X(1)], \end{aligned}$$

où

$$x_k := e^{-\frac{1}{c}} E \left[\sum_{c:0 \leq \tau_n \leq t} Y_n (t + 1 - \tau_n)^k e^{-(t-\tau_n)/c} \middle| X(t), X(t - 1) \right].$$

Si $k = 1$, alors

$$\begin{aligned}
x_1 &= E \left[\sum_{n:0 \leq \tau_n \leq t} Y_n (t+1 - \tau_n) e^{-(t-\tau_n)/c} \middle| X(t), X(t-1) \right] \\
&= E \left[\sum_{n:0 \leq \tau_n \leq t} Y_n (t - \tau_n) e^{-(t-\tau_n)/c} \middle| X(t), X(t-1) \right] \\
&\quad + E \left[\sum_{n:0 \leq \tau_n \leq t} Y_n e^{-(t-\tau_n)/c} \middle| X(t), X(t-1) \right] \\
&= X(t) + y,
\end{aligned}$$

tel que

$$y := E \left[\sum_{n:0 \leq \tau_n \leq t} Y_n e^{-(t-\tau_n)/c} \middle| X(t), X(t-1) \right]$$

exprime l'espérance conditionnelle du processus, lorsque $k = 0$, connaissant son état aux instants t et $t - 1$.

La quantité x_1 peut aussi être calculée d'une autre façon :

$$\begin{aligned}
x_1 &= E \left[\sum_{n:0 \leq \tau_n \leq t} Y_n (t-1 - \tau_n + 2) e^{-(t-\tau_n)/c} \middle| X(t), X(t-1) \right] \\
&= 2 E \left[\sum_{n:0 \leq \tau_n \leq t} Y_n e^{-(t-\tau_n)/c} \middle| X(t), X(t-1) \right] \\
&\quad + e^{-\frac{1}{c}} E \left[\sum_{n:0 \leq \tau_n \leq t-1} Y_n (t-1 - \tau_n) e^{-(t-\tau_n-1)/c} \middle| X(t), X(t-1) \right] \\
&\quad + E \left[\sum_{n:t-1 \leq \tau_n \leq t} Y_n (t-1 - \tau_n) e^{-(t-\tau_n)/c} \middle| X(t), X(t-1) \right].
\end{aligned}$$

Soit

$$z := E \left[\sum_{n:t-1 \leq \tau_n \leq t} Y_n (t-1 - \tau_n) e^{-(t-\tau_n)/c} \middle| X(t), X(t-1) \right].$$

Il s'ensuit que

$$x_1 = 2y + e^{-\frac{1}{c}} X(t-1) + z.$$

Si l'on remplace y par $x_1 - X(t)$ dans cette dernière équation, la valeur de x_1 vaudrait donc

$$x_1 = 2X(t) - e^{-\frac{1}{c}} X(t-1) - z.$$

Il reste à évaluer la quantité z , qui peut être obtenue par indépendance comme suit :

$$\begin{aligned} z &:= E \left[\sum_{n:t-1 \leq \tau_n \leq t} Y_n (t-1-\tau_n) e^{-(t-\tau_n)/c} \middle| X(t), X(t-1) \right] \\ &\stackrel{ind.}{=} E \left[\sum_{n:t-1 \leq \tau_n \leq t} Y_n \middle| X(t), X(t-1) \right] \\ &\quad \times E \left[\sum_{n:t-1 \leq \tau_n \leq t} (t-1-\tau_n) e^{-(t-\tau_n)/c} \middle| X(t), X(t-1) \right]. \end{aligned}$$

Afin de calculer approximativement z , nous supposons que dans l'intervalle de temps $(t-1, t]$, au plus 1 événement s'est produit : soit 0 ou 1 signal survient dans une unité de temps. S'il n'y a pas eu de signal qui fait augmenter le débit à l'instant τ_n dans l'intervalle $(t-1, t]$, le débit à l'instant t serait généralement inférieur au débit précédent à l'instant $t-1$, c'est-à-dire que $X(t) < X(t-1)$. Alors que si un signal se produit dans l'intervalle $(t-1, t]$, le débit à l'instant $t-1$ augmenterait, c'est-à-dire que $X(t) \geq X(t-1)$. L'événement en question s'est produit en moyenne à l'instant $t - \frac{1}{2}$, et son effet est donné par

$$Y_n \left[t - \left(t - \frac{1}{2} \right) \right] e^{-\frac{(t-(t-\frac{1}{2}))}{c}} = \frac{1}{2} Y_n e^{-\frac{1}{2c}}.$$

En outre, pour prendre en considération une éventuelle diminution dans cette période $(t-1, t]$, nous tenons compte du nombre d'unités de temps de diminution du débit dans l'intervalle $(0, t]$, et puis nous calculons la moyenne p de la baisse relative du débit par unité de temps. Il s'ensuit que le premier terme de z serait donné par

$$E \left[\sum_{n:t-1 \leq \tau_n \leq t} Y_n \middle| X(t), X(t-1) \right] \simeq 2e^{\frac{1}{2c}} \left\{ [X(t) - X(t-1)] + pX(t-1) \right\},$$

si $X(t) \geq X(t-1)$.

La moyenne p traduit la pente de la baisse du débit dans la période $(t-1, t]$, et le terme $2e^{\frac{1}{2c}}$ est interprété comme étant la force du signal dans cette période.

Puisqu'il est supposé qu'un seul événement s'est produit dans l'intervalle $(t-1, t]$, nous avons $\tau_n \sim U(t-1, t]$; il est donc possible sous cette hypothèse d'évaluer le second terme de z comme suit

$$\begin{aligned} & E \left[\sum_{n:t-1 \leq \tau_n \leq t} (t-1-\tau_n) e^{-(t-\tau_n)/c} \middle| X(t), X(t-1) \right] \\ &= \int_{t-1}^t (t-1-s) e^{-(t-s)/c} \cdot 1 ds \\ &= (t-1-s) c e^{-(t-s)/c} \Big|_{t-1}^t - \int_{t-1}^t c e^{-(t-s)/c} ds \\ &= -c + c^2 - c^2 e^{-\frac{1}{c}} \\ &:= f(c). \end{aligned}$$

Par déduction,

$$\begin{aligned} z &\simeq 2e^{\frac{1}{2c}} f(c) \left\{ [X(t) - X(t-1)] + pX(t-1) \right\} \mathbf{1}_{\{X(t) \geq X(t-1)\}} \\ &\simeq 2e^{\frac{1}{2c}} f(c) \left\{ X(t) - (1-p)X(t-1) \right\} \mathbf{1}_{\{X(t) \geq X(t-1)\}}, \end{aligned}$$

et

$$x_1 \simeq 2X(t) - e^{-\frac{1}{c}} X(t-1) - 2e^{\frac{1}{2c}} f(c) \left\{ X(t) - (1-p)X(t-1) \right\} \mathbf{1}_{\{X(t) \geq X(t-1)\}}.$$

Finalement, l'estimateur de la valeur prévue du débit à l'instant $t+1$ se simplifie à

$$\begin{aligned}
\widehat{X(t+1)} &= E[X(t+1)|X(t),X(t-1)] \\
&= e^{-\frac{1}{c}} x_1 + E[X(1)] \\
&\simeq e^{-\frac{1}{c}} \left[2X(t) - e^{-\frac{1}{c}} X(t-1) \right. \\
&\quad \left. - 2e^{\frac{1}{2c}} f(c) \left\{ X(t) - (1-p)X(t-1) \right\} \mathbf{1}_{\{X(t) > X(t-1)\}} \right] \\
&\quad + E[X(1)].
\end{aligned} \tag{26}$$

2.3.2 Modèle avec $k = 0$

Dans le cas du modèle avec $k = 0$, l'espérance conditionnelle de $X(t+1)$, étant donnée l'historique du processus dans l'intervalle $[0, t]$, donne l'estimateur de prévision suivant (Lefebvre et Guilbault, 2008) :

$$\begin{aligned}
E[X(t+1)|X(t)] &= E \left[\sum_{n:0 \leq \tau_n \leq t+1} Y_n e^{-(t+1-\tau_n)/c} \middle| X(t) \right] \\
&= E \left[\sum_{n:0 \leq \tau_n \leq t} Y_n e^{-(t+1-\tau_n)/c} + \sum_{n:t \leq \tau_n \leq t+1} Y_n e^{-(t+1-\tau_n)/c} \middle| X(t) \right] \\
&= e^{-1/c} E \left[\sum_{n:0 \leq \tau_n \leq t} Y_n e^{-(t-\tau_n)/c} \middle| X(t) \right] \\
&\quad + E \left[\sum_{n:t \leq \tau_n \leq t+1} Y_n e^{-(t+1-\tau_n)/c} \middle| X(t) \right] \\
&= e^{-1/c} E[X(t)|X(t)] + E[X(1)] \\
&= e^{-1/c} X(t) + E[X(1)] \\
&= e^{-1/c} X(t) + \frac{\lambda c}{\mu} (1 - e^{-1/c}).
\end{aligned} \tag{27}$$

En effet, lorsque $k = 0$, la valeur de $X(t+1)$ ne dépend que de $X(t)$. Lorsque $k = 1$, l'estimateur calculé du débit à l'instant $t+1$ dépend de la valeur du débit aux instants t et $t-1$, ce qui donne une information de plus pour prévoir la valeur du débit à l'instant $t+1$.

CHAPITRE 3 :

APPLICATION ET RÉSULTATS

Une application des formules obtenues précédemment sur des données réelles, tirées des relevés hydrographiques de deux fleuves, va permettre de juger pratiquement des résultats théoriques du modèle. Les stations Montague (No. 014385) du fleuve Delaware au New Jersey et Green Island (No. 013580) du fleuve Hudson à New York serviront à cette fin. Pour la période de temps du 1^{er} octobre 2002 au 30 septembre 2003, les statistiques correspondant à la moyenne, l'écart-type du débit et le coefficient de corrélation entre les débits pour deux jours consécutifs sont résumés dans le tableau 2.1, et la figure 2.2 illustre la trajectoire du débit pour les deux fleuves.

Tableau 2.1. Caractéristiques du débit (en m³/s)

	Delaware River	Hudson River
Moyenne \bar{x}	7517,75342	15482,9863
Écart-type σ	6358,28631	11621,5812
Coef. Corrélation r	0,90121739	0,93868413

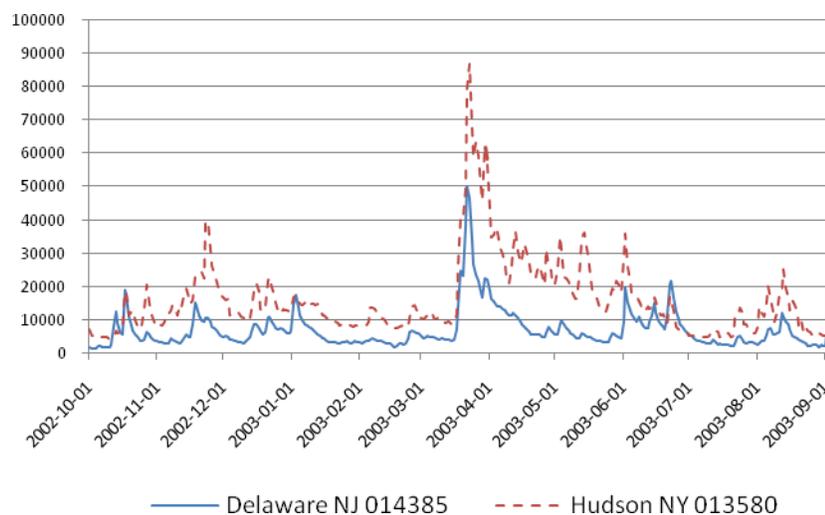


Figure 2.2. Hydrogramme du débit (en m³/s) des fleuves Delaware et Hudson (01/10/2002 au 30/09/2003)

3.1 Modèle avec la valeur du paramètre k estimée

Avec l'utilisation de l'approche du critère du coefficient de corrélation et la méthode alternative d'estimation du paramètre c , une relation de dépendance entre les paramètres c et k est obtenue, ce qui permet par la suite de calculer le coefficient de corrélation $\rho(\delta)$ en fonction de valeurs réelles positives du paramètre k . La valeur du paramètre c lorsque $k = 1$ vaut approximativement en moyenne 2,1923 jours pour que le débit passe du minimum au maximum pour le fleuve Delaware. Comme l'illustre la courbe de la figure 2.3, la valeur du paramètre k optimale qui minimise la somme des différences absolues entre le coefficient de corrélation théorique $\rho(\delta)$ et empirique $r(\delta)$ (sur un horizon δ de 7 jours) vaut 0,75.

Sur 52 signaux ou impulsions produits, selon un processus de Poisson, durant l'année d'étude du fleuve Delaware, la moyenne et l'écart-type temporels séparant deux impulsions consécutives (le début de l'augmentation du débit), sont : $\bar{T} \simeq 8,0392$, $s_T \simeq 4,1374$, ce qui implique que le temps T entre deux signaux consécutifs qui font augmenter le débit n'a pas une distribution exponentielle puisque la moyenne et l'écart-type de cette distribution doivent être égaux. Bien que ce constat signifie que le processus de Poisson filtré n'est peut-être pas le modèle valable théoriquement pour les débits du fleuve Delaware, il est toutefois parfois possible d'appliquer une fonction qui transforme la variable T à une variable aléatoire exponentielle (Lefebvre, 2004). Nous nous contentons dans ce travail de supposer que le modèle du processus de Poisson filtré est valable pour les observations dont nous disposons.

Si l'on calcule la somme des différences absolues entre les coefficients de corrélation théorique $\rho(\delta)$ et empirique $r(\delta)$ sur un horizon δ plus élevé, nous constatons alors qu'au fur et à mesure que δ prend des valeurs plus élevées, la valeur du paramètre k tend à diminuer pour à la fin se stabiliser sur une valeur constante. Effectivement, la valeur du paramètre k se stabilise à 0,63 lorsque les différences absolues des coefficients de corrélation sont calculées sur un horizon de plus de 20 jours pour le fleuve Delaware

(figure 2.4). Cela est, statistiquement parlant, évident du moment que la fonction d'autocorrélation décroît exponentiellement et/ou s'annule après un certain décalage dans le temps. En outre, la valeur du coefficient de corrélation empirique $r(\delta)$ convergera vers la valeur du coefficient de corrélation théorique $\rho(\delta)$; ainsi l'écart entre ces deux coefficients tend à s'annuler dans le temps, c'est-à-dire pour δ très grand.

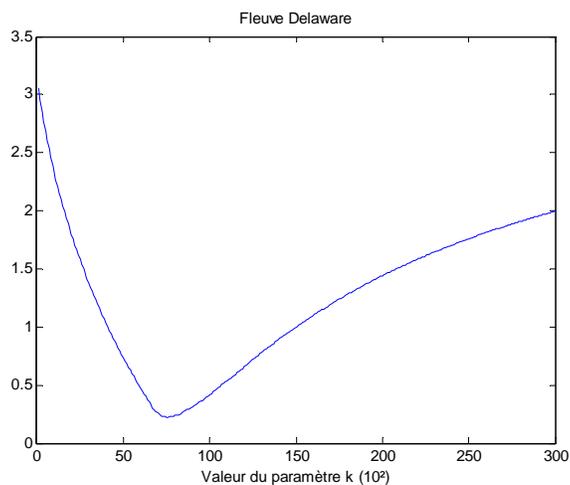


Figure 2.3. Les valeurs de $\sum_{\delta=1}^7 |\rho(\delta) - r(\delta)|$ en fonction des valeurs du paramètre k prises de 0,01 à 3.

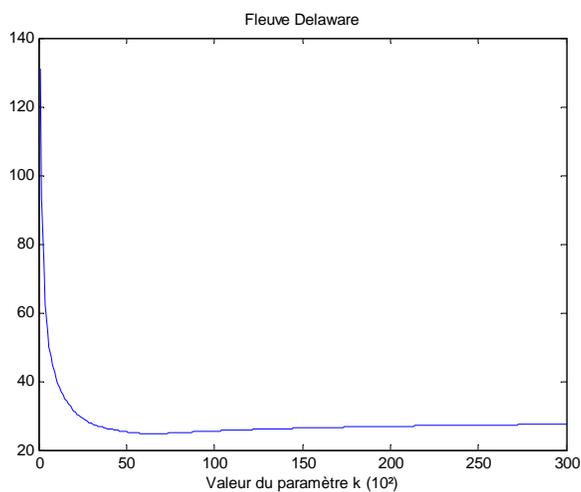


Figure 2.4. Les valeurs de $\sum_{\delta>20} |\rho(\delta) - r(\delta)|$ en fonction des valeurs du paramètre k prises de 0,01 à 3.

Le tableau 2.2 qui suit montre les calculs des coefficients de corrélation théorique et empirique avec la valeur estimée du paramètre k ; la valeur du paramètre c vaut \bar{T}/\hat{k} .

Tableau 2.2 : Écart entre les coefficients de corrélation théorique et empirique des débits $X(t)$ et $X(t + \delta)$ pour $\delta = 1, 2, \dots, 7$, $\hat{k} = 0,63$ et $\hat{c} \simeq 3,4798$ (Delaware).

δ	1	2	3	4	5	6	7
$r(\delta)$	0,9012	0,7470	0,6123	0,5102	0,4247	0,3464	0,2888
$\rho(\delta)$	0,9390	0,8259	0,7022	0,5844	0,4789	0,3883	0,3120
$ \rho(\delta) - r(\delta) $	0,0378	0,0789	0,0899	0,0742	0,0542	0,0419	0,0232

En conclusion, le modèle apparaît estimer très bien la valeur du coefficient de corrélation théorique $\rho(\delta)$. En fait, l'erreur d'estimation la plus élevée est de 14,7% (au 3^e jour suivant l'observation la plus récente), ce qui traduit le fait que le modèle peut être aussi bien approprié pour des prévisions sur une longue période, soit au-delà de sept jours, que sur une brève période. Il faut rappeler que la somme des différences absolues des coefficients de corrélation pour un horizon assez long atteint son minimum lorsque $k = 0,63$. Par ailleurs, la différence des deux coefficients de corrélation théorique et empirique $\rho(\delta) - r(\delta)$ est plus petite qu'avec les modèles avec $k = 0$ ou $k = 1$ comme il sera montré ci-après.

3.2 Comparaison avec les modèles lorsque $k = 0$ et $k = 1$

Pour pouvoir estimer les paramètres du modèle lorsque $k = 0$, il est nécessaire d'estimer la valeur du paramètre c afin d'être capable, par la suite, de calculer les valeurs des paramètres λ et μ en tenant compte des formules asymptotiques précédentes. Le critère crédible utilisé pour calculer la valeur du paramètre c est de supposer l'égalité de la limite asymptotique du coefficient de corrélation théorique de pas 1 : $\rho(\delta = 1)$ et la valeur du coefficient de corrélation empirique de pas 1 : $r(\delta = 1)$.

À partir de l'équation (25), on déduit que

$$\lim_{t \rightarrow \infty} \rho_{X(t), X(t+\delta)} = \begin{cases} e^{-\delta/c} & \text{si } k = 0, \\ \left(\frac{\delta+c}{c}\right) e^{-\delta/c} & \text{si } k = 1. \end{cases}$$

Alors

$$\lim_{t \rightarrow \infty} \rho_{X(t), X(t+1)} = e^{-1/c} \quad \text{lorsque } k = 0.$$

$$e^{-1/c} \simeq 0,9012 \Rightarrow \hat{c} \simeq 9,6128.$$

Par conséquent, les équations (22) et (23) lorsque $k = 0$ sont données par

$$\lim_{t \rightarrow \infty} E[X(t)] = \frac{\lambda c}{\mu},$$

$$\lim_{t \rightarrow \infty} Var[X(t)] = \frac{\lambda c}{\mu^2}.$$

D'où

$$\frac{\lambda(9,6128)}{\mu} \simeq 7517,75 \Rightarrow \hat{\mu} \simeq 1,86 \times 10^{-4},$$

et
$$\frac{\lambda(9,6128)}{\mu^2} \simeq (6358,29)^2 \Rightarrow \hat{\lambda} \simeq 0,1454.$$

- Le modèle avec $k = 0$ obtenu pour le fleuve Delaware est alors

$$X(t) = \sum_{n=1}^{N(t)} Y_n e^{-(t-\tau_n)/9,6128} \quad \text{pour } t \geq 0 \quad (X(t) = 0 \text{ si } N(t) = 0)$$

où $\{N(t), t \geq 0\}$ est un processus de Poisson de taux 0,1454 et les Y_n sont des v.a. i.i.d. de distribution exponentielle de paramètre $1,86 \times 10^{-4}$.

- Dans le cas du modèle avec $k = 1$ pour le fleuve Delaware, l'estimation du paramètre c est $\hat{c} \simeq 2,1923$, et le modèle s'exprime comme suit :

$$X(t) = \sum_{n=1}^{N(t)} Y_n(t - \tau_n) e^{-(t-\tau_n)/2,1923} \quad \text{pour } t \geq 0 \quad (X(t) = 0 \text{ si } N(t) = 0)$$

où $\{N(t), t \geq 0\}$ est un processus de Poisson de taux 0,2908 et les Y_n sont des v.a. i.i.d. de distribution exponentielle de paramètre $1,86 \times 10^{-4}$.

Les tableaux 2.3 et 2.4 ci-après montrent les valeurs des différences des coefficients de corrélation théoriques et empiriques du modèle lorsque $k = 0$ et $k = 1$ respectivement.

Tableau 2.3 : Écart entre les coefficients de corrélation théorique et empirique des débits $X(t)$ et $X(t + \delta)$ pour $\delta = 1, 2, \dots, 7, k = 0$ et $\hat{c} \simeq 9,613$ (Delaware).

δ	1	2	3	4	5	6	7
$r(\delta)$	0,9012	0,7470	0,6123	0,5102	0,4247	0,3464	0,2888
$\rho(\delta)$	0,9012	0,9967	0,7319	0,6596	0,5944	0,5357	0,4828
$ \rho(\delta) - r(\delta) $	0	0,2497	0,1196	0,1494	0,1697	0,1893	0,1940

Tableau 2.4 : Écart entre les coefficients de corrélation théorique et empirique des débits $X(t)$ et $X(t + \delta)$ pour $\delta = 1, 2, \dots, 7, k = 1$ et $\hat{c} \simeq 2,1923$ (Delaware).

δ	1	2	3	4	5	6	7
$r(\delta)$	0,9012	0,7470	0,6123	0,5102	0,4247	0,3464	0,2888
$\rho(\delta)$	0,9228	0,7680	0,6028	0,4556	0,3353	0,2421	0,1721
$ \rho(\delta) - r(\delta) $	0,0216	0,021	0,0095	0,0546	0,0894	0,1043	0,1167

La différence absolue des coefficients de corrélation $|\rho(\delta) - r(\delta)|$ des débits $X(t)$ et $X(t + \delta)$ devrait décroître dans le temps et le modèle serait valable pour prévoir le débit du fleuve à l'horizon δ si l'erreur relative du coefficient de corrélation tend à s'annuler à cet horizon. En guise de comparaison de ces modèles avec le modèle obtenu en estimant

le paramètre k ; il peut être conclu de ces résultats de modélisation que la qualité du modèle semble bien meilleure avec $k = 1$ qu'avec $k = 0$, mais il reste que les résultats de ce dernier sont obtenus avec une valeur fixe du paramètre c calculée à partir du coefficient de corrélation de pas égal à une unité du temps, par convention, car il est le plus fiable statistiquement et qu'à cet horizon-là le modèle devrait être parfaitement réalisable. Par ailleurs, il n'y a pas une très grande différence entre le modèle avec $\hat{k} = 0,63$ et le modèle avec $k = 1$. On voit dans le graphique de la figure 2.3, que la courbe de la somme des différences absolues semble stationnaire à partir du point minimum $0,63$ et que, de ce fait, les valeurs de $\sum_{\delta > 20} |\rho(\delta) - r(\delta)|$ à ce point ne sont pas très élevées ou très différentes de celles au point $k = 1$, ce qui fait que l'ajustement du modèle à moyen terme paraît aussi bon avec $k = 1$ qu'avec $\hat{k} = 0,63$. Mais si on considère un ajustement à court terme, soit à sept jours (voir figure 2.2), la différence des valeurs de $\sum_{\delta=1}^7 |\rho(\delta) - r(\delta)|$ étant importante entre les deux points $0,75$ et 1 , le modèle avec la valeur du paramètre k estimée ($\hat{k} = 0,75$) s'ajuste donc mieux aux observations que le modèle quand k vaut 1 .

3.3 Stabilité du paramètre k

Dans le but de savoir si l'ensemble des valeurs possibles du paramètre k est défini sur un intervalle donné, nous vérifierons la variabilité de ce paramètre en le calculant sur différentes périodes de débits pour d'autres échantillons de données. Le tableau 2.5 résume les valeurs du paramètre k sur différents périodes annuelles de débits; une variabilité du paramètre est immédiatement remarquée pour le fleuve Delaware, contrairement, plus au moins, au fleuve Hudson. Une valeur élevée du paramètre k , dépassant la valeur 2 dans les années 2003-2004 et 2005-2006, est observée pour le fleuve Delaware; même chose dans l'année 2005-2006 pour le fleuve Hudson. Il ne doit y avoir que la nature des données qui influence ces variations, voire le mouvement ou la dynamique du débit dans chaque période, ce qui explique la valeur de k qui est légèrement supérieure à 1 sur deux années consécutives (2002-2004 et 2004-2006) pour le fleuve Delaware. Cependant, la valeur $1,59$ du paramètre k pour le fleuve Hudson

n'apparaît pas trop influencer la variabilité du paramètre sur une période de deux ans (2004-2006 où $k = 0,63$) puisque la valeur est suffisamment inférieure à 1. En examinant le mouvement des fleuves de plus près (voir les figures 2.5 et 2.6), nous constatons que deux crues se sont produites au fleuve Delaware à la période de l'automne et de l'été de 2004 et 2006 respectivement, alors que par opposition aux autres années, le débit fluctue plus aux périodes printanière et hivernale. En outre, la crue hivernale de 2005 n'a pas trop d'impact sur la valeur du paramètre k , peut-être parce que cette crue a été suivie par une stagnation du débit, ce qui a fait alors compenser la moyenne des variations. Par ailleurs, le fleuve Hudson est distingué par l'arrivée de crues très élevées à la période 2005-2006, ce qui a conduit à une valeur de k qui vaut 1,59, soit inférieure à 2, de sorte que la dynamique générale du fleuve est bien stationnaire à l'exception de la crue estivale qui est bien distinguée par rapport aux autres années.

Tableau 2.5 : Variabilité du paramètre k .

	<u>Fleuve Delaware</u>	<u>Fleuve Hudson</u>
	$k(k_{(\delta \rightarrow \infty)})$	$k(k_{(\delta \rightarrow \infty)})$
2002 – 2003	0,75	0,23
2003 – 2004	2,56 (1,97)	0,70
2004 – 2005	0,68	0,31
2005 – 2006	2,37	1,59
2006 – 2007	(1,96)0,40	0,34
2007 – 2008	0,44	0,33
2002 – 2004	1,54	0,42
2004 – 2006	1,15	0,63
2006 – 2008	0,41	0,33
2003 – 2006	1,31	0,63
2002 – 2008	0,88	0,43

Remarque : la valeur du paramètre k entre parenthèses est sa valeur limite (lorsque $\delta \rightarrow \infty$).

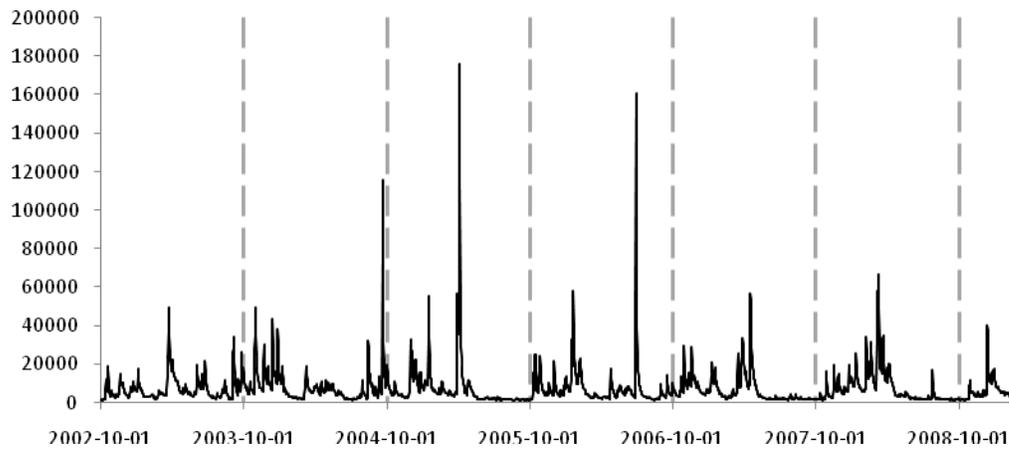


Figure 2.5. Hydrogramme du débit (en m³/s) du fleuve Delaware de 2002 à 2009.

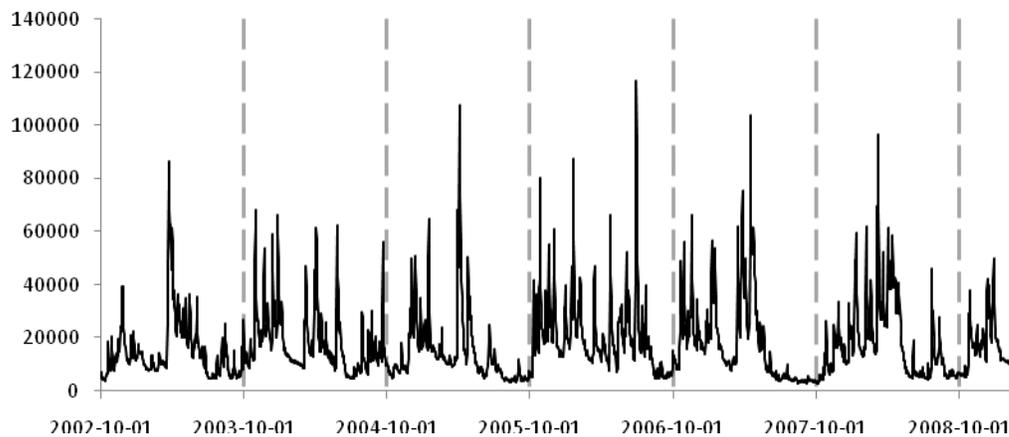


Figure 2.6. Hydrogramme du débit (en m³/s) du fleuve Hudson de 2002 à 2009.

3.4 Capacité prévisionnelle du modèle

Le calcul des coefficients de corrélation théorique et empirique a servi de premier critère d'évaluation de la qualité d'estimation du modèle. Par ailleurs, le second critère primordial de l'adéquation du modèle comme spécification formelle de la dynamique d'un fleuve est son utilité à prévoir le débit du lendemain, voire le débit futur.

3.4.1 Cas lorsque $k = 1$

L'application qui suit se fait sur le fleuve Delaware, en utilisant les données sur les débits du 1^{er} octobre 2002 au 30 septembre 2003.

Le débit a diminué 247 fois pendant cette année, et la diminution moyenne par jour a été d'environ 10,47% ($= p$). Avec $\hat{c} \simeq 2,1923$, le calcul de l'estimateur du débit $\widehat{X}(t+1)$ à l'instant $t+1$ (équation 26) se fait comme suit :

$$E[X(1)] = \frac{\lambda c}{\mu} (c - e^{-1/c}(1+c)) = \frac{7517,75}{2,1923} (0,1689).$$

Alors

$$\begin{aligned} \widehat{X}(t+1) &\simeq e^{-\frac{1}{\hat{c}}} \left[2X(t) - e^{-\frac{1}{\hat{c}}} X(t-1) \right] \\ &\quad - 2e^{\frac{1}{2\hat{c}}} f(c) \left\{ X(t) - (1-p)X(t-1) \right\} \mathbf{1}_{\{X(t) > X(t-1)\}} \\ &\quad + E[X(1)] \\ &\simeq 0,6337 \left[2X(t) - 0,6337 X(t-1) \right] \\ &\quad + 0,6876 \left\{ X(t) - 0,895 X(t-1) \right\} \mathbf{1}_{\{X(t) > X(t-1)\}} \\ &\quad + 580,6924. \end{aligned}$$

Ainsi, si $X(t) \geq X(t+1)$, le terme

$$z \simeq 0,6876 \left\{ X(t) - 0,895 X(t-1) \right\}$$

sera ajouté à l'estimateur du débit $\widehat{X}(t+1)$.

Étant donné que les données sont disponibles pour la période de prévision, le calcul des débits prévus se fait selon la méthode statique de prévision qui tient compte des valeurs

réellement observées des débits; le débit du lendemain serait donc estimé à partir du débit observé aujourd'hui et hier. Avec l'application de cet estimateur, nous effectuons des prévisions sur les quatre premiers mois de l'année 2004, soit du 21 janvier au 30 avril. L'erreur absolue moyenne *mae* (*Mean Absolute Error*), soit la moyenne de la valeur absolue des 101 erreurs de prévision (21/01/2004-30/04/2004) et son écart-type *sdae* (*Standard Deviation of Absolute Error*) sont

$$mae = 717,5432 \quad \text{et} \quad sdae = 1200,2266$$

$$\text{avec } e = X(t+1) - \widehat{X}(t+1).$$

Le modèle de régression linéaire obtenu à partir du même échantillon d'observations est

$$\widehat{X}(t+1) \simeq 1024,89 + 1,2122X(t) - 0,3457X(t-1)$$

$$\text{avec } mae = 808,6454 \quad \text{et} \quad sdae = 948,1041$$

Un autre indicateur d'écart entre les deux modèles de prévision est l'erreur absolue moyenne en pourcentage *mape* (*Mean Absolute Percentage Error*) qui mesure la précision de l'ajustement du modèle aux données, soit

$$mape = 11,29\% \text{ pour l'estimateur du débit } \widehat{X}(t+1),$$

$$mape = 16,36 \text{ pour le modèle de régression,}$$

et ce, sur les 101 erreurs de prévision correspondantes.

Par conséquent, à partir de ce dernier indicateur, le modèle issu du processus de Poisson filtré paraît s'ajuster plus aux vraies données que le modèle de régression linéaire pour prévoir le débit du fleuve Delaware; l'écart d'erreur en pourcentage entre les deux modèles est d'environ 5%. Par contre, en prenant compte de l'erreur quadratique moyenne *mse* (*Mean Square Error*) ou de sa racine carrée *rmse* (*Root Mean Squared Error*),

$mse = 1\,941\,149,2155$, $rmse = 1\,393,2513$ pour l'estimateur du débit $\widehat{X}(t+1)$,

$mse = 1\,543\,908,7757$, $rmse = 1\,242,5413$ pour le modèle de régression.

Nous notons que le mse est un peu plus élevé pour le modèle de prévision. En fait, le mse incorpore la variance et le biais des erreurs de prévision, ce qui permet un arbitrage entre les deux. Alors que le mae pénalise moins les grandes erreurs de prévision en donnant moins de poids aux observations extrêmes, ce n'est pas le cas avec le mse . Quant au $rmse$, il permet d'avoir une unité de mesure comparable à celle des observations. Étant donné que les débits prévus dans cette période des premiers quatre mois de l'année 2004 sont beaucoup moins élevés par rapport aux autres périodes de l'année et que, par conséquent, les erreurs de prévision observées sont également plus faibles, alors les indicateurs pertinents pour comparer les prévisions seraient le mae et le $mape$. Ces derniers sont moins élevés pour l'estimateur du débit $\widehat{X}(t+1)$ que pour le modèle de régression linéaire. Notre modèle est donc plus performant que la régression linéaire.

3.4.2 Cas lorsque $k = 0$

Dans le cas du modèle avec $k = 0$, pour le même échantillon de données du fleuve Delaware, l'estimateur du débit $\widehat{X}(t+1)$ (équation 27) est donné par

$$E[X(t+1)|X(t)] \simeq 0,9012X(t) + 742,7537$$

avec $mae = 837,3837$

$$sdae = 1\,022,0397$$

$$mape = 15,05\%$$

$$mse = 1\,735\,434,3390, rmse = 1\,317,3588.$$

Le modèle de régression linéaire est donné par

$$\widehat{X}(t+1) \simeq 771,1605 + 0,9021X(t)$$

avec $mae = 858,3267$
 $sdae = 1\,009,6543$
 $mape = 15,71\%$
 $mse = 1\,746\,033,4022, rmse = 1\,321,3756$.

Les résultats du modèle lorsque $k = 0$ sont comparables avec ceux du modèle de régression linéaire. Ce qui traduit sa puissance prévisionnelle à court terme (en prévoyant le débit du lendemain à partir de l'observation la plus récente). Cependant, le modèle avec $k = 1$ semble battre le modèle classique du processus de Poisson filtré autant que le modèle de régression correspondant. En fait, ce modèle donne les valeurs les plus petites de mae et $mape$. De plus, 66 fois sur 101 la valeur absolue de l'erreur de prévision de l'estimateur a été inférieure à celle obtenue en utilisant le modèle avec $k = 0$ et la médiane des valeurs absolues des 101 erreurs est égale à 585,87 dans le cas de l'estimateur du débit $\widehat{X}(t+1)$ avec $k = 0$, et à 352,76 avec $k = 1$. Enfin, cela témoigne du fait que le modèle du processus de Poisson filtré serait un bon modèle aux fins de prévision.

Nous voulons maintenant vérifier si l'on obtient des résultats semblables en prévoyant sur une autre période de même longueur que la précédente, soit 101 jours de prévision, du 1^{er} octobre 2003 au 9 janvier 2004 (101 données qui succèdent la période d'estimation du modèle).

Les résultats des calculs obtenus pour chaque critère d'écart entre valeurs observées et prévues du débit sont présentés dans ce qui suit sans le terme z . En fait, l'estimateur du débit $\widehat{X}(t+1)$, avec le terme z , amplifie les valeurs de prévision lorsque le débit est en période de croissance (cela est bien clair graphiquement, voir figures 2.7 à 2.10), cela est dû à l'hypothèse qu'un événement s'est produit durant l'intervalle de temps $(t-1, t]$, ce qui implique évidemment une augmentation du débit. Ce terme z était censé améliorer les prévisions, mais en pratique il les rend moins bonnes.

Le tableau 2.6 résume les résultats obtenus sur la période 01/10/2003 - 09/01/2004, pour le processus de Poisson filtré et pour le modèle de régression linéaire. Les résultats, cette fois-ci, favorisent la régression linéaire plutôt que le processus de Poisson filtré lorsque $k = 1$, quoiqu'on peut prétendre que les résultats sont presque similaires mais avec un degré de plus pour le modèle autorégressif. Le modèle lorsque $k = 0$ demeure comparable avec la régression linéaire correspondante. Le point le plus important à remarquer est que le modèle avec $k = 1$ donne des résultats meilleurs que le modèle classique lorsque $k = 0$ (voir figures 2.7 et 2.8).

En refaisant les mêmes prévisions pour le processus de Poisson filtré sur le fleuve Hudson (tableau 2.7), le modèle lorsque $k = 1$ est meilleur que lorsque $k = 0$, mais la régression linéaire l'emporte dans les deux périodes de prévision, quoique que les résultats du modèle lorsque $k = 0$ demeurent similaires à ceux de la régression correspondante (voir figures 2.9 et 2.10).

Ainsi, ce qui doit être noté, est que les résultats de l'estimateur $\widehat{X}(t+1)$ du processus de Poisson filtré avec $k = 0$ sont similaires à ceux du modèle de régression correspondant, tandis que cet estimateur avec $k = 1$ donne de moins bons résultats comparativement au modèle de régression, mais s'avère meilleur que le modèle classique avec $k = 0$.

Tableau 2.6. Critères de comparaison de prévisions entre modèles pour le fleuve Delaware.

		Estimateur du débit $\widehat{X}(t+1)$	Régression linéaire
Période (21/01/2004-30/04/2004)			
Modèle avec $k = 1$	<i>mae</i>	619,9151	808,6454
	<i>sdae</i>	1 076,2255	948,104116, 36%
	<i>mape</i>	9,54%	1 543 908, 7757
	<i>mse</i>	1 531 088,0801	1 242,5413
	<i>rmse</i>	1 237,3714	
Modèle avec $k = 0$	<i>mae</i>	837,3837	858,3267
	<i>sdae</i>	1 022,0397	1 009,6543
	<i>mape</i>	15,05%	15,71%
	<i>mse</i>	1 735 434,3390	1 746033,4022
	<i>rmse</i>	1 317,3588	1 321,3756
Période (01/10/2003 - 09/01/2004)			
Modèle avec $k = 1$	<i>mae</i>	2 579,25732	2 444,2947
	<i>sdae</i>	5 247,72815	5 167,9355
	<i>mape</i>	12,92%	12,43%
	<i>mse</i>	33 918 558,7870	32 417 702,6619
	<i>rmse</i>	5 823,9642	5 693,6546
Modèle avec $k = 0$	<i>mae</i>	2 959,5348	2 975,4709
	<i>sdae</i>	5 068,7016	5 054,2318
	<i>mape</i>	16,06%	16,12%
	<i>mse</i>	34 196 208,9357	34 145 762,9852
	<i>rmse</i>	5 847,7525	5 843,4376

Tableau 2.7. Critères de comparaison de prévisions entre modèles pour le fleuve Hudson.

		Estimateur du débit $\widehat{X}(t+1)$	Régression linéaire
Période (21/01/2004-30/04/2004)			
Modèle avec $k = 1$	<i>mae</i>	2 296, 9780	2 377, 9165
	<i>sdae</i>	3 957, 4410	3 737, 3510
	<i>mape</i>	9, 28%	9, 73%
	<i>mse</i>	20 782 384, 091	19 483 984, 0464
	<i>rmse</i>	4 558, 7700	4 414, 0666
Modèle avec $k = 0$	<i>mae</i>	2 608, 1403	2 626, 1009
	<i>sdae</i>	3 979, 2372	3 966, 3778
	<i>mape</i>	10, 60%	10, 75%
	<i>mse</i>	22 479 948, 7031	22 472 794, 8554
	<i>rmse</i>	4 741, 3024	4 740, 5479
Période (01/10/2003 - 09/01/2004)			
Modèle avec $k = 1$	<i>mae</i>	4 327, 6741	4 151, 4833
	<i>sdae</i>	7 324, 7111	7 224, 7146
	<i>mape</i>	13, 58%	13, 04%
	<i>mse</i>	71 848 953, 4753	68 914 517, 9778
	<i>rmse</i>	8 476, 3762	8 301, 4769
Modèle avec $k = 0$	<i>mae</i>	4 463, 3310	4 464, 3767
	<i>sdae</i>	7 243, 5280	7 237, 3885
	<i>mape</i>	14, 25%	14, 27%
	<i>mse</i>	71 870 529, 0464	71 791 840, 2566
	<i>rmse</i>	8 477, 6488	8 473, 0066

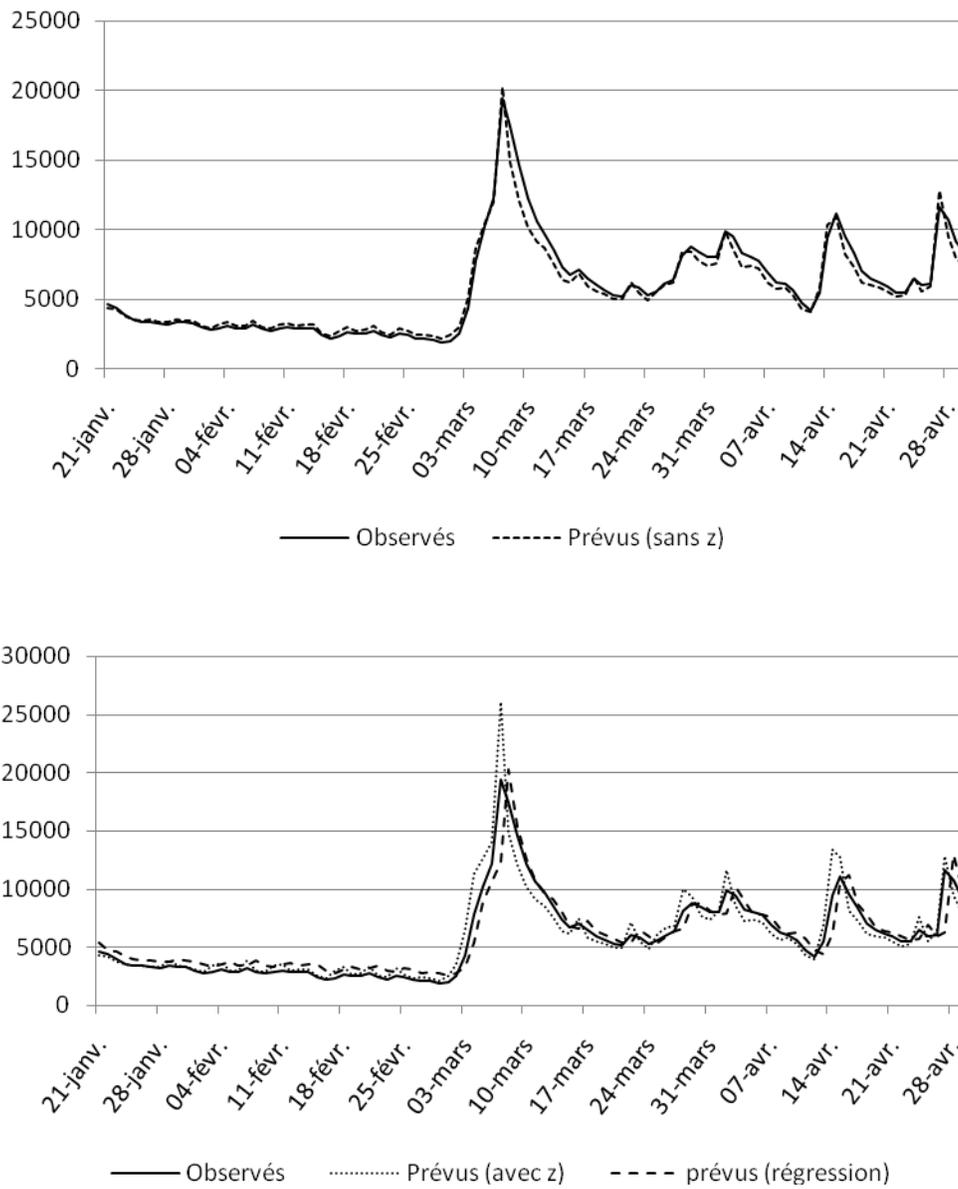


Figure 2.7. Débits (m^3/s) prévus et observés du fleuve Delaware.

(21/01/2004 au 30/04/2004)

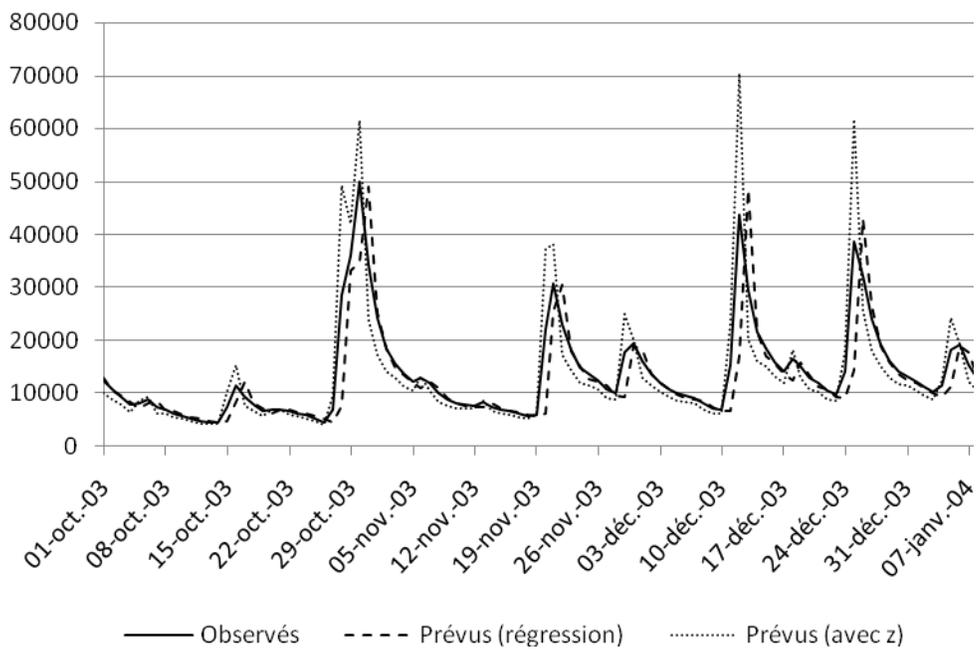
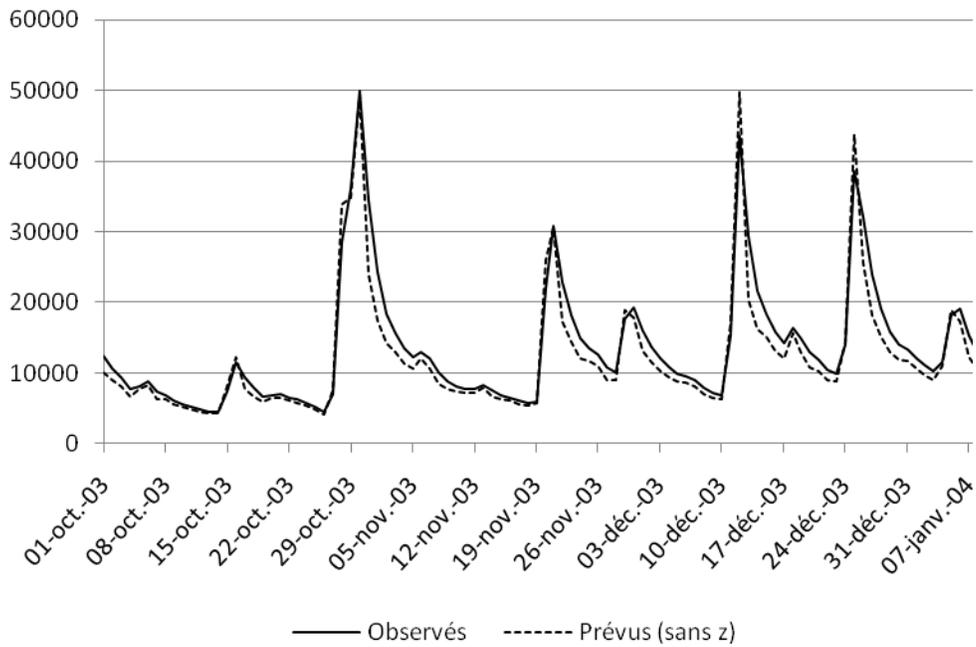


Figure 2.8. Débits (m³/s) prévus et observés du fleuve Delaware.

(01/10/2003 au 09/01/2004)

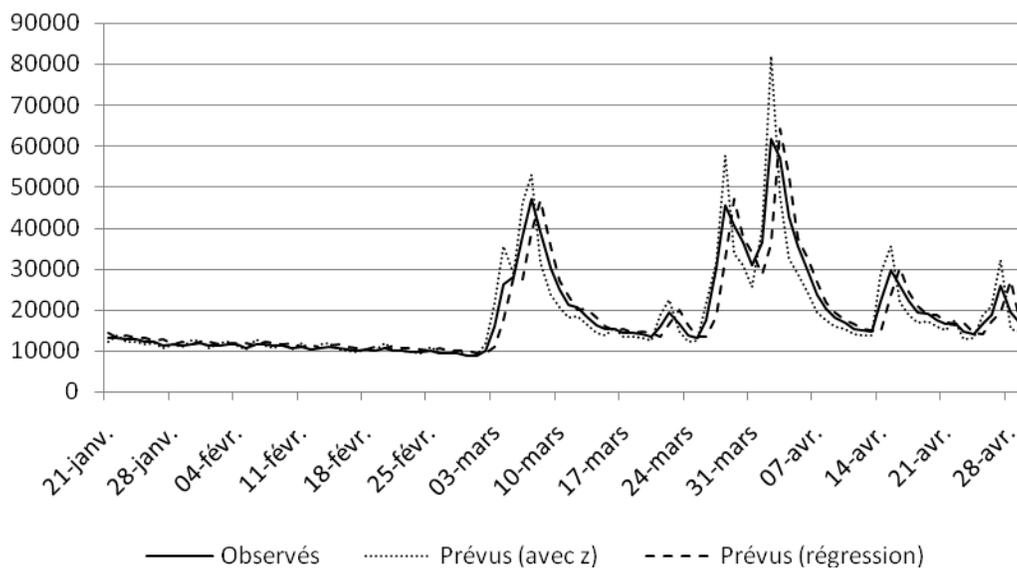
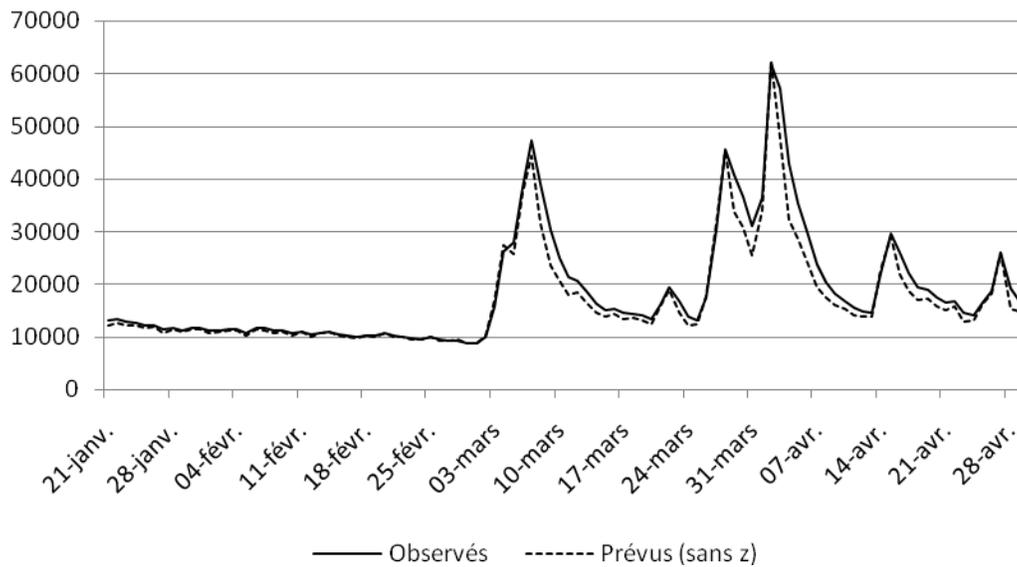


Figure 2.9. Débits (m^3/s) prévus et observés du fleuve Hudson

(21/01/2004 au 30/04/2004).

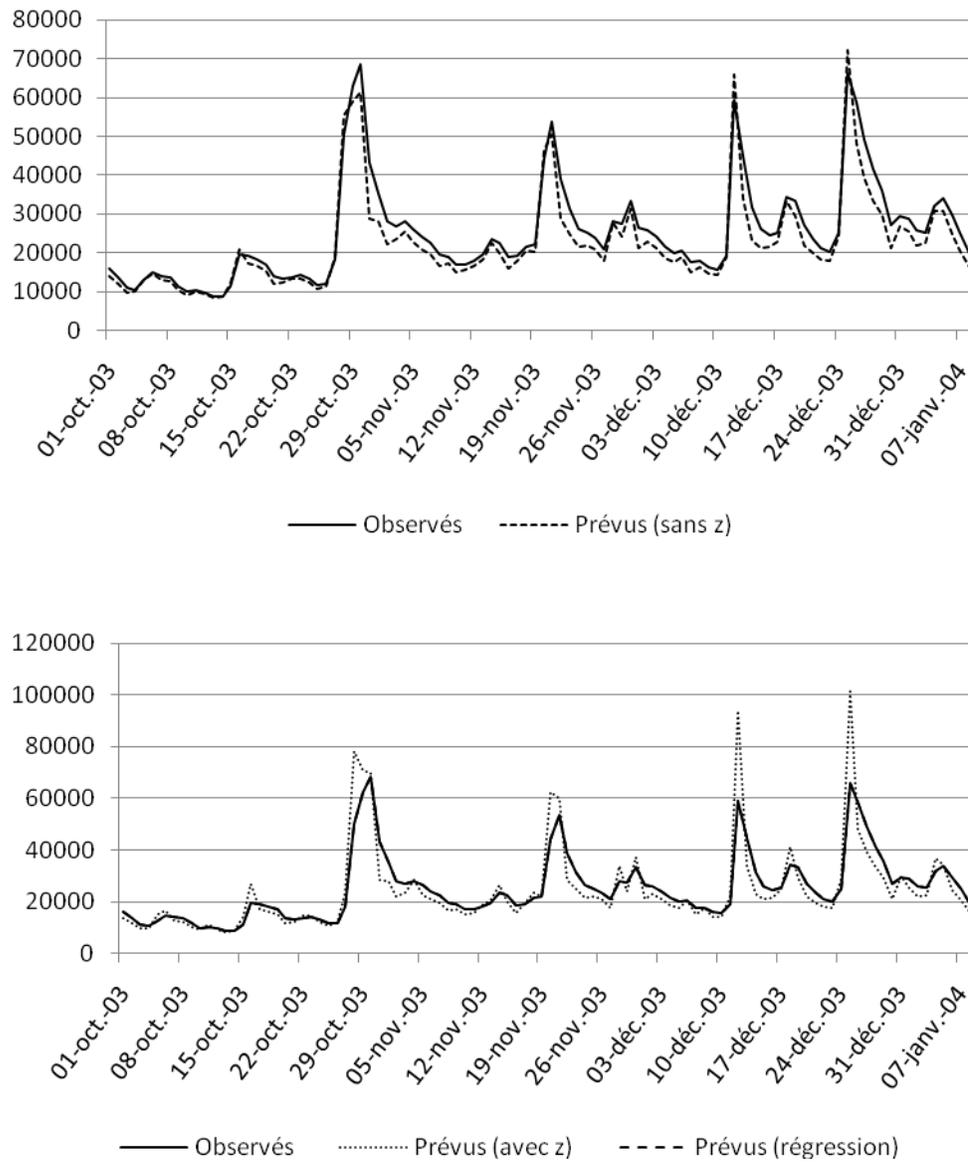


Figure 2.10. Débits (m^3/s) prévus et observés du fleuve Hudson

(01/10/2003 au 09/01/2004).

CONCLUSION

Ce travail donne suite à celui mené par Lefebvre et Guilbault (2008) dans lequel on propose une version plus réaliste du modèle classique du processus de Poisson filtré utilisé pour modéliser le débit journalier $X(t)$:

$$\begin{aligned} X(t) &= \sum_{n=1}^{N(t)} w(t, \tau_n, Y_n) \\ &= \sum_{n=1}^{N(t)} Y_n (t - \tau_n)^k e^{-(t-\tau_n)/c}, \text{ pour } t \geq \tau_n. \end{aligned}$$

Intuitivement, il est évident que la constante k devrait être plus grande que 0. En fait, la fonction de réponse du modèle quand $k = 0$ génère une augmentation instantanée avec une décroissance exponentielle du débit, ce qui est assez loin de la réalité. C'est un modèle qui est simple à utiliser parce que les estimateurs et les formules sont faciles à obtenir, mais les résultats des critères utilisés dans ce travail (voir aussi les résultats de Lefebvre et Guilbault 2008; Guilbault, 2008) prouvent que le modèle avec une valeur de k supérieure à 0 apparaît plus réaliste.

Lefebvre et Guilbault (2008) ont considéré les cas où la constante k prend une valeur nulle ou égale à 1. En analysant la qualité d'ajustement du modèle, les auteurs ont prétendu que la valeur 1/2 pour le paramètre k donnait un meilleur ajustement au modèle. Cependant, dans le cadre de notre travail, nous avons adopté une approche reposant sur la minimisation de l'écart absolu entre les coefficients de corrélation théorique et empirique sur un horizon $\delta (> 0)$ de 7 jours. Étant donné que l'estimation des paramètres est faite en calibrant le modèle sur les observations des débits journaliers au cours d'une année, les variations annuelles de la valeur du paramètre k étaient bien remarquées. Ce paramètre varie selon les fluctuations du débit et peut prendre des valeurs plus élevées (soit $1 \leq k \leq 2$) lorsque la dynamique du fleuve se distingue par l'arrivée de quelques crues exceptionnelles.

En se basant sur le critère de différence des coefficients de corrélation, théorique et empirique, pour comparer la qualité du modèle en tant que telle, nous avons remarqué que la qualité d'ajustement du modèle avec $k = 1$ est équivalente à celle du modèle avec le paramètre estimé \hat{k} ; ces deux derniers modèles s'ajustent plus parfaitement que le modèle classique avec $k = 0$.

Dans le cadre de la capacité prévisionnelle du modèle, nous avons comparé les résultats obtenus du modèle lorsque $k = 1$ avec ceux du modèle classique lorsque $k = 0$ et du modèle autorégressif correspondant. Si les résultats du modèle du processus de Poisson filtré sont comparables à la régression linéaire, c'est déjà un point important marqué par le modèle, puisque le processus autorégressif est un modèle qui fonctionne bien, par défaut, pour les prévisions à court terme d'après l'étude de Lefebvre (2003). En plus, le modèle nous permet de faire plus de prévision en hydrologie; il nous permet de calculer d'autres quantités, des centiles par exemple, de calculer la probabilité que le débit dépasse telle valeur dans certaines périodes, qu'on ne peut pas faire avec la régression. De notre point de vue, le modèle avec $k = 1$ est le meilleur modèle, d'un côté parce qu'il bat le modèle classique avec $k = 0$ et, d'un autre côté, car il donne à peu près la même précision que la régression pour les prévisions à court terme.

On a donc trouvé la valeur du paramètre k : de façon plus précise et on s'est servi du modèle quand k vaut 1 pour les prévisions. Dans l'étude de Lefebvre et Guilbault (2008), pour faire des prévisions avec $k > 0$, il fallait reconstruire toute la séquence de données en introduisant beaucoup d'erreur dans la précision. Mais avec la formule développée, dans ce travail, pour les prévisions avec $k = 1$ (sans le terme z puisqu'il nuit), on améliore les prévisions en les estimant deux jours d'avance. Pour les coefficients de corrélation à moyen terme et pour prévoir, le modèle proposé est meilleur que lorsque $k = 1$. Pour ces deux raisons, l'utilisation du modèle avec $k \in (0, 1)$ apparaît plus réaliste.

L'estimateur $\widehat{X}(t+1)$ lorsque $k = 1$ sans le terme z donne de meilleurs résultats, parce que ce terme fait en sorte qu'il surestime la représentation thématique des données. En

effet, son inclusion dans l'estimateur fait en sorte que la valeur du débit est plus élevée lors de son augmentation par rapport aux données observées. Mais, en pratique, on s'aperçoit que ceci rend les résultats moins bons que lorsque $k = 0$.

Ce qui peut être envisagé pour continuer cette étude est de refaire l'estimation du modèle ainsi que le calcul du débit à l'instant $t + 1$ sur des périodes plus courtes : trimestres, mois, pour prendre en considération les variations saisonnières du débit, ce qui est généralement le cas en hydrologie.

Un autre point important est d'essayer de trouver une formule très explicite pour les prévisions $\widehat{X}(t + 1)$ en tenant compte du paramètre estimé \hat{k} .

BIBLIOGRAPHIE

- Allamano P., Claps P. et Laio F. (2007). Continuous time modelling of hydrologic time series: Shot noise models. In: La Loggia G., Aronica T., Ciruolo G. Water resources assessment under water scarcity scenarios. (pp. 185-204). ISBN: 978-88-900282-8-1. Milano: CSDU (Italy).
- Bado B. A. et Unny T. E. (1990). Modèles linéaires stochastiques théoriques pour la réponse des petits bassins. *Revue des sciences de l'eau*, 3(2), 151-182.
- Bernier, J., Morlat, G., O'Connell, P., O'Donnell, T., Sneyers, R., Delaporte, P. et Borgman, L. (1970). Inventaire des modèles de processus stochastiques applicables à la description des débits journaliers des rivières. *Revue de l'Institut International de Statistique*, 49-104.
- Claps, P., Giordano, A. et Laio, F. (2005). Advances in shot noise modeling of daily streamflows. *Advances in Water Resources*, 28(9), 992-1000.
- Évora, N. D. (1997). *Modélisation stochastique des débits journaliers : une nouvelle approche adaptée aux écoulements sahéliens*. Thèse de doctorat en génie civil, École Polytechnique de Montréal. 180 p.
- Gradshteyn, S. et Ryzhik, I. M. (2000). *Table of Integrals, Series, and Products*, 6e ed. Academic Press, San Diego.
- Guilbault, J.L. (2008). *Processus de Poisson filtrés, chaînes de Markov et applications*. Thèse de doctorat en mathématiques appliquées, École Polytechnique de Montréal. 80 p.
- Kavvas, M. et Delleur, J. (1984). A statistical analysis of the daily streamflow hydrograph. *Journal of Hydrology*, 71(3-4), 253-275.

- Kelman, J. (1980). A stochastic model for daily streamflow. *Journal of Hydrology*, 47(3-4), 235-249.
- Koch, R. W. (1985). A stochastic streamflow model based on physical principles. *Water Resources Research*, 21(4), 545-553.
- Konecny, F. (1992). On the shot-noise streamflow model and its applications. *Stochastic Hydrology and Hydraulics*, 6(4), 289-303.
- Lawrance, A. et Kottegoda, N. (1977). Stochastic modelling of riverflow time series. *Journal of the Royal Statistical Society. Series A (General)*, 1-47.
- Lefebvre, M (2003). Prévisions hydrologiques à court terme obtenues en utilisant la régression linéaire. *Revue des sciences de l'eau*, 16(2), p. 255-265.
- Lefebvre, M. (2005a). A filtered renewal process as a model for a river flow. *Mathematical Problems in Engineering*, DOI 10.1155/MPE.2005.49, p. 49-59.
- Lefebvre, M. (2005b). *Processus Stochastiques Appliqués*: Presses internationales Polytechnique xii, 430 p.
- Lefebvre, M. et Guilbault, J. L. (2008). Using filtered Poisson processes to model a river flow. *Applied Mathematical Modelling*, 32(12), 2792-2805.
- Lefebvre, M. et Guilbault, J.L. (2009). On a process derived from a filtered Poisson process. *Revue Roumaine de Mathématiques Pures et Appliquées*, 54(2), 147-159.
- Lefebvre, M., Ribeiro, J., Rousselle, J., Seidou, O. et Lauzon, N. (2003). *Probabilistic prediction of peak food discharges*, in : A. Der Kiureghian, S. Mandanat and J.M. Pestana. Proceedings of the ICASP9 Conference, Millpress, Rotterdam, Netherlands.

- Murrone, F., Rossi, F. et Claps, P. (1997). Conceptually-based shot noise modeling of streamflows at short time interval. *Stochastic Hydrology and Hydraulics*, 11(6), 483-510.
- Nash, J. (1957). The form of the instantaneous unit hydrograph, *Int. Assoc. Sci. Hydrol., Gen. Assess. Toronto*, 3, 114–121.
- Salas, J. D. (1980). *Applied Modeling of Hydrologic Time Series*: Water Resources Publication.
- Seidou, O., Rousselle, J., Lefebvre, M., Lauzon, N. et Ribeiro, J. (2002). Modélisation de l'incertitude sur les séquences futures de débits en rivière. *Hydrological Sciences Journal*, 47(3), 367-385.
- Treiber, B. et Plate, E. J. (1977). A stochastic model for the simulation of daily flows. *Hydrol. Sci. Bull.*, 22(1), 175-192.
- Weiss, G. (1977). Shot noise models for the generation of synthetic streamflow data. *Water Resources Research*, 13(1), 101-108.
- Yue, S., Hashino, M., Bobee, B., Rasmussen, P. et Ouarda, T. (1999). Derivation of streamflow statistics based on a filtered point process. *Stochastic Environmental Research and Risk Assessment*, 13(5), 317-326.

ANNEXE A

ILLUSTRATION DES TRAJECTOIRES DE PROCESSUS

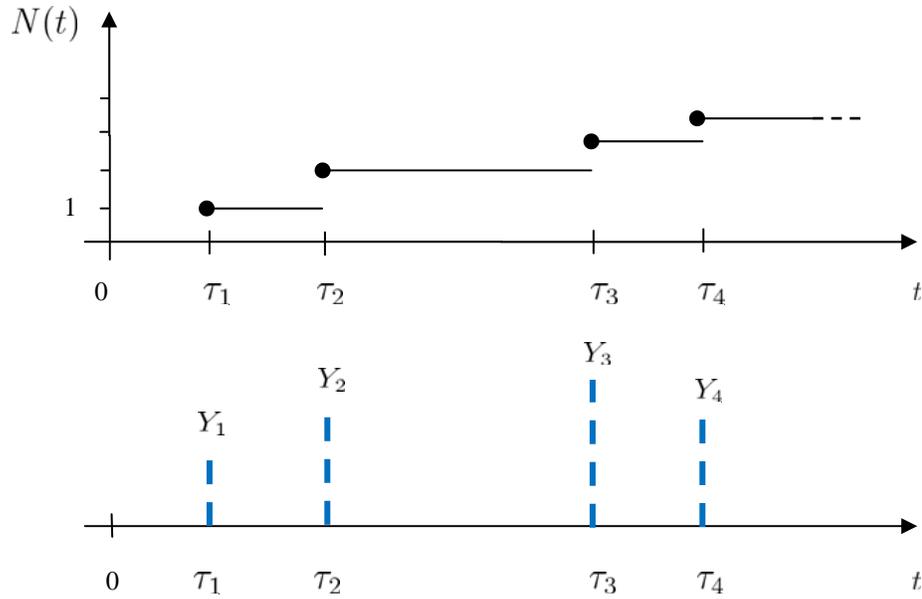


Figure A.1. Exemple de la trajectoire d'un processus de Poisson

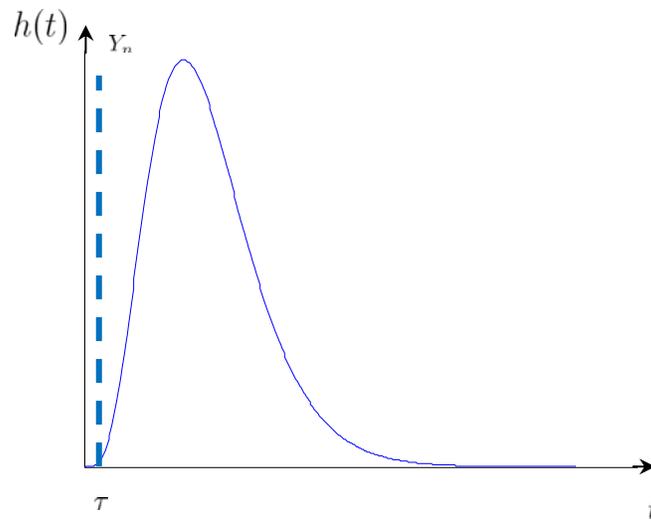


Figure A.2. Exemple de la trajectoire de la fonction de réponse unitaire de

l'hydrogramme : $h(t, \tau) = (t - \tau)^k e^{-(t-\tau)/c}$.

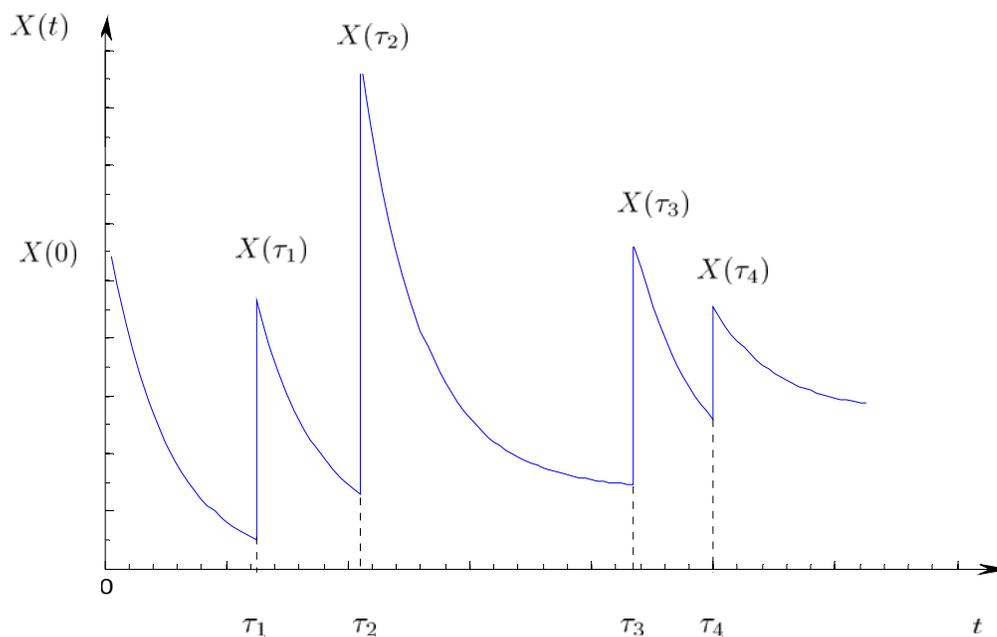


Figure A.3. Exemple de la trajectoire d'un processus de Poisson filtré (*shot noise process*)

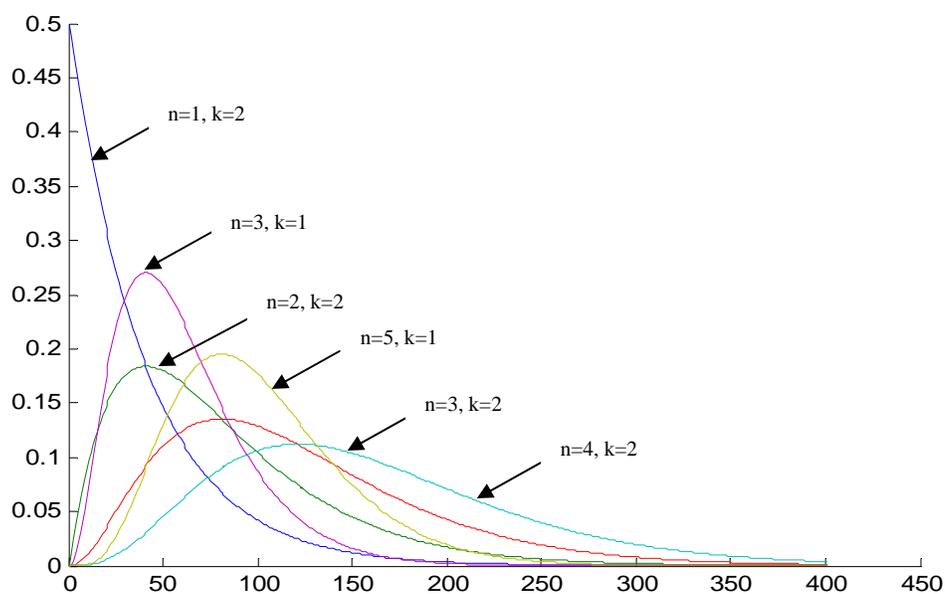


Figure A.4. Forme de l'HUI par le modèle de Nash (fonction gamma $\Gamma(n, k)$)