Revue des sciences de l'eau
Journal of Water Science

Échantillonnage par valeurs supérieures à un seuil : modélisation des occurrences par la méthode du renouvellement
Over-threshold sampling : Modeling of occurrences by renewal processes

M. Lang, P. Rasmussen, G. Oberlin and B. Bobée

Volume 10, Number 3, 1997

URI: https://id.erudit.org/iderudit/705281ar
DOI: https://doi.org/10.7202/705281ar

See table of contents

Publisher(s)

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

ISSN

0992-7158 (print)
1718-8598 (digital)

Explore this journal

Cite this article

Lang, M., Rasmussen, P., Oberlin, G. & Bobée, B. (1997). Échantillonnage par valeurs supérieures à un seuil : modélisation des occurrences par la méthode du renouvellement. *Revue des sciences de l'eau / Journal of Water Science*, *10*(3), 279–320. https://doi.org/10.7202/705281ar

Article abstract

The principle of over-threshold sampling is to consider all the events in a time-series that exceed a given threshold. The probabilistic analysis implies estimating two statistical models, one describing the occurrence of events (date of the events), the other describing their magnitude (value of the local maximum). These two models are then combined to obtain the distribution of annual maximum flows.

The theory of renewal processes can be used to study the occurrence of flood events. We present here properties of the well-known Poisson distribution (stationary or non-stationary process), and certain new results for the binomial and negative binomial distributions. These results concern the distribution of the number of events in a given time interval and the distribution of the waiting time, defined as the time span between two successive exceedances of the threshold.

The relationship between the distribution of a variable and its corresponding return period are then studied in more detail. We review the results for the case where the variable of interest is obtained by selection of one or more events per year, or from a Poisson point process. New results are presented for the case of the binomial and negative binomial processes.

Finally, we establish the analytical relationship between the two types of sampling, annual maximum sampling and peaks-over-threshold sampling, in terms of return period, distribution of the annual maximum, and sampling variance.

Échantillonnage par valeurs supérieures à un seuil : modélisation des occurrences par la méthode du renouvellement

Over-threshold sampling: Modeling of occurrences by renewal processes

M. LANG[1], P. RASMUSSEN[2], G. OBERLIN[1] et B. BOBEE[2]

Reçu le 12 août 1996, accepté le 13 mars 1997*.

Cet article est publié intégralement en français et en anglais ; les références bibliographiques communes aux deux versions, sont placées aprés le texte en anglais, voir p.318.
This paper is published integrally in both French and English; see p. 300.

RÉSUMÉ

L'échantillonnage par valeurs supérieures à un seuil consiste à retenir tous les événements d'une chronique, définis par l'existence d'un maximum local supérieur à un seuil critique. L'étude probabiliste est alors menée par calage de deux lois de probabilité, une sur le processus d'occurrence de ces événements (date des événements), une autre sur la marque des événements (valeur du maximum local), puis par recomposition de ces deux lois pour obtenir la loi de probabilité associée au maximum annuel.

La théorie du renouvellement permet d'étudier le processus d'occurrence d'événements. Les propriétés générales de la loi le plus souvent utilisée, la loi de Poisson (stationnaire ou non), sont présentées, ainsi que des éléments nouveaux concernant la loi Binomiale et la loi Binomiale négative. Ces propriétés sont relatives à la distribution du nombre d'événements sur un intervalle de temps donné, et à la distribution de la durée de retour, définie comme l'intervalle de temps séparant deux occurrences successives d'événements.

Les relations existant entre la loi de probabilité d'une variable et la période de retour de l'événement associé sont ensuite détaillées. Il s'agit d'un rappel de résultats lorsque la variable étudiée est obtenue par sélection d'un ou de plusieurs maximums par an, ou dans le cas d'un processus marqué de Poisson ; et d'éléments nouveaux dans le cas d'un processus représenté par une loi Binomiale ou une loi Binomiale négative.

Pour finir, on trouvera les correspondances entre les deux types d'échantillonnage précédents (par maximum annuel ou par valeurs supérieures à un seuil),

1. Cemagref Lyon – Division Hydrologie-Hydraulique, 3 bis quai Chauveau 69336 Lyon Cedex 09, France.
2. INRS-EAU, Université du Québec, 2800 rue Einstein, Sainte-Foy (PQ), G1V4C7, Canada.

* Les commentaires seront reçus jusqu'au 20 mars 1998.

en terme de période de retour, de distribution et de variance d'échantillonnage.

Mots-clés : *valeurs supérieures à un seuil, échantillonnage, théorie du renouvellement, processus d'occurence, distribution des crues.*

SUMMARY

The principle of over-threshold sampling is to consider all the events in a time-series that exceed a given threshold. The probabilistic analysis implies estimating two statistical models, one describing the occurrence of events (date of the events), the other describing their magnitude (value of the local maximum). These two models are then combined to obtain the distribution of annual maximum flows.

The theory of renewal processes can be used to study the occurrence of flood events. We present here properties of the well-known Poisson distribution (stationary or non-stationary process), and certain new results for the binomial and negative binomial distributions. These results concern the distribution of the number of events in a given time interval and the distribution of the waiting time, defined as the time span between two successive exceedances of the threshold.

The relationship between the distribution of a variable and its corresponding return period are then studied in more detail. We review the results for the case where the variable of interest is obtained by selection of one or more events per year, or from a Poisson point process. New results are presented for the case of the binomial and negative binomial processes.

Finally, we establish the analytical relationship between the two types of sampling, annual maximum sampling and peaks-over-threshold sampling, in terms of return period, distribution of the annual maximum, and sampling variance.

Key words : *flood frequency analysis, partial duration series, threshold values, sampling techniques, renewal process.*

1 – INTRODUCTION

L'analyse fréquentielle des crues est le plus souvent réalisée à partir du calage des paramètres d'une loi de probabilité sur l'échantillon formé de la valeur maximum de chaque année (ASHKAR *et al.*, 1994). Une alternative consiste à retenir tous les événements d'une chronique, définis par l'existence d'un maximum local supérieur à un seuil critique (en anglais, Peak Over-Treshold Values ou Partial Duration Series). L'étude probabiliste est alors menée par calage de deux lois de probabilité, une sur le processus d'occurrence de ces événements (date des événements), une autre sur la marque des événements (valeur du maximum local), puis par recomposition de ces deux lois pour obtenir la loi de probabilité associée au maximum annuel.

L'objet de cet article est relatif à la modélisation des occurrences, en liaison avec la méthode du renouvellement (COX, 1965, FELLER, 1966), dont les premières applications dans le domaine de l'hydrologie sont à mettre à l'actif de BORGMAN (1963), SHANE et LYNN (1964) et BERNIER (1967). Le lecteur intéressé à la modélisation des dépassements se reportera utilement aux travaux théoriques de PICKANDS (1975), DAVISON et SMITH (1990), ou à des considérations pratiques sur le type de loi à utiliser (2 ou 3 paramètres), le choix du seuil, l'indépendance des valeurs présentées par ROSBJERG *et al.* (1991, 1992) ; ROSBJERG et MADSEN (1992) ; LANG (1995a).

Nous passons en revue quelques lois utilisées pour décrire les processus, puis nous présentons les relations existant entre la loi de probabilité d'une variable et la période de retour de l'événement considéré. Ces relations sont spécifiques du type d'échantillonnage utilisé : par sélection de valeurs maximum par épreuve ou supérieures à un seuil. Nous indiquons ensuite des éléments permettant de comparer ces deux types d'échantillonnage.

2 – ÉTUDE DES PROCESSUS

On considère un processus d'occurrence d'événements E, décrit soit par la durée θ séparant deux occurrences successives d'un événement, appelée durée de retour, soit par le nombre d'événements m_t survenus dans l'intervalle [0;t]. On associe à chaque variable une fonction de répartition, une densité et éventuellement sa valeur moyenne :

– durée de retour θ : $F(d) = \mathrm{Prob}[\theta < d]$ et $f(x) \cdot dx = \mathrm{Prob}[x < \theta < x + dx]$

On suppose que $F(0) = 0$ de façon à ce qu'il ne soit pas possible d'avoir simultanément deux événements. On définit également la période de retour de l'événement par :

$$T = E(\theta) = \int_0^{+\infty} \theta \cdot f(\theta) \cdot d\theta$$

– nombre d'événements m_t sur [0;t] : $w_k\,[0;t] = \mathrm{Prob}[m_t = k]$

On définit également le nombre moyen d'événements N(t) sur [0;t] : $N(t) = E(m_t)$, l'intensité du processus $\mu(t) = dN(t)/dt$, et un indice de dispersion $I_t = \mathrm{Var}(m_t) / E(m_t)$.

2.1 Flux de Poisson

On suppose que le processus d'occurrence des événements respecte 4 hypothèses :

(i) homogénéité dans le temps des événements,

(ii) la probabilité d'avoir un événement pendant une courte durée dt est très faible, du même ordre que dt,

(iii) la probabilité d'avoir plus d'un événement pendant une courte durée dt est infime, négligeable devant dt,

(iv) indépendance successive des événements.

On peut montrer (BASS ,1974, p. 145-148, cf. annexe 1), que ces hypothèses conduisent aux relations :

$$w_k\,[0;t] = \exp[-\mu t] \cdot (\mu t)^k/k! \tag{1}$$

$$F(d) = \mathrm{Prob}[\theta < d] = 1 - \exp[-\mu \cdot d] \tag{2}$$

Ainsi, le nombre d'événements E pendant l'intervalle de temps [0;t] suit une loi de Poisson, de moyenne $N(t) = \mu \cdot t$ et de variance $Var(m_t) = \mu \cdot t$ (eq. 1). L'intensité du processus $\mu(t)$ est dans ce cas constante et égale à μ. La durée de retour θ séparant deux événements suit une loi exponentielle simple (eq. 2), la période de retour de l'événement vaut $T = 1 / \mu$, et l'indice de dispersion est égal à 1 ($I_t = 1$). Si on remplace le paramètre μ de la loi de Poisson par son estimation $\hat{\mu} = \hat{E}(m_t)/t$, on obtient les relations :

$$F(d) = 1 - \exp[-d/\hat{\alpha}] \tag{3}$$

$$T = E(\theta) = \hat{\alpha} \tag{4}$$

avec

$$(1/\hat{\alpha}) = \hat{\mu} \tag{5}$$

2.2 Flux de Poisson non-stationnaire

On suppose dans ce cas qu'il n'y a plus homogénéité dans le temps des événements. On retient seulement les trois dernières hypothèses du flux de Poisson. L'hypothèse (ii) s'écrit alors : $w_1[t;t + dt] = \mu(t) \cdot dt$. Par un raisonnement analogue à celui du flux de Poisson (VENTSEL, 1973, p. 510-511), on arrive aux relations :

$$w_k\,[t;t'] = \exp[-N(t,t')] \cdot (N(t,t'))^k/k! \tag{6}$$

où $N(t, t') = \int_t^{t'} \mu(\tau) \cdot d\tau$ représente le nombre moyen d'événements pendant l'intervalle de temps [t ;t'].

$$F_t(d) = \mathrm{Prob}[\theta(t) < d] = 1 - \exp[-\int_t^{t+d} \mu(\tau) \cdot d\tau] \tag{7}$$

Le nombre d'événements E *sur l'intervalle [t;t'] suit une loi de Poisson*, de moyenne N(t,t'). L'intensité du processus est fonction du temps :

$$\lim_{\Delta t \to 0^+} N(t,t+\Delta t)/\Delta t = \mu(t)$$

BORGMAN (1963) a traité plus en détail le cas du flux de Poisson non stationnaire avec variations saisonnières, mais stabilité inter-annuelle. Il est possible *alors de se ramener au flux de Poisson simple, par l'intermédiaire d'un change*ment d'échelle sur le temps :

$$\tilde{t} = \int_0^t \mu(\tau) \cdot d\tau$$

On a alors :

$$w_k[t_1;t_2] = \exp[-(\tilde{t}_2 - \tilde{t}_1)] \cdot (\tilde{t}_2 - \tilde{t}_1)^k/k! \tag{8}$$

(processus réduit de Poisson, avec $\mu = 1$),

$$\tilde{T} = 1, \text{soit} : \int_0^T \mu(\tau) \cdot d\tau = 1$$

2.3 Loi Binomiale négative

Plusieurs auteurs, CUNNANE (1979), MIQUEL (1984), BOIRET (1987), VUKMIROVIC (1990), BEN-ZVI (1991), proposent la loi Binomiale Négative pour décrire le processus d'occurrence :

$$w_k[0;t] = C^k_{\gamma t + k - 1} \delta^{\gamma t} \cdot (1 - \delta)^k \tag{9}$$

où δ est un paramètre compris entre 0 et 1. La combinaison $C^k_{\gamma t + k - 1}$ est calculée par l'expression des factorielles en terme de fonction Gamma.

Le nombre m_t a pour valeur moyenne $N(t) = \gamma t\ (1 - \delta)/\delta$, pour variance $Var(m_t) = \gamma t\ (1 - \delta)/\delta^2$ (eq. 9). L'indice de dispersion est supérieur à 1 : $I_t = 1/\delta > 1$.

On peut montrer (cf. annexe I) que la durée de retour a alors les propriétés suivantes :

$$F(d) = Prob\ [\theta < d] = 1 - \delta^{\gamma d} \tag{10}$$

$$T = E(\theta) = \int_0^{+\infty} (-\gamma \cdot Log\delta \cdot \delta^{\gamma\theta}) \cdot \theta \cdot d\theta = -1/(\gamma \cdot Log\delta) \tag{11}$$

Si on remplace les paramètres γ et δ par leurs estimations $\hat{\gamma} = \hat{\mu}/(\hat{I}_t - 1)$ et $\hat{\delta} = 1/\hat{I}_t$, les équations (10) et (11) ont la même expression que les équations (3) et (4), avec :

$$1/\hat{\alpha} = \hat{\mu} \cdot (Log\hat{I}_t/(\hat{I}_t - 1)) \tag{12}$$

Cette loi constitue une alternative au flux de Poisson, lorsque l'indice de dispersion est plus grand que 1.

2.4 Loi Binomiale

VUKMIROVIC (1990) propose également la loi Binomiale :

$$w_k[0;t] = C^k_{\gamma t} \delta^k \cdot (1 - \delta)^{\gamma t - k} \tag{13}$$

La moyenne et la variance du nombre m_t sont respectivement, $N(t) = \gamma t \cdot \delta$, et $Var(m_t) = \gamma t \cdot \delta \cdot (1 - \delta)$ (eq. 13). L'indice de dispersion est inférieur à 1 : $I_t = 1 - \delta < 1$.

On peut montrer (cf. annexe I) que la durée de retour a alors les propriétés suivantes :

$$F(d) = Prob\ [\theta < d] = 1 - (1 - \delta)^{\gamma d} \tag{14}$$

$$T = E(\theta) = \int_0^{+\infty} (-\gamma \cdot Log(1 - \delta) \cdot (1 - \delta)^{\gamma\theta}) \cdot \theta \cdot d\theta = -1/(\gamma \cdot Log(1 - \delta)) \tag{15}$$

Si on remplace les paramètres γ et δ par leurs estimations $\hat{\gamma} = (\hat{\mu}/(1 - \hat{I}_t))$ et $\hat{\delta} = 1 - \hat{I}_t$, les équations (14) et (15) ont la même expression que les équations (3) et (4), avec la même valeur du paramètre $\hat{\alpha}$ (eq. 12) que la loi Binomiale négative.

Cette loi constitue une alternative au flux de Poisson, lorsque l'indice de dispersion est plus petit que 1.

2.5 Autres lois

Il existe d'autres alternatives pour traiter le cas de processus plus complexes. BERNIER (1967) présente certaines généralisations du processus de renouvellement : le flux d'Erlang d'ordre k, obtenu à partir du flux de Poisson en ne retenant qu'un événement tous les (k + 1) événements, les lois d'Erlang généralisées, où l'ordre k est quelconque (pas seulement un entier), le processus retardé, où l'origine du temps est fixée entre deux événements, le processus cumulatif, où on associe une deuxième variable à la variable de base, le processus alterné, où la variable étudiée peut prendre seulement deux états (également traité par KUNDZEWICZ, 1989), le processus semi-markovien (généralisation du processus alterné à plusieurs états). KARR (1976), ROSBJERG (1977b), MADSEN et ROSBJERG (1995a) ont étudié le processus d'occurrence d'événements représentés par une chaîne de Markov (avec prise en compte de l'autocorrélation de rang 1). DAVISON et SMITH (1990) font référence à des travaux de LEADBETTER (1983), LINDGREN et ROOTZEN (1987), LEADBETTER et ROOTZEN (1988) permettant de traiter le cas d'un processus de Poisson pour des groupes d'événements liés. KAVVAS (1982a et b) et CERVANTES *et al.* (1983) proposent un modèle très intéressant de processus de Poisson non stationnaire à deux niveaux : le premier niveau simule l'arrivée de perturbations météorologiques indépendantes, le deuxième niveau permet de modéliser plusieurs événements d'une perturbation.

3 – ÉCHANTILLONNAGE PAR SÉLECTION DE VALEUR MAXIMUM PAR ÉPREUVE OU SUPÉRIEURES À UN SEUIL

Le mode d'échantillonnage le plus utilisé pour l'étude des risques de crues consiste à sélectionner la crue la plus forte de chaque année. On trouvera dans ASHKAR *et al.* (1994) une revue des différents problèmes liés à ce type d'analyse : contrôles sur les échantillons, critères de choix d'une distribution et d'une méthode d'estimation des paramètres, régionalisation... Une alternative consiste à retenir toutes les valeurs supérieures à un seuil, en utilisant la méthode du renouvellement pour décrire le processus d'occurrence des crues supérieures à un seuil. RASMUSSEN (1991), ROSBJERG (1993), RASMUSSEN *et al.* (1994), puis LANG (1995a) ont dressé un état de l'art des différents travaux effectués sur le sujet.

Nous présentons les caractéristiques principales de ces deux types d'échantillonnage, puis nous donnons les relations permettant de passer d'un type à l'autre.

3.1 Échantillonnage par sélection de valeurs maximum sur une épreuve de durée fixe

3.1.1 Sélection d'une valeur maximum par épreuve

Soient X, une variable aléatoire (V.A) quelconque, X^* une V.A définie comme la valeur maximum de X sur une épreuve (l'année, la saison, le mois...). La durée de l'épreuve sur laquelle on extrait les valeurs de X^* est notée t.

Si on définit une occurrence d'événement comme le dépassement d'une valeur donnée x par la valeur maximum X^* sur l'épreuve de durée t ($X^* > x$), on peut associer à l'événement les éléments suivants :

– Période de retour : $T_x(x)$ (l'indice x pour échantillonnage par maximum).
– Quantile : X (T_x), la valeur de X = x relative à la période de retour T_x.
– Probabilité de non-dépassement : $F_x(x) = \text{Prob}\,[X^* < x]$.

En supposant que le flux d'événements suit une loi de Poisson, on peut relier $F_x(x)$ et $T_x(x)$:

$$\text{Prob}[\text{k événements en N épreuves}] = C_N^k [F_x(x)]^{N-k} \cdot [1 - F_x(x)]^k$$

C'est une loi Binomiale, de moyenne $N \cdot (1 - F_x(x))$. Le nombre moyen d'événements sur l'intervalle [0;t] vaut :

$$N(x,t) = t \cdot [1 - F_x(x)]$$

Comme la période de retour d'un événement suivant un flux de Poisson vaut $T = 1/\mu$, où $\mu = dN(x,t)/dt$, on a :

$$F_x(x) = 1 - t/T_x(x) \tag{16}$$

3.1.2 Sélection de k valeurs maximum par épreuve

Soient X une variable aléatoire, et X_k une des k plus fortes valeurs de X sur une épreuve de durée t.

Si on définit une occurrence d'événement comme le dépassement d'une valeur donnée x par une des k plus fortes valeurs de X sur l'épreuve de durée t ($X_k > x$), on peut associer à l'événement les éléments suivants :

– Période de retour : $T_x(x)$ (l'indice x pour échantillonnage par maximum).
– Quantile : X (T_x), la valeur de X = x relative à la période de retour T_x.
– Probabilité de non-dépassement : $G_{kx}(x) = \text{Prob}\,[X_k < x]$

On peut relier facilement F_x et G_{kx}, puis G_{kx} et T_x :

$$\begin{aligned} F_x(x) &= \text{Prob}\,[X^* < x] \\ &= \text{Prob}\,[(X_k)_1 < x \text{ et } (X_k)_2 < x \text{ et} \ldots (X_k)_k < x] \\ &= [G_{kx}(x)]^k \text{ (en supposant l'indépendance des k plus fortes valeurs)} \end{aligned}$$

ce qui donne :

$$G_{kx}(x) = (1 - t/T_x(x))^{1/k} \tag{17}$$

La relation (17) a été établie en supposant l'indépendance des valeurs X_k, ce qui conduit à rejeter, parmi les k plus fortes valeurs de chaque épreuve, celles qui sont supposées appartenir au même événement. L'implantation d'un critère d'indépendance est souvent difficile à réaliser, et on se contente en général d'un critère d'espacement (durée minimum entre deux valeurs).

Ce type d'échantillonnage permet d'augmenter le nombre de valeurs de l'échantillon, mais, comme le précédent, il présente l'inconvénient de prendre un nombre identique de valeurs pour chaque épreuve, sans s'intéresser au processus d'occurrence des événements par épreuve. Ainsi en hydrologie, où il existe des années « sèches » et des années « humides », cet échantillonnage fournit une population pas toujours homogène, avec des valeurs extrêmes peu intéressantes pour les années « sèches » et la non-sélection de valeurs intéressantes pour les années « humides ». C'est pour cette raison que nous préconisons l'échantillonnage suivant où l'on sélectionne toutes les plus fortes valeurs sur l'ensemble de la chronique.

3.2 Échantillonnage par sélection de valeurs supérieures à un seuil

3.2.1 Formulation générale

Soient X, une variable aléatoire (V.A) et X_S, la V.A définie comme la valeur maximum de X sur un épisode. L'épisode est relatif à une valeur seuil S : il commence quand X > S et finit quand X < S.

Si on définit une occurrence d'événement comme le dépassement d'une valeur donnée x par la valeur maximum X_S d'un épisode ($X_S > x$), on peut associer à l'événement les éléments suivants :

– Période de retour : $T_s(x)$ (l'indice s pour échantillonnage par valeur sup-seuil).

– Quantile : X (T_S), la valeur de X = x relative à la période de retour T_S.

– Probabilité de non-dépassement : $G_S(x)$ = Prob [$X_S < x$].

On peut relier la loi de probabilité F_x avec w_k (t) et G_s :

$F_x(x)$ = Prob [$X^* < x$], où X^* est la valeur maximum de X sur une épreuve de durée t.

$w_k(t)$ = Prob [k épisodes supérieurs au seuil S pendant t].

$G_S(x)$ = Prob [$X_S < x$] où X_S est la valeur maximum de l'épisode.

$$F_x(x) = \sum_{k=0}^{\infty} w_k(t) \cdot [G_s(x)]^k \tag{18}$$

(en supposant l'indépendance des maximums de chaque épisode).

Par rapport à la présentation sur les processus, il faut noter qu'il y a maintenant deux sortes d'événements : l'événement « X supérieur au seuil S », appelé dorénavant épisode sup-seuil, auquel on rattache la loi de probabilité $w_k(t)$; et l'événement « X_S supérieur à x », appelé événement, auquel on rattache la loi de probabilité $G_S(x)$. L'introduction de la loi G_S va permettre d'extrapoler la distribution expérimentale vers des périodes de retour élevées, ce qui n'était pas possible avec la méthode du renouvellement proprement dite. En effet, la période de retour d'un événement E vaut T(E) = 1/μ, où μ est le nombre moyen d'événements observés par épreuve. Pour les événements rares, on ne dispose pas d'une estimation fiable de μ.

Deux difficultés relatives à l'échantillonnage par valeurs supérieures à un seuil concernent l'introduction de nouveaux paramètres liés au processus des crues, et les hypothèses sous-jacentes à l'utilisation de ces modèles. Le premier aspect sera partiellement abordé au paragraphe 4, pour évaluer l'effet de l'incertitude de la modélisation du processus sur l'estimation des quantiles. Le deuxième aspect est fonction du processus retenu. Dans le cas du flux de Poisson, les deux hypothèses les plus contraignantes concernent l'indépendance successive des événements et la stationnarité du phénomène. Le choix du seuil résulte en général d'un compromis entre un nombre élevé de valeurs dont certaines liées (seuil bas) et un faible nombre de valeurs indépendantes (seuil haut). Si les hypothèses du flux de Poisson ne peuvent être respectées, avant d'utiliser des processus plus complexes, il est conseillé de tester les lois Binomiale et Binomiale négative. L'existence d'un caractère saisonnier du processus peut être traitée simplement en raisonnant sur les maximums des épisodes appartenant à chacune des saisons mises en évidence, ou en utilisant un flux de Poisson non stationnaire.

Nous reprenons les quatre processus les plus simples qui viennent d'être décrits, sachant que d'autres lois sont disponibles (§ 2.5).

3.2.2 Cas de la loi Binomiale

On se propose de déduire les propriétés du nombre de dépassement de la valeur x (x > S) à partir du nombre de dépassement du seuil S. Il suffit de remarquer que l'on passe du niveau de base au niveau x par effacement aléatoire (événement effacé avec la probabilité $G_S(x)$; événement supérieur à x conservé avec la probabilité $1 - G_S(x)$). Le processus des événements conservés est ainsi une sous-classe (déterminée si x est fixé) du processus des événements de base.

Le processus des événements X_S suit une loi Binomiale :

$$B_k(\delta, \gamma t) = \text{Prob}\,[k \text{ valeurs } X_S \text{ sur } \gamma t \text{ valeurs disponibles ; avec } \delta = \text{Prob}\,[X > S]]$$

avec un nombre moyen d'épisodes sup-seuil sur la durée t égal à : $N(t) = \gamma t \cdot \delta$

Le tirage sur une sous-classe donne nécessairement une loi Binomiale :

$$B_K(\delta', \gamma t) = \text{Prob}\,[k \text{ valeurs } X_S > x \text{ sur } \gamma t \text{ valeurs disponibles ;}$$
$$\text{avec } \delta' = \text{Prob}\,[X_S > x]]$$

avec un nombre moyen d'événements ($X_S > x$) sur la durée t égal à : $N(x, t) = \gamma t \cdot \delta'$.

Or : $\delta' = \text{Prob}\,[X_S > x] = \text{Prob}\,[X > x] \cdot \text{Prob}\,[X_S > x \mid X > S] = \delta \cdot [1 - G_S(x)]$

D'où :

$$N(x,t) = \gamma t \cdot \delta \cdot [1 - G_S(x)] \tag{19}$$

$$\text{Prob}[k \text{ valeurs } X_S > x \text{ pendant } t] = C_{\gamma t}^{k} A^{k} \cdot (1 - A)^{\gamma t - k} \tag{20}$$

avec : $A = \delta \cdot [1 - G_S(x)]$

$$\begin{aligned} F_X(x) &= \text{Prob}\,[X^* < x \text{ sur } [0;t]] \\ &= \text{Prob}\,[0 \text{ valeurs } X_S > x \text{ sur } [0;t]] \\ &= [\,1 - \delta \cdot (1 - G_S(x))\,]^{\gamma t} \qquad \text{(d'après (20))} \end{aligned} \tag{21}$$

qui donne, en remplacant γ et δ par leurs estimations $\hat{\gamma}$ et $\hat{\delta}$:

$$F_X(x) = [\hat{I}_t - (\hat{I}_t - 1) \cdot G_S(x)]^{\hat{\mu} \cdot t/(1 - \hat{I}_t)} \tag{22}$$

La combinaison des relations (15) et (19) donne la liaison entre la période de retour $T_s(x)$ d'un quantile x et sa probabilité au non-dépassement $G_s(x)$:

$$G_s(x) = 1 - (1 - \exp[-1/(\gamma \cdot T_s(x)])/\delta \qquad (23)$$

qui devient :

$$G_s(x) = \{\hat{I}_t - \exp[(\hat{I}_t - 1)/(\hat{\mu} \cdot T_s(x))]\}/(\hat{I}_t - 1) \qquad (24)$$

La combinaison des relations (21) et (23) donne la liaison entre la période de retour $T_s(x)$ d'un quantile x et sa probabilité au non-dépassement $F_x(x)$:

$$F_x(x) = \exp[-t/T_s(x)] \qquad (25)$$

3.2.3 Cas de la loi de Poisson

La loi de Poisson est une limite de la loi Binomiale (cf. annexe I). Le résultat précédent montre que la loi du nombre d'événements ($X_s > x$) de la sous-classe est une loi de Poisson, puisque $\mu_s = \delta \cdot \gamma$. D'où :

$$N(x, t) = \mu_s t \cdot [1 - G_s(x)] \qquad (26)$$

$$\text{Prob [k valeurs } X_s > x \text{ pendant t]} = \exp[-N(x,t)] \cdot [N(x,t)]^k/k! \qquad (27)$$

$$\begin{aligned} F_X(x) &= \text{Prob [0 valeurs } X_s > x \text{ pendant t]} \\ &= \exp[-\mu_s t \cdot (1 - G_s(x))] \qquad \text{(d'après (27))} \end{aligned} \qquad (28)$$

Comme $T = 1/\mu$, où $\mu = dN(x,t)/dt$, il vient donc :

$$G_s(x) = 1 - 1/(\mu_s \cdot T_s(x)) \qquad (29)$$

La combinaison des relations (28) et (29) donne la même relation (25) que pour la loi Binomiale.

3.2.4 Cas de la loi Binomiale négative

La loi Binomiale Négative peut être caractérisée comme une loi de Poisson de moyenne μ', où μ' est une variable aléatoire suivant une loi Gamma, de paramètre d'échelle $\delta/(1 - \delta)$, de paramètre de forme γt. Pour chaque t et γ fixé :

1) tirer au hasard une valeur μ' aléatoire issue d'une loi gamma de densité :

$$g(\mu') = [\delta/(1-\delta)]^{\gamma t} \cdot [1/\Gamma(\gamma t)] \cdot \exp[-\delta\mu'/(1-\delta)] \cdot (\mu')^{\gamma t - 1}$$

2) engendrer un processus de Poisson sur [0;t] de paramètre μ' :

$$P_{\mu'}(k) = \exp[-\mu'] \cdot (\mu')^k/k\,!$$

Le résultat est un processus dont le nombre suit une loi Binomiale négative :

$$\begin{aligned} w_k(t) &= \text{Prob [k valeurs } X_s \text{ sur [0;t]]} \\ &= \int_0^{+\infty} P_{\mu'}(k) \cdot g(\mu') \cdot d\mu' \\ &= C^k_{\gamma t + k - 1} \delta^{\gamma t} \cdot (1 - \delta)^k \qquad \text{(formule (9))} \end{aligned}$$

Si on réalise la procédure d'effacement sur l'échantillon de valeurs sup-seuil ($X_s > x$), le processus restant est de Poisson, de paramètre $\mu' \cdot [1 - G_s(x)]$. D'où :

$$N(x,t) = \gamma t \cdot [(1-\delta)/\delta)] \cdot [1 - G_S(x)] \quad (30)$$

$$\text{Prob}[k \text{ valeurs } X_S > x \text{ pendant } t] = C^k_{\gamma t+k-1}(\delta/B)^{\gamma t} \cdot (1-\delta/B)^k \quad (31)$$

avec : $B = 1 - (1-\delta) \cdot G_S(x)$

$$F_X(x) = \text{Prob}\,[0 \text{ valeurs } X_S > x \text{ pendant } t]$$

$$= [\delta/(1-(1-\delta) \cdot G_S(x))]^{\gamma t} \quad \text{(d'après (31))} \quad (32)$$

qui donne, en remplacant γ et δ par leurs estimations $\hat{\gamma}$ et $\hat{\delta}$ la même relation (22) que pour la loi Binomiale.

La combinaison des relations (11) et (30) donne la liaison entre la période de retour $T_S(x)$ d'un quantile x et sa probabilité au non-dépassement $G_S(x)$:

$$G_S(x) = (1 - \delta \cdot \exp[1/(\gamma \cdot T_S(x)])/(1-\delta) \quad (33)$$

qui donne, en remplacant γ et δ par leurs estimations la même relation (23) que pour la loi Binomiale. La combinaison des relations (32) et (33) donne la même relation (25) que pour la loi Binomiale.

3.2.5 Cas du flux de Poisson non stationnaire

Ce cas a été développé en détail par NORTH (1980) et BERNIER (1981). On définit maintenant la probabilité de non-dépassement des valeurs sup-seuil par rapport à un temps donné : $G_S(x,t) = \text{Prob}\,[X_S < x$ à un instant t]. On obtient :

$$F_X(x,t) = \exp\left[-\int_0^t [1 - G_S(x,\tau)] \cdot \mu_S(\tau) \cdot d\tau\right] \quad (34)$$

$$N(x,t) = \int_0^t \mu_S(\tau) \cdot [1 - G_S(x,\tau)] \cdot d\tau \quad (35)$$

BERNIER (1981) a montré que, si la loi $G_S(x,t)$ conservait le même caractère au cours du temps (par exemple toujours une exponentielle simple, mais avec des paramètres évoluant avec le temps), il était possible de se ramener au cas du flux de Poisson, en estimant les paramètres de la loi $G_S(x)$ à partir de tous les événements sup-seuil, ce qui revient à moyenner les paramètres de la loi $G_S(x,t)$ sur la période disponible. On obtient alors :

$$G_S(x) = 1 - 1/(\tilde{T}_S(\tilde{x})) \quad (36)$$

$$F_X(x) = \exp[-\tilde{t}/\tilde{T}_S(x)], \text{ avec } \tilde{T}_S(x) = \int_0^{T_S(x)} \mu_S(\tau) \cdot d\tau \quad 2(37)$$

Le cas particulier d'un flux de Poisson stationnaire à l'échelle inter-annuelle et variations saisonnières a fait l'objet d'une abondante littérature : BORGMAN (1963) donne un exemple d'application de cette méthode pour l'évaluation du risque d'occurrence d'ouragans sur la côte est des États-Unis, connaissant la répartition saisonnière de ces ouragans. TODOROVIC et ZELENHASIC (1970), TODOROVIC et ROUSSELLE (1971), TODOROVIC et WOOLHISER (1972), GUPTA *et al.* (1976), TODOROVIC (1978), NORTH (1980), ASHKAR et ROUSSELLE (1981), BERNIER (1981), KONECNY et NACHTNEBEL (1985), NACHTNEBEL et KONECNY (1987) ont proposé différentes applications du flux de Poisson non stationnaire, dont on retiendra le modèle le plus récent, de NACHTNEBEL et KONECNY (1994), avec :

– une fonction périodique pour l'intensité du processus :

$$\mu(t) = \theta_0 \cdot \exp\left(\sum_{j=1}^{q} \theta_j \cdot \sin(2\pi j(t/365) + \theta_{q+j})\right)$$

– une loi exponentielle simple pour les valeurs sup-seuil, avec une fonction périodique pour le paramètre d'échelle :

$$G_s(x,t) = 1 - \exp[-\beta(t) \cdot x], \text{ avec } \beta(t) = \gamma_0 \cdot \exp\left(\sum_{k=1}^{r} \gamma_k \cdot \sin(2\pi k(t/365) + \gamma_{r+k})\right)$$

4 – CORRESPONDANCE ENTRE LES MODES D'ÉCHANTILLONNAGE

4.1 Correspondance entre les périodes de retour T_s et T_x

Dans l'absolu, les périodes de retour $T_x(x)$ et $T_s(x)$ ne sont pas comparables, puisqu'elles correspondent à deux variables aléatoires différentes : X^* et X_s. Ainsi, on note une différence d'interprétation entre les deux périodes de retour, évoquée par NADEN et BAYLISS (1993) : la période de retour T_x est la durée moyenne séparant deux années successives ayant leur maximum annuel supérieur à X (T_x), alors que la période de retour T_s est la durée moyenne séparant deux valeurs sup-seuil successives supérieures à X (T_s). La période de retour T_x est largement utilisée en hydrologie, du fait de l'emploi majoritaire de l'échantillonnage par maximum annuel, alors qu'elle semble assez artificielle dans sa définition. Dans la majorité des cas, l'usager est intéressé par l'estimation « d'un risque d'occurrence de tous les événements successifs », et donc par le quantile X (T_s). Le seul cas où l'usage de la période de retour T_x serait pertinent serait celui d'un usager ne tolérant pas plus d'une occurrence de l'événement par épreuve : ainsi une zone agricole avec des cultures annuelles où l'agriculteur devra attendre l'année suivante pour replanter si une inondation vient à endommager son champ. Ainsi, il semble préférable de présenter systématiquement les résultats avec la période de retour T_s, quel que soit le mode d'échantillonnage utilisé.

En combinant les relations (16) et (25) (obtenue dans le cas de la loi de Poisson, Binomiale négative ou Binomiale), on obtient la formule de correspondance (38), déjà développée pour un flux de Poisson par BORGMAN (1963) (eq. (47)) :

$$T_x/t = 1/(1 - \exp[-t/T_s]) = (T_s/t) + 1/2 + (t/T_s)/12 + 0(t/T_s)^2 \qquad (38)$$

Le tableau 1 montre que T_x tend rapidement vers $T_s + 1/2$, et X (T_x) tend en général vers X (T_s) à partir de $T_x = 10$ ans. Ce tableau avait déjà été présenté par LANGBEIN (1949), mais il avait été obtenu par une méthode rudimentaire, sans aucune mention du processus d'occurrence des valeurs supérieures à un seuil. TAKEUCHI (1984), par une approche analogue à celle de BORGMAN, a d'ailleurs clairement mis en évidence ce dernier point. Le terme en 1/2 de la relation (38) peut s'expliquer de la façon suivante. Le nombre N d'années séparant deux années successives avec une occurrence d'événement est relié à la partie entière de la durée de retour θ par :

$$N = \text{Ent}(\theta) \text{ ou } N = \text{Ent}(\theta) + 1,$$

suivant la position respective des deux événements dans l'année. Les deux alternatives étant équiprobables, le passage à l'espérance mathématique donne (pour T élevée) :

$$T_x = (1/2) \cdot T_s + (1/2) \cdot (T_s + 1) = T_s + 1/2$$

Tableau 1 Correspondance entre les périodes de retour T_s et T_x.

T_s/t	1	2	5	10	20	100
T_x/t	1,58	2,54	5,52	10,51	20,5	100,5

BERAN et NOZDRYN-PLOTNICKI (1977) ont comparé, sur 40 stations hydrométriques du Royaume-Uni, les périodes de retour T_s et T_x. La relation (38) n'est en général pas vérifiée :

$$T_x > 1/(1 - \exp[-1/T_s])$$

Ceci peut s'expliquer soit par la dépendance entre les valeurs supérieures à un seuil, soit par la dépendance des durées de retour. Les données expérimentales utilisées par les deux auteurs privilégient la deuxième explication, avec un calcul biaisé de la période de retour T = E(θ).

ROSBJERG (1987) a étudié la relation $T_s - T_x$ lorsque les valeurs supérieures à un seuil présentent une autocorrélation non nulle $\rho \neq 0$ (eq. (45) p. 12-13). Le cas général se situe entre les deux situations extrêmes :

– indépendance stricte ($\rho = 0$) : relation (38)

– dépendance complète ($\rho = 1$) : $T_x/t = (\mu_s/(1 - \exp[-\mu_s]) \cdot (T_s/t)$ (39)

Lorsque les événements sont indépendants ($\rho = 0$), il est possible d'utiliser la période de retour T sans préciser sa définition exacte (T_s ou T_x) seulement pour les événements rares (ROSBJERG, 1977a). Pour les événements fréquents, il est préférable de raisonner avec la période de retour T_s (avec une loi F_x ou G_s) et d'utiliser la relation (38) pour connaître la correspondance $T_s - T_x$.

Lorsque les événements sont complètement liés ($\rho = 1$), la relation (38) pose un problème d'interprétation des périodes de retour T_x et T_s. Soit toutes les valeurs de la chronique sont liées, auquel cas il n'y a qu'une occurrence d'événement ($X^* > x$) et aucune observation possible de l'intervalle de temps entre deux événements successifs (T_x n'est pas définie), soit les valeurs supérieures à un seuil de chaque année sont liées, auquel cas l'échantillonnage sup-seuil ne présente pas d'intérêt (T_s n'est pas définie).

Dans le cas général ($0 < \rho < 1$) ou lorsque le processus n'est pas homogène (existence de plusieurs saisons hydrologiques dans l'année), il est difficile de connaître la relation exacte entre T_s et T_x. Il est alors préférable de rester homogène dans le choix du mode d'échantillonnage et de la période de retour étudiée : si l'hydrologue s'intéresse à au plus une occurrence de crue par épreuve (période de retour T_x), il doit utiliser un échantillonnage par sélection de k maximums par saison ; s'il s'intéresse à toutes les occurrences successives, l'échantillonnage sup-seuil est à utiliser.

4.2 Correspondance entre les distributions F_x, G_{kx} et G_s

L'estimation d'un quantile X(T) peut s'effectuer suivant deux modes d'échantillonnage : sélection de k valeurs maximum par épreuve de durée fixe, ou sélection de la valeur maximum de chaque épisode sup-seuil. D'une façon générale, nous préconisons le deuxième mode d'échantillonnage, qui permet de sélectionner une population plus représentative du phénomène étudié.

– Pour les événements rares (T > 10), les deux modes d'échantillonnages aboutissent pratiquement aux mêmes résultats pour le calcul des quantiles, puisque $T_x/t \approx (T_s/t) + 1/2$. Des différences peuvent s'observer lorsque la chronique comporte des années sans occurrence d'événement ou au contraire avec une succession d'événements remarquables. Dans ces deux cas de figure, l'échantillonnage G_{kx} peut conduire à obtenir des valeurs peu représentatives (rajout de valeurs très basses les années « sèches », oubli de valeurs intéressantes les années « humides »).

– Pour les événements intermédiaires (1 < T < 10), il est nécessaire d'utiliser une correspondance entre les deux distributions. Nous résumons, dans le tableau 2, les formules de correspondance, obtenues en supposant un processus suivant une loi de Poisson, une loi Binomiale négative ou une loi Binomiale.

Tableau 2 Correspondance entre les distributions G_{kx} et G_s.

Échantillonnage sur une épreuve de durée t	Distribution G_{kx}	Distribution G_s	
Période de retour T_x	$(1 - t/T_x)^{1/k}$	$1 - 1/(\hat{\mu} \cdot T_{eq\,2})$	(a)
		$\{\hat{I}_t - \exp[(\hat{I}_t - 1)/(\hat{\mu} \cdot T_{eq\,2})]\}/(\hat{I}_t - 1)$	(b)
Période de retour T_s	$\exp(-t/k \cdot T_s)$	$1 - 1/(\hat{\mu} \cdot T_s)$	(a)
	ou : $1 - 1/(\hat{\mu} \cdot T_{eq\,1})$	$\{\hat{I}_t - \exp[(\hat{I}_t - 1)/(\hat{\mu} \cdot T_s)]\}/(\hat{I}_t - 1)$	(b)
(a) Flux de Poisson	(b) Loi Binomiale négative ou Binomiale		
$T_{eq\,1} = t/1 - \exp(-t/T_s))$	$T_{eq\,2} = -t/\mathrm{Log}(1 - t/T_x)$	$\hat{\mu} = \hat{E}(m_t)/t$	$\hat{I}_t = \hat{V}ar(m_t)/\hat{E}(m_t)$

– Pour les événements fréquents (T < 1), l'apport d'une loi de probabilité G_s est assez minime. La méthode du renouvellement donne directement une bonne estimation du quantile, sans introduire de loi de probabilité sup-seuil. Comme $G_s(x) = \mathrm{Prob}\,[X_s < x]$, où X_s est la valeur maximum d'un épisode où toutes les valeurs sont supérieures à S, on a $G_s\,[S] = 0$. On obtient une estimation du quantile $X(T_s)$ avec $\hat{X}(T_s) = S$, où S est recherché par approximations successives de façon à avoir $\hat{\mu} = 1/T_s$ (flux de Poisson), ou $(\hat{I}_t - 1)/(\hat{\mu} \cdot \mathrm{Log}\hat{I}_t) = T_s$ (loi Binomiale négative ou loi Binomiale).

4.3 Comparaison de la distribution d'échantillonnage des quantiles des lois G_{kx} and G_s

CUNNANE (1973), TAVARES et DA SILVA (1983), ROSBJERG (1985), WANG (1991), MADSEN *et al.* (1997) ont comparé la variance d'échantillonnage des quantiles issus d'un échantillonnage par maximum annuel (F_X) ou par valeurs sup-seuil (G_S). Ils ont trouvé un léger avantage de la loi F_x vis-à-vis de la loi G_s lorsque $\mu_S = 1$; un avantage de la loi G_s sur la loi F_x lorsque $\mu_S > 1{,}7$, une similitude de performances quand la période de retour devient élevée (T > 10 ans). BUISHAND (1989) a comparé de façon théorique l'écart quadratique moyen (EQM) des quantiles de crues issus du binôme (Loi de Poisson ; Loi exponentielle simple), obtenus par échantillonnage sup-seuil de deux façons différentes, soit en sélectionnant toutes les valeurs supérieures à un seuil, soit en retenant les NEV plus forts événements sur NEP épreuves. Les valeurs théoriques EQM étant du même ordre, Buishand conclut à l'équivalence des deux approches. Il est toutefois préférable d'utiliser la première, qui donne implicitement des valeurs sup-seuils séparées par des épisodes inférieurs au seuil, alors que la deuxième nécessite de rajouter une condition d'espacement entre les NEV plus forts événements.

LANG (1995b) a complété ces résultats par une analyse comparative des quantiles issus de lois G_{kx} et G_s (avec $k = \mu_S$ variant de 1 à 5). Le critère de comparaison est l'écart EQM des quantiles issus de la paire (Poisson ; exponentielle simple) ou d'une loi de Gumbel (k maximums par an). L'échantillonnage G_S donne en moyenne un écart EQM deux à trois plus faible que celui obtenu avec l'échantillonnage G_{kx}.

On retiendra de ces simulations un net avantage à utiliser l'échantillonnage G_s plutôt que l'échantillonnage classique F_X. S'il y a un léger avantage de la loi F_x sur la loi G_S (avec $\mu_S = 1$), il suffit de baisser le seuil pour améliorer les résultats de la loi G_S, alors que l'augmentation du nombre de valeurs de la loi G_{kx} n'apporte globalement pas d'amélioration, en fait plutôt une détérioration des résultats.

5 – CONCLUSION

La technique de l'échantillonnage par valeurs supérieures à un seuil a fait l'objet de nombreuses publications. On retiendra des travaux antérieurs la possibilité de modéliser toute une gamme de processus à l'aide de la méthode du renouvellement. Les éléments nouveaux apportés ici sont d'une part la formulation complète du cas où le processus suit une loi Binomiale négative et une loi Binomiale, d'autre part la comparaison des échantillonnages G_{kx} et G_S, jusqu'alors limitée à la sélection d'une valeur par an (k = 1 et $\mu_S = 1$). Il faut insister sur le caractère artificiel de la période de retour T_X liée à l'échantillonnage classique F_X, où « T_X est la durée moyenne séparant deux années successives ayant leur maximum annuel supérieur à X (T_X) », alors que la période de retour T_S liée à l'échantillonnage G_S, où « T_S est la durée moyenne séparant deux valeurs sup-seuil successives supérieures à X(T_S) », correspond beaucoup mieux à la demande sociale de prédétermination du risque.

Un aspect complémentaire, non abordé ici, concerne l'utilisation conjointe de l'échantillonnage sup-seuil et de l'analyse bayésienne : WOOD et RODRIGUEZ-ITURBE (1975a et b) ont étudié l'incertitude due au choix d'un modèle probabiliste, ROUSSELLE et HINDIE (1976) l'incertitude sur les paramètres des lois de probabilité, MIQUEL (1984) la prise en compte de l'information des crues historiques, RASMUSSEN et ROSBJERG (1991), MADSEN *et al.* (1994) puis MADSEN et ROSBJERG (1995b) la prise en compte d'une information régionale, BERNIER (1981) puis ABI-ZEID (1995) l'estimation de l'intensité d'un processus de Poisson non stationnaire par méthode bayésienne.

Pour finir, il faut noter qu'une des difficultés de l'échantillonnage par valeurs supérieures à un seuil vient du fait que la prise en compte de l'aspect processus des crues introduit des paramètres supplémentaires à déterminer. LANG *et al.* (1997) présentent une revue bibliographique sur les tests ou méthodes disponibles pour aider au choix de ces paramètres.

REMERCIEMENTS

Les auteurs remercient le Professeur Jacques BERNIER de ses remarques, et notamment de sa suggestion de déduire les propriétés exposées au paragraphe 3.2 à l'aide d'un effacement aléatoire. La démarche antérieure est exposée en annexe 2.

NOTATIONS

$B_k(\delta,\gamma t)$: Loi Binomiale, de paramètres δ et γt.

E : Événement.

E(X) : Espérance de la variable aléatoire $X; E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) \cdot dx$.

Ent(x) : Partie entière de x.

EQM : Écart Quadratique Moyen de l'estimateur de $\hat{X}$ de X ; $EQM^2 = b^2 + d^2$, où $b = E(\hat{X}) - X$ est le biais de l'estimateur, et $d = (\mathrm{Var}\ \hat{X})^{1/2}$ est l'écart-type de l'estimateur.

f(x) : Densité de probabilité d'une variable.

F(x) : Fonction de répartition d'une variable.

F_x : Fonction de répartition de la variable X^*.

G_{kx} : Fonction de répartition de la variable X_k.

G_s : Fonction de répartition de la variable X_s.

I_t : Indice de dispersion ; $I_t = \mathrm{Var}(m_t)/E(m_t)$.

m_t : Nombre d'événements survenus dans l'intervalle [0;t] ; de fonction de répartition $w_k(t)$, d'espérance N(t)

$N(t)$: Nombre moyen d'événements sur $[0;t]$; $N(t) = E(m_t)$.

$N(t,t')$: Nombre moyen d'événements sur $[t;t']$; $N(t,t') = \int_t^{t'} \mu(\tau) \cdot d\tau$.

$N(x,t)$: Nombre moyen d'événements dépassant la valeur x sur $[0;t]$.

T : Période de retour $T = E(\theta)$; espérance de la durée de retour θ.

T_s : Période de retour de la variable X_s.

T_x : Période de retour de la variable X^*.

Var (X) : Variance de la variable aléatoire X ; $\mathrm{Var}(X) = \int_{-\infty}^{+\infty} (x - E(x))^2 \cdot f(x) \cdot dx$.

$w_k[0;t]$: Fonction de répartition de m_t ; notée $w_k(t)$ lorsque le processus est stationnaire.

X : Variable aléatoire.

$\hat{X}$: Estimation de la variable X.

$\tilde{X}$: Changement d'échelle sur la variable X ; $\tilde{X} = \int_0^X \mu(\tau) \cdot d\tau$.

X^* : Valeur maximum de X sur une épreuve de durée fixe.

X_k : Une des k plus fortes valeurs de X sur une épreuve de durée fixe.

X_s : Valeur maximum de X sur un épisode supérieur au seuil S.

$X(T)$: Quantile de la variable aléatoire X, associé à la période de retour T.

α : Paramètre d'échelle de la fonction de répartition de la durée de retour θ.

δ, γ : Paramètres de la loi Binomiale ou Binomiale négative.

$\Gamma(x)$: Fonction Gamma ; $\Gamma(n) = (n-1)!$ si n est entier ;

$$\Gamma(z) = \int_0^{+\infty} x^{z-1} \cdot e^{-x} \cdot dx .$$

$\mu(t)$: Paramètre décrivant l'intensité du processus ; $\mu(t) = dN(t)/dt$, noté μ si le processus est stationnaire.

$\mu_s(t)$: Paramètre $\mu(t)$ lié au processus des événements sup-seuil X_s ; noté μ_s si le processus est stationnaire.

θ : Durée de retour, séparant deux occurences successives d'un événement ; notée $\theta(t)$ si le processus est non stationnaire.

ρ : Coefficient d'autocorrélation de la série des maximums.

ANNEXE 1 : ÉTUDE DE LA DURÉE DE RETOUR

Flux de Poisson

Les 4 hypothèses définissant le flux de Poisson s'écrivent :

(i) $w_k\,[0;t]$ indépendant de la position de $[0;t]$ dans le temps.

(ii) $w_1\,[0;dt] = \mu \cdot dt$

(iii) $\sum_{k=2}^{\infty} w_k[0;dt]$ négligeable devant dt

(iv) $w_k\,[t_1;t_2]$ et $w_k\,[t_3;t_4]$ sont indépendants si les intervalles de temps respectifs sont disjoints.

On peut établir une relation de récurrence entre $w_k[0;t]$ et $w_{k-1}[0;t]$:

$$w_k\,[0;t+dt] = \sum_{j=0}^{k} w_{k-j}\,[0;t] \cdot w_j[t;\,t+dt] \qquad \text{à cause de (iv)}$$

$$\approx w_k[0;\,t] \cdot w_0\,[t;t+dt] + w_{k-1}\,[0;t] \cdot w_1[t;t+dt] \qquad \text{à cause de (iii)}$$

On obtient $w_1[t;t+dt] = \mu \cdot dt$, à cause de (ii) et (i), et :

$$w_0[t;t+dt] = 1 - \sum_{k=1}^{\infty} w_k[t;t+dt]$$

$$\approx 1 - w_1[t;t+dt] \qquad \text{à cause de (iii)}$$

$$= 1 - \mu \cdot dt$$

Il vient donc :

$$(w_k[0;t+dt] - w_k[0;t])/(\mu \cdot dt) = w_{k-1}[0;t] - w_k[0;t]$$

soit :

$$(1/\mu) \cdot w'_k[0;t] = w_{k-1}[0;t] - w_k[0;t] \qquad (k > 1) \qquad \text{(A.1)}$$

Pour $k = 0$, on a : $w_0[0;t+dt] = w_0[0;t] \cdot w_0[t;t+dt]$

soit :

$$w'_0[0;t] = -\mu \cdot w_0[0;t]$$

Comme $w_0[0;0] = 1$, on a $w_0[0;t] = \exp\,[-\mu \cdot t]$. Si on pose $w_k[0;t] = u_k[0;t] \cdot \exp\,[-\mu \cdot t]$, (A.1) devient :

$$u'_k[0;t] = \mu \cdot u_{k-1}[0;t]$$

On arrive facilement à :

$$w_k[0;t] = \exp\,[-\mu t] \cdot (\mu t)^k/k! \qquad \text{(A.2)}$$

On s'intéresse maintenant à la durée séparant deux occurrences successives d'un événement. Soient E_k l'événement survenant à $t = t_k$, E_{k+1} l'événement survenant à $t = t_{k+1}$, et $\theta_k = t_{k+1} - t_k$ la durée séparant deux occurrences successives.

$$\begin{aligned}\text{Prob}\,[\theta_k > d] &= \text{Prob}\,[0 \text{ événement sur } [t_k;\, t_{k+d}]]\\ &= w_0[t_k;\, t_{k+d}]\\ &= w_0[0;d] \qquad \text{à cause de (i)}\end{aligned}$$

soit :

$$F(d) = \text{Prob}[\theta < d] = 1 - w_0[0;\, d] = 1 - \exp[-\mu \cdot d] \qquad (A.3)$$

$$f(\theta) = \mu \cdot \exp\,[-\mu \cdot \theta] \qquad (A.4)$$

Loi Binomiale négative

$$w_k[0;t] = C^k_{\gamma t + k - 1}\delta^{\gamma t} \cdot (1-\delta)^k \qquad (A.5)$$

avec : $E\,[m_t] = \gamma t\,(1 - \delta)/\delta$ et $\text{Var}\,[m_t] = \gamma t\,(1 - \delta)/\delta^2$

Quand l'indice de dispersion I_t tend vers 1, c'est-à-dire lorsque $\delta \to 1^-$ et $\gamma \to +\infty$, la loi Binomiale négative tend vers la loi de Poisson. En effet, si on note $\hat{\mu} = E(m_t)/t$, δ est estimé par $\hat{\delta} = 1/(1 + \hat{\mu}/\gamma)$, et on a :

$$w_k[0;t] = (\gamma t + k - 1) \cdot (\gamma t + k - 2)\ldots(\gamma t) \cdot \delta^{\gamma t} \cdot (1-\delta)^k/k! \qquad \text{à cause de (A.5)}$$

$$\begin{aligned}\lim_{\gamma \to +\infty} w_k[0;t] &\approx (\gamma t)^k \cdot (1 + \hat{\mu}/\gamma t)^{-\gamma t} \cdot (\hat{\mu}/\gamma t)^k \cdot (1 + \hat{\mu}/\gamma t)^{-k}/k!\\ &\approx (\hat{\mu})^k \cdot (1 + \hat{\mu}/\gamma t)^{-\gamma t}/k!\\ &\approx \exp\,[-\hat{\mu}] \cdot (\hat{\mu})^k/k!\end{aligned}$$

On retrouve l'expression (A.2) de la loi de Poisson.

De façon analogue au flux de Poisson, on peut calculer la densité de probabilité de la durée θ :

$$F(d) = 1 - w_0[0;d] = 1 - \delta^{\gamma d} \qquad (A.6)$$

$$f(\theta) = -\gamma \cdot \text{Log}\delta \cdot \delta^{\gamma\theta} \qquad (A.7)$$

Loi Binomiale

$$w_k[0;t] = C^k_{\gamma t}\delta^k \cdot (1-\delta)^{\gamma t - k} \qquad (A.8)$$

avec : $E[m_t] = \gamma t \cdot \delta$ et $\text{Var}\,[m_t] = \gamma t \cdot \delta\,(1 - \delta)$.

Lorsque l'indice de dispersion I_t tend vers 1, c'est-à-dire lorsque $\delta \to 0^+$ et $\gamma \to +\infty$, la loi Binomiale tend vers la loi de Poisson. En effet, si on note $\hat{\mu} = E(m_t)/t$, δ est estimé par $\hat{\delta} = \hat{\mu}/\gamma$ et on a :

$$w_k\,[0;t] = (\gamma t) \cdot (\gamma t - 1)\ldots(\gamma t - k + 1) \cdot \delta^k \cdot (1-\delta)^{\gamma t - k}/k! \qquad \text{à cause de (A.8)}$$

$$\begin{aligned}\lim_{\gamma \to +\infty} w_k[0;t] &\approx (\gamma t)^k \cdot (\hat{\mu}/\gamma t)^k \cdot (1 - \hat{\mu}/\gamma t)^{\gamma t - k}/k!\\ &\approx (\hat{\mu})^k \cdot (1 - \hat{\mu}/\gamma t)^{\gamma t}/k!\\ &\approx \exp\,[-\hat{\mu}] \cdot (\hat{\mu})^k/k!\end{aligned}$$

On retrouve l'expression (A.2) de la loi de Poisson.

De façon analogue au flux de Poisson, on peut calculer la densité de probabilité de la durée θ :

$$F(d) = 1 - w_0\,[0;d] = 1 - (1-\delta)^{\gamma d} \tag{A.9}$$

$$f(\theta) = -\gamma \cdot \mathrm{Log}\,(1-\delta) \cdot (1-\delta)^{\gamma\theta} \tag{A.10}$$

ANNEXE 2 : LIAISON ENTRE LA PÉRIODE DE RETOUR D'UN QUANTILE X ET LA PROBABILITÉ AU DÉPASSEMENT DU NIVEAU X

Flux de Poisson

La combinaison des relations (1) et (18) donne la correspondance (28), décrite pour la première fois par SHANE et LYNN (1964). La liaison entre la période de retour $T_s(x)$ d'un quantile x et sa probabilité au non-dépassement $G_s(x)$ peut s'obtenir de la façon suivante :

Prob [k valeurs $X_s > x$ pendant t]

$= \sum_{i=0}^{\infty}$ Prob [(k + i) valeurs X_s avec k valeurs supérieures à x pendant t et i valeurs inférieures à x pendant t]

$$= \sum_{i=0}^{\infty} C_{k+i}^{i} w_{k+i}(t) \cdot [1 - G_s(x)]^k \cdot [G_s(x)]^i$$

$$= \sum_{i=0}^{\infty} \frac{(k+i)!}{k! \cdot i!} \cdot \frac{\exp[-\mu_s \cdot t] \cdot (\mu_s \cdot t)^{(k+i)}}{(k+i)!} \cdot [1 - G_s(x)]^k \cdot [G_s(x)]^i$$

$= \exp[-N(x,t)] \cdot [N(x,t)]^k / k!$ (formule (27))

Loi Binomiale négative

La combinaison des relations (9) et (18) donne la correspondance (32), décrite par VUKMIROVIC et PETROVIC (1995). La liaison entre la période de retour $T_s(x)$ d'un quantile x et sa probabilité au non-dépassement $G_s(x)$ peut s'obtenir de la façon suivante :

Prob [k valeurs $X_s > x$ pendant t]

$$= \sum_{i=0}^{\infty} C_{k+i}^{i} C_{k+i+\gamma t-1}^{k+i} \delta^{\gamma t} \cdot (1-\delta)^{k+i} \cdot [1 - G_s(x)]^k \cdot [G_s(x)]^i$$

$$= C_{k+\gamma t-1}^{k} \delta^{\gamma t} \cdot (B-\delta)^k \cdot B^{-(k+\gamma t)} \cdot \sum_{i=0}^{\infty} C_{i+k+\gamma t-1}^{i} (1-B)^i \cdot B^{k+\gamma t}$$

(avec $B = 1 - (1-\delta) \cdot G_s(x)$)

$$= C^{k}_{\gamma t+k-1}(\delta/B)^{\gamma t}\cdot((1-\delta)/B)^{k}$$ (formule (31)).

Loi Binomiale

La combinaison des relations (13) et (18) donne la correspondance (21) décrite par VUKMIROVIC et PETROVIC (1995). La liaison entre la période de retour $T_s(x)$ d'un quantile x et sa probabilité au non-dépassement $G_s(x)$ peut s'obtenir de la façon suivante :

Prob [k valeurs $X_s > x$ pendant t]

$$= \sum_{i=0}^{\infty} C^{i}_{k+i} w_{k+i}(t)\cdot[1-G(x)]^{k}\cdot[G_s(x)]^{i}$$

$$= \sum_{i=0}^{\gamma t-k} C^{i}_{k+i} C^{k+i}_{\gamma t}\delta^{k+i}\cdot(1-\delta)^{\gamma t-k-i}\cdot[1-G_s(x)]^{k}\cdot[G_s(x)]^{i}$$

$$= C^{k}_{\gamma t}A^{k}\cdot(1-\delta)^{\gamma t-k}\cdot\sum_{i=0}^{\gamma t-k} C^{i}_{\gamma t-k}[(\delta-A)/(1-\delta)]^{i}$$ (avec $A = \delta\cdot(1-G_s(x))$)

$$= C^{k}_{\gamma t}A^{k}\cdot(1-\delta)^{\gamma t-k}\cdot[1+(\delta-A)/(1-\delta)]^{\gamma t-k}$$

$$= C^{k}_{\gamma t}A^{k}\cdot(1-A)^{\gamma t-k}$$ (formule (20)).

*

* *

1 – INTRODUCTION

Flood frequency analysis is often carried out by fitting a probability distribution to a sample of maximum annual floods (ASHKAR *et al.*, 1994). An alternative method consists of retaining all flood peaks in a time-series that exceed a given threshold (peaks-over-threshold or partial duration series). This method involves the use of two distributions, one describing the number of events, the other describing the magnitude of events. The two distributions are combined to obtain the distribution of annual maximum floods.

The present paper is concerned with the modeling of occurrences of events for use in the method of renewal processes (COX, 1965; FELLER, 1966). The first applications of this method in hydrology must be credited to BORGMAN (1963), SHANE and LYNN (1964), and BERNIER (1967). The interested reader may consult PICKANDS (1975), and DAVISON and SMITH (1990) for a more theoretical treatment of exceedance modeling. Operational aspects concerning the choice of exceedance distribution (two or three parameters), the choice of threshold, and the independence of successive exceedances have been studied by ROSBJERG *et al.* (1991, 1992), ROSBJERG and MADSEN (1992), and LANG (1995a).

In the following, we review different distributions used to describe the occurrence of events, and discuss the relationship between the distribution of a variable and the corresponding return period of a quantile. These relationships depend on the type of sampling considered: a fixed number of events per sampling interval or over-threshold sampling. Finally, we explain how the two sampling methods may be compared.

2 – MODELING OF RENEWAL PROCESSES

We consider in the following a process of occurrences of events E, described either by the time θ between successive events, denoted waiting time, or by the number of events m_t in the interval [0 ; t]. Each variable may be characterized by its distribution function, its density, and its mean value

– waiting time θ: $F(d) = \text{Prob}[\theta < d]$ and $f(x)dx = \text{Prob}[x < \theta < x + dx]$

We assume that $F(0) = 0$, so that it is impossible to have two simultaneous events. The return period of the event is defined as

$$T = E(\theta) = \int_0^{+\infty} \theta f(\theta) d\theta$$

– number of events m_t in $w_t[0;t] = \text{Prob}[m_t = k]$

We define in the same way the mean number of events N(t) in [0;t]: $N(t) = E(m_t)$; the intensity of the process: $\mu(t) = dN(t)/dt$; and a dispersion index defined as: $I_t = \text{Var}(m_t)/E(m_t)$

2.1 Stationary Poisson process

We suppose that the occurrence process satisfies four hypotheses:

(i) homogeneity in time;

(ii) the probability of having one event in a short period dt is small and of the same order as dt;

(iii) the probability of having more than one event in a short period dt is negligible; and

(iv) successive events are independent.

It can be shown (BASS, 1974, p. 145-148, cf. Appendix 1) that these hypotheses lead to the following relationships:

$$w_k[0;t] = \frac{(\mu t)^k \exp(-\mu t)}{k!} \tag{1}$$

$$F(d) = \mathrm{Prob}(\theta < d) = 1 - \exp(-\mu d) \tag{2}$$

Therefore the number of events E in the time interval [0;t] is Poisson distributed with mean $N(t) = \mu t$ and variance $\mathrm{Var}(m_t) = \mu t$, c.f. (1). The intensity of the process $\mu(t)$ is constant and equal to μ. The waiting time θ separating two events is exponentially distributed (eq. 2), the return period is equal to $T = 1/\mu$, and the dispersion index is equal to 1 ($I_t = 1$). If the parameter μ is replaced by its estimator $\hat{\mu} = \hat{E}(m_t)/t$ we obtain the following relationships:

$$F(d) = 1 - \exp[-d/\hat{\alpha}] \tag{3}$$

$$T = E(\theta) = \hat{\alpha} \tag{4}$$

where

$$1/\hat{\alpha} = \hat{\mu} \tag{5}$$

2.2 Non-stationary Poisson process

We assume in this case that the occurrence of events is not homogeneous in time, that is, only the last three hypotheses leading to the homogeneous Poisson process are retained. In the non-stationary case hypothesis (ii) states that $w_1[t;t+dt] = \mu(t)dt$. By considerations similar to those above for the homogenous Poisson process (VENTSEL, 1973, p.510-511) one is led to the following relationships:

$$w_k[t;t'] = \frac{[N(t,t')]^k \exp[N(t,t')]}{k!} \tag{6}$$

where $N(t;t') = \int_t^{t'} \mu(\tau)d\tau$ represents the mean number of events during the time interval $[t;t']$.

$$F_t(d) = \mathrm{Prob}[\theta(t) < d] = 1 - \exp\left[-\int_t^{t+d} \mu(\tau)d\tau\right] \tag{7}$$

The number of events E in the time interval [t; t'] is Poisson distributed with mean N (t;t'). The intensity of the process depends on time:

$$\lim_{\Delta t \to 0^+} N(t,t+\Delta t)/\Delta t = \mu(t)$$

BORGMAN (1963) has studied in more detail the case of a non-stationary process with seasonal variations, but inter-annual stability. In that case, it is possible to obtain a stationary Poisson process by an appropriate change of time scale:

$$\tilde{t} = \int_0^t \mu(\tau)d\tau$$

We then obtain:

$$w_k[t_1;t_2] = \frac{(\tilde{t}_2-\tilde{t}_1)^k \exp[-(\tilde{t}_2-\tilde{t}_1)]}{k!} \tag{8}$$

(reduced Poisson process with $\mu = 1$)

$$\tilde{T} = 1 \text{ ,that is: } \int_0^T \mu(\tau)d\tau = 1$$

2.3 Negative binomial distribution

Several authors (CUNNANE, 1979; MIQUEL, 1984; BOIRET, 1987; VUKMIROVIC, 1990; BEN-ZVI, 1991) have proposed the negative binomial distribution for describing the occurrence of events:

$$w_k[0;t] = C^k_{\gamma t+k-1}\delta^{\gamma t}(1-\delta)^k \tag{9}$$

where δ is a parameter whose value is between 0 and 1. The coefficient $C^k_{\gamma t+k-1}$ is computed as a ratio of factorials expressed as gamma functions.

The mean value of m_t is $N(t) = \gamma t(1-\delta)/\delta$, and its variance is $Var(m_t) = \gamma t(1-\delta)/\delta^2$ (eq. 9). The dispersion index is greater than one: $I_t = 1/\delta > 1$. It can be shown that the waiting time has the following properties (cf Appendix 1):

$$F(d) = Prob[\theta < d] = 1 - \delta^{\gamma d} \tag{10}$$

$$T = E(\theta) = \int_0^{+\infty} (-\gamma \cdot \log\delta \cdot \delta^{\gamma\theta})\theta \cdot d\theta = -1/(\gamma \log\delta) \tag{11}$$

If one replaces the parameters θ and δ by their estimators $\hat{\gamma} = \hat{\mu}/(\hat{I}_t - 1)$ and $\hat{\delta} = 1/\hat{I}_t$, the above equations become identical to equations (3) and (4), with

$$1/\hat{\alpha} = \hat{\mu}(\log\hat{I}_t/(\hat{I}_t - 1)) \tag{12}$$

This distribution constitutes an alternative to the Poisson process when the dispersion index is greater than 1.

2.4 *Binomial distribution*

VUKMIROVIC (1990) also proposed the binomial distribution:

$$w_k[0;t] = C_{\gamma t}^k \delta^k (1-\delta)^{\gamma t - k} \tag{13}$$

The mean and the variance of m_t are, respectively, $N(t) = \gamma t \delta$ and $Var(m_t) = \gamma t \delta(1-\delta)$ (eq. 13). The dispersion index is smaller than 1: $I_t = 1 - \delta < 1$. It can be shown (see Appendix 1) that the waiting time has the following properties:

$$F(d) = Prob[\theta < d] = 1 - (1-\delta)^{\gamma d} \tag{14}$$

$$T = E(\theta) = \int_0^{+\infty} (-\gamma \log(1-\delta)(1-\delta)^{\gamma\theta})\theta d\theta = -1/(\gamma \log(1-\delta)) \tag{15}$$

If the parameters γ and δ are replaced by their estimators $\hat{\gamma} = (\hat{\mu}/(1 - \hat{I}_t))$ and $\hat{\delta} = 1 - \hat{I}_t$ the above expressions become identical to (3) and (4), with the same value of the parameter α as in the case of the negative binomial distribution (eq. 12).

This distribution constitutes an alternative to the Poisson process when the dispersion index in smaller than 1.

2.5 *Other distributions*

Other distributions have been proposed to describe more complex processes. BERNIER (1967) presented certain generalizations of the renewal process: a k-order Erlang process derived from the Poisson process by considering only one event for every (k + 1) events, generalized Erlang distributions where the order k is arbitrary, *i.e.* not limited to integer values, a delayed processes where the time origin is fixed between two events, a cumulative process where a secondary variable is associated with the basic variable, an alternating process where the variable in consideration can take one of two values (see also KUNDZEWICZ, 1989), as well as a semi Markov process (generalization of the alternating process with several states). KARR (1976), ROSBJERG (1977b), and MADSEN and ROSBJERG (1995a) studied the occurrence of events when the underlying process is MARKOVIAN (lag-1 autocorrelation taken into account). DAVISON and SMITH (1990) cite the work of LEADBETTER (1983), LINDGREN and ROOTZEN (1987), LEADBETTER and ROOTZEN (1988) in which Poisson processes with groups of dependent events are considered. KAVVAS (1982a, b) and CERVANTES *et al.* (1983) proposed an interesting two-level non-stationary Poisson process: at the first level, the occurrence of a meteorological perturbation is modeled, at the second level several events pertaining to a given perturbation are modeled.

3 – ANNUAL MAXIMUM OR PEAKS-OVER-THRESHOLD SAMPLING

The most commonly used sampling method for flood risk analysis consists of retaining the largest flood of each year. ASHKAR *et al.* (1994) discuss several problems related to this type of analysis: tests of basic hypotheses, criteria for choice of distribution and estimation method, regionalisation, etc. An alternative is to consider all values that exceed a threshold using renewal processes to describe the occurrence of events above the threshold. RASMUSSEN (1991), ROSBJERG (1993), RASMUSSEN *et al.* (1994), and LANG (1995a) provide state-of-the-art reviews of research on this subject. We present below the main characteristics of the two types of sampling and the relationships that exist between them.

3.1 Sampling a fixed number of maximums per period

3.1.1 Selection of one maximum per period

Let X be a random variable and X* another random variable defined as the maximum of X in a given period (year, season, month). The length of the period from which the maximum X* is extracted is denoted t. We define the occurrence of an event as the exceedance of a given level x by X*, *i.e.* X*> x, and associate the following elements with the event:

- Return period: $T_x(x)$, where index x indicates sampling of maximums;
- Quantile: $X(T_x)$, *i.e.* the value of X = x that has return period $T_x(x)$;
- Non-exceedance probability: $F_x(x) = \text{Prob}[X^* < x]$.

Assuming that events arise from a Poisson process, one can obtain the relationship between $F_x(x)$ and $T_x(x)$ as follows. First note that

Prob [k events in N periods] = $C_N^k[F_x(x)]^{N-K}[1-F_x(x)]^k$

which is a binomial distribution with mean $N[1-F_x(x)]$. The mean number of events in the interval [0;t] is

$$N(x, t) = t[1 - F_x(x)]$$

Since the return period of an event described by a Poisson process is $T = 1/\mu$ where $\mu = dN(x,t)/dt$, we get:

$$F_x(x) = 1 - t/Tx\,(x) \qquad (16)$$

3.1.2 Selection of k maximums per period

Let again X be a random variable, and X_k one of the k largest observations in a period of length t. We define the occurrence of an event as the exceedance of x by one of the k largest values observed in the period of length $t(X_k > x)$, and associate the following elements with the event:

- Return period: $T_x(x)$, where index *x* indicates sampling of maximums;
- Quantile: $X(T_x)$, *i.e.* the value of X = x that has return period $T_x(x)$;
- Non-exceedance probability: $G_{kx}(x) = \text{Prob}[X_k < x]$.

One readily obtains the following relationships between F_x and G_{kx}, and G_{kx} and T_x:

$$\begin{aligned} F_x(x) &= \text{Prob}[X^* < x] \\ &= \text{Prob}[(X_k)_1 < x \text{ and } (X_k)_2 < x \text{ and } (X_k)_k < x] \\ &= [G_{kx}(x)]^k \text{ (assuming that the k largest values are independent)} \end{aligned}$$

which gives

$$G_{kx}(x) = (1 - t/T_x\,(x))^{1/k} \qquad (17)$$

Equation (17) was obtained assuming independence of the values of X_k which leads us to reject, among the k largest values, observations belonging to the same hydrological event. The implementation of a criterion for testing independence is often difficult; in practice, one usually imposes a minimum time span between successive events.

The above sampling scheme allows us to increase the number of observations. However, as in the previous case, it has the inconvenience of considering a fixed number of peaks per period, without paying any attention to the process that generates the events. Since some years are dry and others humid, this type of sampling leads to populations that are not always homogeneous, with selection of dry-year extremes that are of little interest and non-selection of wet-year extremes that may contain valuable information. For this reason, we recommend use of the following sampling scheme in which the largest values in the whole time series are considered.

3.2 Peaks-over-threshold sampling

3.2.1 General formulation

Let X be a random variable and X_s its maximum during a threshold exceedance. In the present context, a threshold exceedance event is defined with reference to the threshold S: the event begins when X > S and ends when X < S. Let the occurrence of an event be defined as the exceedance of a given level x by the maximum X_s of the threshold exceedance event. We shall associate the following elements with the event:

- Return period: $T_s(x)$, where index s indicates sampling of threshold exceedances;
- Quantile: $X(T_s)$, *i.e.* the value of X = x that has return period $T_s(x)$;
- Non-exceedance probability: $G_s(x) = \text{Prob}[X_s < x]$

The distributions F_x is related to w_k (t) and G_s in the following way:

$F_x(x) = \text{Prob}[X^* < x]$, where X^* is the maximum of X in a time period of length t

$w_k(t) = \text{Prob}[k$ exceedances of the threshold S in t years$]$

$G_s(x) = \text{Prob}[X_s < x]$ where X_s is the maximum value of a threshold exceedance event

$$F_x(x) = \sum_{k=0}^{\infty} w_k(t)[G_s(x)]^k \tag{18}$$

(assuming independence of the maximums of different exceedance events).

It should be emphasized that contrary to our presentation of occurrence processes there is now two types of events involved: the event "X exceeding the threshold S", called a threshold exceedance event, with which we associate the distribution $w_k(t)$; and the event "X exceeding the level x", simply called an event which has the distribution $G_s(x)$. The introduction of the distribution G_s will allow us to extrapolate the empirical distribution to high return periods which is not possible with the classical renewal method. In fact, the return period of an event E is $T(E) = 1/\mu$ where μ is the mean number of observed events per time period. In the case of rare events, one cannot reliably estimate μ.

There are two difficulties related to the use of over-threshold sampling: one concerns the introduction of new parameters describing the distribution of floods, the other concerns the underlying hypotheses of the sampling technique. The first aspect will be discussed in section 4 where we evaluate the effect of uncertainty in the modeling of the process on quantile estimates. The second aspect is a function of the type of process considered. In the case of the Poisson process, the two most critical hypotheses concern the independence of successive events and the stationarity of the process. The choice of threshold will in general be a compromise between a large number of events of which some may be correlated (low threshold) and a small number of independent events (high threshold). If the hypotheses of the Poisson process are not respected, it is recommended to try the binomial and negative binomial distributions before turning to more complex models. A seasonal periodicity in the process can be treated by considering the maximums in each identified season or by using a non-stationary Poisson process.

We elaborate on the four occurrence processes described previously, although other distributions could also be considered (see section 2.5).

3.2.2 Binomial distribution

The statistical properties of the number of exceedances of the level x (x > S) may be deduced from the number of exceedances of the threshold S. In fact, by raising the level from S to x, one randomly censors the sample so that with probability $G_s(x)$ an event is eliminated, and with probability $1 - G_s(x)$ it is retained. The process of the events retained is therefore a subclass (well-defined when x is fixed) of the basic occurrence process.

Assume that the occurrence of the events X_s follows a binomial distribution:

$$B_k(\delta', \gamma t) = \text{Prob}[k \text{ values } X_s \text{ of a total of } \gamma t; \text{ with } \delta = \text{Prob}(X_s > x)]$$

with a mean number of exceedance events during a period of length t equal to $N(t) = \gamma t \delta$. The sub-model obtained by censoring is evidently a binomial distribution:

$$B_k(\delta', \gamma t) = \text{Prob}[k \text{ values } X_s \text{ of a total of } \gamma t; \text{ with } \delta' = \text{Prob}(X_s > x)]$$

with a mean number of events ($X_s > x$) during a period of length t equal to $N(t) = \gamma t \delta'$. However,

$$\delta' = \text{Prob}[X_s > x] = \text{Prob}[X > x] \, \text{Prob}[X_s > x \mid X > S] = \delta[1 - G_s(x)]$$

so

$$N(x, t) = \gamma t \delta[1 - \bar{G}_s(\dot{x})] \tag{19}$$

$$\text{Prob}[k \text{ values of } X_S > x \text{ in } t] = C_{\gamma t}^{k} A^{k} \cdot (1 - A)^{\gamma t - k} \tag{20}$$

with: $A = \delta[1 - G_s(x)]$

$$\begin{aligned} F(x) &= \text{Prob}[X^* < x \text{ in } [0;t]] \\ &= \text{Prob}[0 \text{ values of } X_s > x \text{ in } [0;t]] \\ &= [1 - \delta(1 - G_s(x))]^{\gamma t} \qquad \text{(see eq. 20)} \end{aligned} \tag{21}$$

which, after replacing γ and δ with their estimators $\hat{\gamma}$ and $\hat{\delta}$:

$$F_x(x) = [\hat{I}_t - (\hat{I}_t - 1) G_s(x)]^{\hat{u}t/(1 - \hat{I}_t)} \tag{22}$$

Combination of (15) and (19) yields the relationship between $T_s(x)$ of a quantile x and its non-exceedance probability $G_s(x)$:

$$G_s(x) = 1 - (1 - \exp[-1/(\gamma T_s(x)])/\delta \quad (23)$$

which gives:

$$G_s(x) = \{\hat{l}_t - \exp[(\hat{l}_t - 1)/(\hat{\mu} \cdot T_s(x))]\}/(\hat{l}_t - 1) \quad (24)$$

Combination of (21) and (23) yields the relationship between the return period $T_s(x)$ of a quantile and its non-exceedance probability $F_x(x)$:

$$F_x(x) = \exp[-t/T_s(x)] \quad (25)$$

3.2.3 Poisson distribution

The Poisson distribution is a limiting form of the binomial distribution (see Appendix 1). The preceding result shows that the distribution of the number of events ($X_s > x$) in the subclass is a Poisson distribution, since $\mu_s = \delta\gamma$. Therefore:

$$N(x, t) = \mu_s t[1 - G_s(x)] \quad (26)$$

$$\text{Prob[k values of } X_S > x \text{ in t]} = \exp[-N(x,t)]\,[N(x,t)]^k/k! \quad (27)$$

$$F_X(x) = \text{Prob[0 values } X_s > x \text{ in t]}$$

$$= \exp[-\mu_S t\,(1 - G_s(x))] \qquad \text{(see eq. 27)} \quad (28)$$

Since $T = 1/\mu'$ where $\mu = dN(x,t)/dt$, we have

$$G_s(x) = 1 - 1/(\mu_s T_s(x)) \quad (29)$$

The combination of (28) and (29) yields the same relationship (25) than in the case of the binomial distribution.

3.2.4 Negative binomial distribution

The negative binomial distribution can be characterized by a Poisson distribution with mean μ', where μ' is a random variable having a gamma distribution with scale parameter $\delta/(1-\delta)$ and shape parameter γt. For arbitrary t and fixed γ one may proceed as follows

1) randomly draw a value μ' from a gamma distribution:

$$g(\mu') = [\delta/(1-\delta)]^{\gamma t}\,[1/\Gamma(\gamma t)]\,\exp[-\delta\mu'/(1-\delta)]\,(\mu')^{\gamma t - 1}$$

2) generate a Poisson process with parameter μ' in [0;t]:

$$P_{\mu}'(k) = \exp[-\mu'].(\mu')^k/k!$$

The result is a process in which the number of occurrences has a negative binomial distribution.

$$w_k(t) = \text{Prob[k values } X_s \text{ in (0;t)]}$$

$$= \int_0^{+\infty} P'_{\mu}(k) \cdot g(\mu') \cdot d\mu'$$

$$= C^{k}_{\gamma t+k-1}\delta^{\gamma t}(1-\delta)^{k} \qquad \text{(see eq. 9)}$$

If we censor the sample of over-threshold events ($X_s > x$), the resulting process is still Poissonian with parameter $\mu'[1 - G_s(x)]$. Therefore:

$$N(x,t) = \gamma t[(1-\delta)/\delta)]\,[1 - G_s(x)] \tag{30}$$

$$\text{Prob[k values of } X_s > x \text{ in t]} = C^{k}_{\gamma t+k-1}(\delta/B)^{\gamma t}(1-\delta/B)^{k} \tag{31}$$

where $B = 1 - (1 - \delta)\, G_s(x)$, and

$$F_x(x) = \text{Prob [0 valeurs } X_s > x \text{ pendant t]}$$

$$= [\delta/(1-(1-\delta)\,G_s(x))\,]^{\gamma t} \qquad \text{(cf. eq. 31)} \tag{32}$$

which, upon replacing γ and δ by their estimators $\hat{\gamma}$ and $\hat{\delta}$, yields expression (22), the same as for the binomial distribution.

Combination of (11) and (30) gives the relationship between the return period $T_s(x)$ of the quantile x and its non-exceedance probability $G_s(x)$:

$$G_s(x) = (1 - \delta \exp[1/(\gamma T_s(x)])/(1-\delta) \tag{33}$$

By replacing γ and δ in the above expression with their estimators, we obtain the same result as for the binomial distribution (23). Combination of (32) and (33) also gives the same expression (25) as for the binomial distribution.

3.2.5 Non-stationary Poisson process

The non-stationary Poisson process has been studied in detail by NORTH (1980) and BERNIER (1981). In this case, one defines the non-exceedance probability of values above the threshold as a function of time:

$$G_s(x,t) = \text{Prob }[X_s < x \text{ at time t}].$$

We obtain:

$$F_x(x,t) = \exp\left[-\int_0^t [1 - G_s(x,\tau)]\mu_s(\tau)d\tau\right] \tag{34}$$

$$N(x,t) = \int_0^t \mu_s(\tau)[1 - G_s(x,\tau)]d\tau \tag{35}$$

BERNIER (1981) showed that, if $G_s(x,t)$ has the same form in time, for example always exponential but with time-dependent parameters, then one will obtain a Poisson process by estimating the parameters of $G_s(x,t)$ from all the events exceeding the threshold which is equivalent to averaging the parameters of $G_s(x,t)$ over the available period. More specifically, we obtain:

$$G_s(x) = 1 - 1/(\tilde{T}_s(x)) \tag{36}$$

$$F_x(x) = \exp[-\bar{t}/\tilde{T}_s(x)], \text{ with } \tilde{T}_s(x) = \int_0^{T_s(x)} \mu_s(\tau)d\tau \tag{37}$$

The case of Poisson processes stationary at the annual time scale but with seasonal variations has been thoroughly studied in the literature: BORGMAN (1963) gives an example of the use of this method to evaluate the risk of hurricanes at the East cost of the United States, exploiting the knowledge of the seasonal distribution of hurricanes. TODOROVIC and ZELENHASIC (1970), TODOROVIC and ROUSSELLE (1971), TODOROVIC and WOOLHISER (1972), GUPTA *et al.* (1976), TODOROVIC (1978), NORTH (1980), ASHKAR and ROUSSELLE (1981), BERNIER (1981), KONECNY and NACHTNEBEL (1985), and NACHTNEBEL et KONECNY (1987) have proposed different applications of the non-stationary Poisson process. A recent model, described by NACHTNEBEL and KONECNY (1994), is given by:

– a periodic intensity function:

$$\mu(t) = \theta_0 \exp \sum_{j=1}^{q} \theta_j \sin(2\pi j(t/365) + \theta_{q+j})$$

– an exponential distribution for over-threshold events, with a scale parameter described by a periodic function:

$$G_s(x,t) = 1 - \exp[-\beta(t)\,x] \text{ with } \beta(t) = \gamma_0 \cdot \exp\left(\sum_{k=1}^{r} \gamma_k \cdot \sin(2\pi k(t/365) + \gamma_{r+k})\right)$$

4 – RELATIONSHIP BETWEEN THE DIFFERENT TYPES OF SAMPLING

4.1 Correspondence between the return periods T_s and T_x

Strictly speaking, the return periods T_s and T_x can not be directly compared, because they correspond to two different variables: X* and X_s. Hence, the two return periods also have a different interpretation, as noted for example by NADEN and BAYLISS (1993): the return period T_x is the mean time separating two successive years with annual maximum above $X(T_x)$, while the return period T_s is the mean time separating two successive over-threshold values that exceed $X(T_s)$. The return period T_x is extensively used in hydrology because the predominance of annual maximum sampling, although its basis seems artificial. In the majority of cases, the user is interested in the estimation of "the risk of occurrence of all successive events", and hence in the quantile $X(T_s)$. The only case where the use of the return period T_x would be relevant is when the user is indifferent as to whether there is one or more exceedances per sampling period, as for example in a rural region with annual crops where farmers will have to postpone replanting till the following year if an inundation damages the fields. Hence, it appears reasonable to present the results in terms of the return period T_s whatever the sampling scheme being considered.

Combining (16) and (25) (which applies to the Poisson, the negative binomial, and the binomial distributions), we obtain the following relationship, which was also obtained by BORGMAN (1963) for the Poisson process:

$$T_x/t = 1/(1 - \exp[-t/T_s]) = (T_s/t) + 1/2 + (t/T_s)/12 + o\,(t/T_s)^2 \qquad (38)$$

Table 1 shows that T_x quickly converges to T_s + 1/2, and $X(T_x)$ is generally close to $X(T_s)$ for T_x greater than 10 years. Table 1 was also presented by

LANGBEIN (1949), but he obtained it using a somewhat simplistic technique that did not take into account the process of peaks above a threshold. TAKEUCHI (1984), using an approach similar to that of BORGMAN, clearly demonstrated this point. The constant of 1/2 appearing in (38) can be explained in the following way. The number of years, N, separating two successive years with occurrences of an event is related to the integer part of the return period θ as:

$$N = \mathrm{Int}(\theta) \text{ or } N = \mathrm{Int}(\theta) + 1$$

depending on the specific dates at which the events occur. The two alternatives being equally likely, one may obtain the average as:

$$T_x = (1/2)\ T_s + (1/2)\ (T_s + 1) = T_s + 1/2$$

Table 1 Correspondence between the returns periods T_s and T_x.

T_s/t	1	2	5	10	20	100
T_x/t	1.58	2.54	5.52	10.51	20.5	100.5

BERAN and NOZDRYN-PLOTNICKI (1977) have compared the return periods T_s and T_x from 40 gauging stations in the United Kingdom and found that (38) was not generally valid. Rather:

$$T_x > 1/(1 - \exp[-1/T_s])$$

This can be explained either by the dependence between peaks above the threshold or by the dependence between the inter-event times. The empirical data used by BERAN and NOZDRYN-PLOTNICKI suggest the second explanation, because of a biased estimate of the return period $T = E[\theta]$.

ROSBJERG (1987) studied the relationship between T_s and T_x when the peaks above the threshold has non-zero autocorrelation (ROSBJERG's equation 45). The general case lies in between the two extreme situations:
- strict independence ($\rho = 0$): equation (38)
- complete dependence ($\rho = 1$): $T_x/t = (\mu_s/(1 - \exp[-\mu_s])\ (T_s/t)$

When the events are independent ($\rho = 0$), one can only for rare events use the return period T without specifying its exact definition T_s or T_x (ROSBJERG, 1977a). For more frequent events, it is preferable to use the return period T_s (with the distribution F_x or G_s) and then (38) to determine the $T_s - T_x$ relationship.

When events are completely dependent ($\rho = 1$), the interpretation of the return periods in (38) presents a problem. Either all events of the series are linked in which case only one event ($X^* > x$) occurs, that is, there is no observation of the time interval between two successive events (T_x undefined), or all exceedances of the threshold are linked in which case over-threshold sampling is of no interest (*T_s undefined*).

In the general case where $0 < \rho < 1$, or when the process is non homogeneous (several seasons within the year), it is difficult to determine the exact relationship between T_s and T_x. It is preferable that the hydrologist be consistent in the choice of sampling scheme and return period: if one is interested in no more than one flood event per sampling period (return period T_X), then one should sample k maximums per period; if one is interested in all successive events, then one should use over-threshold sampling.

4.2 Relationship between F_x, G_{kx}, and G_x

The estimation of a quantile X(T) can be accomplished using two types of sampling schemes: selection of the k largest values per sampling period, or selection of the peak flows of all exceedance events. In general, we recommend the second type of sampling which allows a more realistic representation of the phenomenon considered.

• For rare events (T > 10), the two sampling schemes yield essentially the same quantile estimates, since $T_X/t \approx (T_S/t) + 1/2$. Certain differences may be observed if the time series has years without extreme events, or, alternatively, contains clusters of extreme events. In these cases, sampling of G_{kx} may lead to observations that are not representative (small values in dry years, and omission of important values in wet years).

• For events with intermediate recurrence interval (1 < T < 10), it is necessary to use the correspondence between the two distributions. We summarize in Table 2 these relationships for the Poisson, the negative binomial, and the binomial processes.

Table 2 Correspondence between the distributions G_{kx} and G_S.

Sampling on intervals of length t	Distribution G_{kx}	Distribution G_s	
Return period T_x	$(1 - t/T_x)^{1/k}$	$1 - 1/(\hat{\mu} \cdot T_{eq\,2})$	(a)
		$\{\hat{I}_t - \exp[(\hat{I}_t - 1)/(\hat{\mu} \cdot T_{eq\,2})]\}/(\hat{I}_t - 1)$	(b)
Return period T_s	$\exp(-t/k \cdot T_s)$	$1 - 1/(\hat{\mu} \cdot T_s)$	(a)
	ou : $1 - 1/(\hat{\mu} \cdot T_{eq\,1})$	$\{\hat{I}_t - \exp[(\hat{I}_t - 1)/(\hat{\mu} \cdot T_s)]\}/(\hat{I}_t - 1)$	(b)
(a) Poisson Process	(b) Binomial or negative binomial distribution		
$T_{eq\,1} = t/1 - \exp(-t/T_s))$	$T_{eq\,2} = -t/\mathrm{Log}(1 - t/T_x)$	$\hat{\mu} = \hat{E}(m_t)/t$	$\hat{I}_t = \hat{V}ar(m_t)/\hat{E}(m_t)$

• For frequent events (T < 1), there is little gain in using the distribution G_S. The renewal method provides a direct and good estimate of the quantile, without the necessity to introduce the distribution of over-threshold events. Since $G_S(x)$= Prob[X_S < x], where X_s is the peak of an event exceeding the threshold S, we have $G_S(S) = 0$. One may obtain an estimate of the quantile $X(T_S)$ as $\hat{X}(T_S) = S$, where S is determined by successive approximation to give $\hat{\mu} = 1/T_S$ (Poisson process), or $(\hat{I}_t - 1)/(\hat{\mu} \cdot \log \hat{I}_t) = T_s$ (negative binomial or binomial distribution).

4.3 Comparison of the sampling distribution of quantiles from G_{kx} and G_s

CUNNANE (1973), TAVARES and DA SILVA (1983), ROSBJERG (1985), WANG (1991), and MADSEN *et al.* (1997) have compared the sampling variance of quantile estimators based on maximum annual floods (F_x) and on peaks-over-threshold (G_s). They found a slight advantage of the distribution F_x over G_s when $\mu_s = 1$; an advantage of G_s over F_x when $\mu_s > 1.7$, and a comparable performance when the return period is large (T > 10 years). Buishand (1989) compared analytically the mean square error of flood quantile estimates from the Poisson-Exponential model for threshold exceedances with two types of sampling: selection of all values above a threshold and selection of the n largest events in the series. The theoretical values of the mean square errors being of the same order of magnitude, Buishand concluded that the two approaches are essentially equivalent. It is preferable, however, to use the first approach which implicitly assures that the events are separated by observations below the threshold, whereas the second requires considerations as to the time that should separate peaks.

LANG (1995b) extended these result by an analysis of quantiles from G_{kx} and G_s with $k = \mu_s$ varying from 1 to 5. The comparison criterion was the root mean square error of quantiles obtained from the Poisson-Exponential model and the Gumbel distribution corresponding to k maximums per year. Sampling of G_s gives on the average a root mean square error that is two to three times smaller than that corresponding to sampling of G_{kx}.

From the above mentioned simulation studies, one must conclude that sampling of G_s rather than the classical F_x is advantageous. If there appears to be an advantage of the distribution F_x over G_s (with $\mu_s = 1$), one may lower the threshold to improve the results based on the distribution G_s, whereas increasing the number of values to estimate G_{kx} generally does not lead to an improvement. In fact it generally leads to a deterioration of the results.

5 – CONCLUSION

Numerous papers have dealt with the sampling of over-threshold values. These studies have shown that a variety of phenomena can be modeled by renewal processes. The contribution of this study is a detailed study of the case where the process can be described by the negative binomial and binomial distributions, and a comparison of the sampling of G_{kx} and G_s which in previous studies has been restricted to the case of one event per year (k = 1 or $\mu_s = 1$). We emphasize the somewhat artificial character of the return period T_x in the classical sampling of F_x: T_x is the mean time separating two successive years with maximums above $X(T_x)$ whereas the return period T_s related to the sampling of G_s, that is, the mean time separating two successive over-threshold events exceeding $X(T_s)$, better corresponds the public understanding of the notion of risk.

An additional aspect which has not been discussed here concerns the joint use of over-threshold sampling and Bayesian analysis: WOOD and RODRIGUEZ-IRTURBE (1975a, b) have studied the uncertainty associated with model choice, and ROUSSELLE and HINDIE (1976) have studied the uncertainty of estimated parameters. MIQUEL (1984) has examined the inclusion of information on historical floods, RASMUSSEN and ROSBJERG (1991), MADSEN *et al.* (1994), and MADSEN and ROSBJERG (1995b) the inclusion of regional information. Finally, BERNIER (1981) and ABI-ZEID (1995) have studied the estimation of the intensity of a non-stationary Poisson process using a Bayesian approach.

As a final remark, let us mention that one of the difficulties associated with over-threshold sampling concerns the introduction of extra parameters that must be determined. LANG *et al.* (1997) present a review of tests and techniques available to guide the choice of these parameters.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the constructive comments of professor Jacques BERNIER, in particular his suggestion to derive the properties in section 3.2 by a random censoring. The previous approach is described in Appendix 2.

NOTATION

$B_k(\delta, \gamma t)$:Binomial distribution with parameters δ and γt.

E :Event.

E(X) :Mathematical expectation of the random variable X; $E(X) = \int_{-\infty}^{+\infty} x f(x)dx$.

EQM : Mean square error of the estimator $\hat{X}$ of X; $EQM^2 = b^2 + d^2$, where $b = E(\hat{X}) - X$ is the bias of the estimator, and $d = (\mathrm{Var}\hat{X})^{1/2}$ is the standard deviation of the estimator.

f(x) : Probability density function.

F(x) : Cumulative density function.

F_x : Cumulative density function of the variable X*.

G_{kx} : Cumulative density function of the variable X_k.

G_s : Cumulative density function of the variable X_s.

Int(x) : Integer part of x.

I_t : Dispersion index; $I_t = \mathrm{Var}(m_t)/E(m_t)$.

m_t : Number of events in the interval [0;t]; with distribution function $w_k(t)$ and expectation N(t).

N(t) : Mean number of events in the interval [0;t]; $N(t) = E(m_t)$.

N(t,t') : Mean number of events in the interval [t;t']; $N(t,t') = \int_t^{t'} \mu(\tau)d\tau$

$N(x,t)$	: Mean number of events exceeding x in the interval [0;t].
T	: Return period $T = E(\theta)$; expected waiting time θ.
T_s	: Return period of the variable X_s.
T_x	: Return period of the variable X^*.
$Var(X)$	: Variance of the random variable X; $Var(X) = \int_{-\infty}^{+\infty} (x - E(x))^2 f(x) dx$.
$w_k[0;t]$	: Distribution function of m_t; denoted $w_k(t)$ when the process is stationary.
X	: Random variable.
$\hat{X}$	: Estimator of the variable X.
$\tilde{X}$	: Change of scale of the variable X; $\tilde{X} = \int_0^X \mu(\tau) \cdot d\tau$
X^*	: Maximum of X in a sampling period of fixed length.
X_k	: One of the k largest values of X in a sampling period of fixed length.
X_s	: Maximum value of X during an over-threshold event where S is the threshold.
$X(T)$	: Quantile of the random variable X, with associated return period T.
α	: Scale parameter of the cumulative distribution function of the waiting time θ.
δ,γ	: Parameters of the Binomial and the negative Binomial distribution.
$\Gamma(x)$	: Gamma function; $\Gamma(n) = (n-1)!$ if n is an integer; $\Gamma(z) = \int_0^{+\infty} x^{z-1} e^{-x} dx$
$\mu(t)$	: Parameter describing the intensity of the process; $\mu(t) = dN(t)/dt$, denoted μ if the process is stationary.
$\mu_S(t)$	: Parameter $\mu(t)$ relation to the process of threshold exceedances X_s; denoted μ_s if the process is stationary.
θ	: Waiting time, *i.e.* the time between to successive occurrences of an event; denoted $\theta(t)$ if the process is non-stationary.
ρ	: Autocorrelation coefficient of the series of maximum.

APPENDIX 1 – PROPERTIES OF THE WAITING TIME

Poisson process

The four hypotheses defining the Poisson process are:

(i) $w_k[0;t]$ is independent of time t.

(ii) $w_1[0;dt] = \mu\, dt$

(iii) $\sum_{k=2}^{\infty} W_k[0;dt]$ negligible compared to dt

(iv) $w_k[t_1; t_2]$ and $w_k[t_3;t_4]$ are independent if the intervals are disjunct.

One can establish a recurrence relationship between $w_k[0;t]$ and $w_{k-1}[0;t]$:

$$w_k[0;t+dt] = \sum_{j=0}^{k} w_{k-j}[0;t]w_j[t;t+dt] \qquad \text{in virtue of (iv)}$$

$$\approx w_k[0;t]\,w_0[t;t+dt] + w_{k-1}[0;t]\,w_1[t;t+dt] \qquad \text{in virtue of (iii)}$$

We obtain $w_1[t;t+dt] = \mu dt$, in virtue of (ii) and (i), and:

$$\begin{aligned} w_0[t;t+dt] &= 1 - w_k[t;t+dt] \\ &\approx 1 - w_1[t;t+dt] \qquad \text{in virtue of (iii)} \\ &= 1 - \mu dt \end{aligned}$$

Therefore:

$$(w_k[0;t+dt] - w_k[0;t])/(\mu\, dt) = w_{k-1}[0;t] - w_k[0;t]$$

or:

$$(1/\mu)\ \ w'_k[0;t] = w_{k-1}[0;t] - w_k[0;t]\ (k > 1) \tag{A.1}$$

For k = 0, we have: $w_0[0;t+dt] = w_0[0;t]\,w_0[t;t+dt]$

or:

$w'_0[0;t] = -\mu w_0[0;t]$

Since $w_0[0;0] = 1$, we have $w_0[0;t] = \exp[-\mu t]$. If we let $w_k[0;t] = u_k[0;t]\exp[-\mu t]$, (A.1) becomes:

$u'_k[0;t] = \mu u_{k-1}[0;t]$

It is readily seen that:

$$w_k[0;t] = \exp[-\mu t]\,(\mu t)^k/k! \tag{A.2}$$

We are in particular interested in the time separating two successive occurrences of an event. Let E_k be an event occurring at time $t = t_k$, E_{k+1} an event occurring at time $t = t_{k+1}$, and $\theta_k = t_{k+1} - t_k$ the time separating the two successive occurrences.

$\text{Prob}[\theta_k > d] = \text{Prob}[0 \text{ events in } [t_k;t_{k+d}]]$

$= w_0[t_k;t_{k+d}]$

$= w_0[0;d]$ in virtue of (i)

or:

$$F(d) = \text{Prob}[\theta < d] = 1 - w_0[0;d] = 1 - \exp[-\mu d] \tag{A.3}$$

$$f(\theta) = \exp[-\mu\theta] \tag{A.4}$$

Negative Binomial distribution

We have

$$w_k\,[0;t] = C^k_{\gamma t+k-1}\delta^{\gamma t}\cdot(1-\delta)^k \tag{A.5}$$

with: $E\,[m_t] = \gamma t\,(1-\delta)/\delta$ etVar $[m_t] = \gamma t\,(1-\delta)/\delta^2$

When the dispersion index tends to 1, that is, when $\delta \to 1^-$ and $\gamma \to +\infty$, the negative Binomial distribution tends to the Poisson distribution. In fact, if we denote $\hat{\mu} = E\,(m_t)/t$, δ is estimated as $\hat{\delta} = 1/(1+\hat{\mu}/\gamma)$, and we have:

$$w_k[0;t] = (\gamma t+k-1)\,(\gamma t+k-2)\ldots(\gamma t)\delta^{\gamma t}\,(1-\delta)^k/k! \qquad \text{in virtue of (A.5)}$$

$$\lim_{\gamma\to+\infty} w_k[0;t] \approx (\gamma t)^k(1+\hat{\mu}/\gamma t)^{-\gamma t}(\hat{\mu}/\gamma t)^k(1+\hat{\mu}/\gamma t)^{-k}/k!$$

$$\approx (\hat{\mu})^k\,(1+\hat{\mu}/\gamma t)^{-\gamma t}/k!$$

$$\approx \exp\,[-\hat{\mu}]\;(\hat{\mu})^k/k!$$

which is identical to the expression (A.2) for the Poisson distribution.

In the same way as for the Poisson process, one can calculate the probability density function for the waiting time θ.

$$F(d) = 1 - w_0\,[0;d] = 1 - \delta^{\,\gamma d} \tag{A.6}$$

$$f(\theta) = -\gamma \log\delta\;\delta^{\,\gamma\theta} \tag{A.7}$$

Binomial distribution

$$w_k[0;t] = C^k_{\gamma t}\delta^k(1-\delta)^{\gamma t-k} \tag{A.8}$$

with: $E\,[m_t] = \gamma t\,\delta$ and $\mathrm{Var}\,[m_t] = \gamma t\,\delta\,(1-\delta)$

When the dispersion index I_t tends to 1, that is, when $\delta \to 0^+$ and $\gamma \to +\infty$, the Binomial distribution tends to the Poisson distribution. In fact, if we define $\hat{\mu} = E\,(m_t)/t$, δ is estimated as $\hat{\delta} = \hat{\mu}/\gamma$, and we have:

$$w_k\,[0;t] = (\gamma t)\,(\gamma t-1)\ldots(\gamma t-k+1)\,\delta^k\,(1-\delta)^{\gamma t-k}/k! \qquad \text{in virtue of (A.8)}$$

$$\lim_{\gamma\to+\infty} w_k[0;t] \approx (\gamma t)^k(\hat{\mu}/\gamma t)^k(1-\hat{\mu}/\gamma t)^{\gamma t-k}/k!$$

$$\approx (\hat{\mu})^k(1-\hat{\mu}/\gamma t)^{\gamma t}/k!$$

$$\approx \exp\,[-\hat{\mu}]\,(\hat{\mu})^k/k!$$

which is identical to the expression (A.2) for the Poisson distribution.

In the same way as for the Poisson process, one can calculate the probability density function for the waiting time θ:

$$F(d) = 1 - w_0\,[0;d] = 1 - (1-\delta)^{\gamma d} \tag{A.9}$$

$$f(\theta) = -\gamma\log\,(1-\delta)\,(1-\delta)^{\gamma\theta} \tag{A.10}$$

APPENDIX 2 – RELATIONSHIP BETWEEN THE RETURN PERIOD O A QUANTILE X AND THE EXCEEDANCE PROBABILITY OF X

Poisson process

Combination of (1) and (18) yields the relationship (28), first described by SHANE and LYNN (1964). The relationship between the return period T_s (x) of a quantile x and its non-exceedance probability G_s (x) can be obtained as follows:

Prob [k values X_s > x during time t]

$= \sum_{i=0}^{\infty}$ Prob [(k + i) values X_s with k values greater than x and i values smaller than x during time t]

$$= \sum_{i=0}^{\infty} C^i_{k+i} w_{k+i}(t)[(1 - G_s(x))]^k [G_s(x)]^i$$

$$= \sum_{i=0}^{\infty} \frac{(k+i)!}{k!i!} \frac{\exp[-\mu_s t](\mu_s \cdot t)^{(k+i)}}{(k+i)!}[1 - G_s(x)]^k [G_s(x)]^i$$

$= \exp[-N(x, t)]\,[N(x, t)]^k/k!$ (equation 27)

Negative Binomial distribution

Combination of (9) and (18) yields the relationship (32), described by VUKMIROVIC and PETROVIC (1995). The relationship between the return period T_s (x) of a quantile x and its non-exceedance probability G_s (x) can be obtained as follows:

Prob [k values X_s > x during time t]

$$= \sum_{i=0}^{\infty} C^i_{k+i} C^{k+i}_{k+i+\gamma t-1} \delta^{\gamma t}(1-\delta)^{k+i}[1 - G_s(x)]^k [G_s(x)]^i$$

$$= C^k_{k+\gamma t-1} \delta^{\gamma t}(B-\delta)^k B^{-(k+\gamma t)} \sum_{i=0}^{\infty} C^i_{i+k+\gamma t-1}(1-B)^i B^{k+\gamma t}$$

(with $B = 1 - (1 - \delta)\, G_s(x)$)

$= C^k_{\gamma t+k-1}(\delta/B)^{\gamma t}(1 - \delta/B)^k$ (equation 31).

Binomial distribution

Combination of (13) and (18) yields the relationship (21) described by VUKMIROVIC and PETROVIC (1995). The relationship between the return period $T_s(x)$ of a quantile x and its non-exceedance probability $G_s(x)$ can be obtained as follows:

Prob [k values $X_s > x$ during time t]

$$= \sum_{i=0}^{\infty} C^i_{k+i} w_{k+i}(t)[1 - G(x)]^k [G_s(x)]^i$$

$$= \sum_{i=0}^{\gamma t - k} C^i_{k+i} C^{k+i}_{\gamma t} \delta^{k+i} (1-\delta)^{\gamma t - k - i} [1 - G_s(x)]^k [G_s(x)]^i$$

$$= C^k_{\gamma t} A^k (1-\delta)^{\gamma t - k} \sum_{i=0}^{\gamma t - k} C^i_{\gamma t - k} [(\delta - A)/(1-\delta)]^i \qquad \text{(with } A = \delta\,[1 - G_s(x)])$$

$$= C^k_{\gamma t} A^k (1-\delta)^{\gamma t - k} [1 + (\delta - A)/(1-\delta)]^{\gamma t - k}$$

$$= C^k_{\gamma t} A^k (1 - A)^{\gamma t - k} \qquad \text{(equation (20)).}$$

REFERENCES

ABI-ZEID, I. (1995). L'estimation de l'intensité d'un processus de Poisson inhomogène par une méthode bayésienne. Statistical and Bayesian Methods in Hydrological Sciences, Intern. Conf. in Honor of Jacques Bernier, Paris, UNESCO, 11-13 sept., 13 p.

ASHKAR, F., ROUSSELLE, J. (1981). Design discharge as a random variable: a risk study. *Water Resour. Res., 17 (3)*, 577-591.

ASHKAR, F., BOBEE, B., RASMUSSEN, P. F, ROSBJERG, D. (1994). A perspective on the annual maximum flood approach to flood frequency analysis. *Stochastic and Statistical Methods in Hydrology and Environmental Engineering, Vol.1, Extreme Values: Floods and Droughts*, Ed. by K. W Hipel, Water Science and Technology Library, Kluwer Academic Publishers, 3-14.

BASS, J. (1974). *Éléments de calcul des probabilités*. Ed. Masson et Cie, Paris, 3e ed., 275p.

BEN-ZVI, A. (1991). Observed advantage for negative binomial over Poisson distribution in partial duration series. *Stochastic Hydrol. Hydraul., 5*, 135-146.

BERAN, M.A., NOZDRYN-PLOTNICKI, M.J. (1977). Estimation of low return period floods. *Hydrol. Sci. Bull., XXII, 2, 6*, 275-282.

BERNIER, J. (1967). Sur la théorie du renouvellement et son application en hydrologie. *Rapport EDF*, Chatou, HYD 67, n° 10, 22 p.

BERNIER, J. (1981). Le modèle de renouvellement non stationnaire. *Rapport EDF*, Chatou, HE 40 81-11, 36 p.

BOIRET, P. (1987). Analyse des précipitations de 6 mn à 24 h par une méthode du type renouvellement. Note de travail du service central d'exploitation de la météorologie, n° 19, MELATT, Météo Nationale, Paris, 34 p.

BORGMAN, L.E. (1963). Risk criteria. *J. Waterw. Harbors Div., ASCE, 80*, 1-35.

BUISHAND, T.A. (1989). The partial duration series method with a fixed number of peaks. *J. Hydrol., 109*, 1-9.

CERVANTES, J.E., KAVVAS, M.L., DELLEUR, J.W. (1983). A cluster model for flood analysis. *Water Resour. Res., 19 (1)*, 209-224.

COX, D.R. (1965). *Renewal Theory*. Methuen & Co, LTD, London, 186 p.

CUNNANE, C. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *J. Hydrol., 18*, 257-271.

CUNNANE, C. (1979). A note on the Poisson assumption in partial duration series models. *Water Resour. Res., 15 (2)*, 489-494.

DAVISON, A. C, SMITH, R.L. (1990). Models for Exceedances over High Thresholds. *J. Roy. Statist. Soc., B, 52, n° 3*, 393-442.

FELLER, W. (1966). *An Introduction to Probability Theory and its Applications.* John Wiley & Sons, Inc., New York, vol. II.

GUPTA, V.K., DUCKSTEIN, L., PEEBLES, R.W. (1976). On the joint distribution of the largest flood and its time of occurrence. *Water Resour. Res., 12 (2)*, 295-304.

KARR, A.F. (1976). Two extreme value processes arising in hydrology. *J. Appl. Prob., 13*, 190-194.

KAVVAS, M.L. (1982a). Stochastic trigger model for flood peaks. 1: Development of model. *Water Resour. Res., 18 (2)*, 383-398.

KAVVAS, M.L. (1982b). Stochastic trigger model for flood peaks. 2: Application of the model to the flood peaks of Goksu-Karahacili. *Water Resour. Res., 18 (2)*, 399-411.

KONECNY, F., NACHTNEBEL, H.P. (1985). Extreme value processes and the evaluation of risk in flood analysis. *Appl. Math. Modelling, 9*, 11-15.

KUNDZEWICZ, Z.W. (1989). Renewal theory criteria of evaluation of water-resource systems: reliability and resilience. *Nordic Hydrol., 20*, 215-320.

LANG, M. (1995a). Les chroniques en hydrologie : Modélisation comparée par un système de gestion de bases de données relationnel et orienté-objet ; Traitements de base et intervalles de confiance des quantiles de crues ; Techniques d'échantillonnage par la méthode du renouvellement – Thèse de doctorat, Université Joseph Fourier Grenoble, Cemagref Lyon, 296 p.

LANG, M. (1995b). Techniques d'échantillonnage par la méthode du renouvellement – Séminaire inter-chercheurs, Les modèles au Cemagref, tome 1, Gif-sur-Yvette, ed. Cemagref Antony, 89-98.

LANG, M., OUARDA, T.B.M.J., BOBEE, B. (1997). Towards operational guidelines for over-threshold modeling. *J. Hydrol., submitted.*

LANGBEIN, W.B. (1949). Annual floods and the partial duration flood series. *Transactions, AGU, 30 (6)*, 879-881.

LEADBETTER, M.R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Ver. Geb., 65*, 291-306.

LEADBETTER, M.R., ROOTZEN, H. (1988). Extremal theory for stochastic processes. *Ann. Probab., 16*, 431-478.

LINDGREN, G., ROOTZEN, H. (1987). Extreme values: theory and technical applications. *Scand. J. Statist., 14*, 241-279.

MADSEN, H., RASMUSSEN, P.F., ROSBJERG, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events, 1, At-site modeling. *Water Resour. Res., 33 (4)*, 747-757.

MADSEN, H., ROSBJERG, D. (1995a). On the modelling of extreme droughts. Modelling and management of sustainable basin-scale water resource systems. *IAHS Publ. n° 231*, 377-385.

MADSEN, H., ROSBJERG, D. (1995b). A regional Bayesian method for estimation of extreme streamflow droughts. Statistical and Bayesian Methods in Hydrological Sciences, Intern. Conf. in Honor of Jacques Bernier, Paris, UNESCO, 11-13 sept., 23 p.

MADSEN, H., ROSBJERG, D., HARREMOËS, P. (1994). PDS-modelling and regional Bayesian estimation of extreme rainfalls. *Nordic Hydrol., 25*, 279-300.

MIQUEL, J. (1984). *Guide pratique d'estimation des probabilités de crue.* Ed. Eyrolles, Paris, 160 p.

NACHTNEBEL, H.P., KONECNY, F. (1987). Risk analysis and time-dependent flood models. *J. Hydrol., 91*, 295-318.

NACHTNEBEL, H.P., KONECNY, F. (1994). Risk estimation in partial duration series of seasonal floods. *Stochastic and Statistical Methods in Hydrology and Environmental Engineering, Vol.1, Extreme Values: Floods and Droughts*, Ed. by K. W Hipel, Water Science and Technology Library, Kluwer Academic Publishers, 73-83.

NADEN, P. S, BAYLISS, A.C. (1993). Flood estimation: peak-over-threshold techniques. MAFF Conference of River and Coastal Engineers, Univ. of Loughborough, 5-7 July, 9.1.1-9.1.18.

NORTH, M. (1980). Time-dependent stochastic model of floods. *J. Hydraul. Div., ASCE, 106 (HY5)*, 649-665.

PICKANDS, J. (1975). Statistical inference using extreme order statistics. *Ann. Stat., 3 (1)*, 119-131.

RASMUSSEN, P.F. (1991). The partial duration series approach to flood frequency analysis. Ph.D. thesis, Inst. of Hydrodynamics and Hydraulic Engineering, Tech. Univ. Denmark, Series Paper 55, 138 p.

RASMUSSEN, P.F., ASHKAR, F., ROSBJERG, D., BOBEE, B. (1994). The POT method for flood estimation: a review. *Stochastic and Statistical Methods in Hydrology and Environmental Engineering, vol. 1, Extreme Values: Floods and Droughts.* Ed. by K. W Hipel, Water Science and Technology Library, Kluwer Academic Publishers, 15-26.

RASMUSSEN, P.F., ROSBJERG, D. (1991). Application of Bayesian principles in regional flood frequency estimation. *Adv. Water Resour. Tech.*, Ed. by G. Tsakaris, Balkema, Rotterdam, 65-75.

ROSBJERG, D. (1977a). Return periods of hydrological events. *Nordic Hydrol., 8*, 57-61.

ROSBJERG, D. (1977b). Crossings and extremes in dependent annual series. *Nordic Hydrol., 8*, 257-266.

ROSBJERG, D. (1985). Estimation in partial duration series with independent and dependent peak values. *J. Hydrol., 76*, 183-195.

ROSBJERG, D. (1987). On the annual maximum distribution in dependent partial duration series. *Stochastic Hydrol. Hydraul., 1*, 3-16.

ROSBJERG, D. (1993). Partial duration series in water resources. Inst. of Hydrodynamics and Hydraulic Engineering, Tech. Univ. Denmark, 33 p. + 13 notes.

ROSBJERG, D., MADSEN, H. (1992). On the choice of threshold level in partial durations series. XVII Nordic Hydrological Conference, Alta, Norway, *NHP report n° 30*, 604-615.

ROSBJERG, D., RASMUSSEN, P.F., MADSEN, H. (1991). Modelling of exceedances in partial duration series. *Proc. Int. Hydrol. Water Resour. Symp., Perth, 2-4 oct.*, 755-760.

ROSBJERG, D., MADSEN, H., RASMUSSEN, P.F. (1992). Prediction in partial duration series with generalized Pareto distributed exceedances. *Water Resour. Res., 28 (11)*, 3001-3010.

ROUSSELLE, J., HINDIE, F. (1976). Incertitude dans les débits de crues: approche bayésienne. *J. Hydrol., 30*, 341-349.

SHANE, R. M, LYNN, W.R. (1964). Mathematical model for flood risk evaluation. *J. Hydraul. Div., ASCE, 90 (HY6)*, 1-20.

TAKEUCHI, K. (1984). Annual maximum series and partial duration series: evaluation of Langsbein's formula and Chow's discussion. *J. Hydrol., 68*, 275-284.

TAVARES, L.V., DA SILVA, J.E. (1983). Partial duration series method revisited. *J. Hydrol., 64*, 1-14.

TODOROVIC, P. (1978). Stochastic models of floods. *Water. Resour. Res., 14 (2)*, 345-356.

TODOROVIC, P., ROUSSELLE, J. (1971). Some problems of flood analysis. *Water Resour. Res., 7 (5)*, 1144-1150.

TODOROVIC, P., WOOLHISER, D.A. (1972). On the time when the extreme flood occurs. *Water Resour. Res., 8 (6)*, 1433-1438.

TODOROVIC, P., ZELENHASIC, E. (1970). A stochastic model for flood analysis. *Water Resour. Res., 6 (6)*, 1641-1648.

VENTSEL, H. (1973). *Théorie des probabilités.* Ed. Mir, Moscou, 563 p.

VUKMIROVIC, V. (1990). Analiza verovatnoce pojave hidroloskih velicina (Analysis of probability of occurrence of hydrological variables; in Serbian). Faculty of Civil Eng. and "Naucna knjiga", Belgrade, 173 p.

VUKMIROVIC, V., PETROVIC, J. (1995). Flood flow analysis using renewal processes. FRIEND-AMHY Annual Conference, Thessaloniki; 25-27 sept., 12 p.

WANG, Q.J. (1991). The POT model described by the generalized Pareto distribution with Poisson arrival rate. *J. Hydrol., 129*, 263-280.

WOOD, E.F., RODRIGUEZ-ITURBE, I. (1975a). Bayesian inference and decision making for extreme hydrological events. *Water Resour. Res., 11 (4)*, 533-542.

WOOD, E.F., RODRIGUEZ-ITURBE, I. (1975b). A Bayesian approach to analyzing uncertainty among flood frequency models. *Water Resour. Res., 11 (4)*, 839-843.