

Article

« Revue de processus ponctuels et synthèse de tests statistiques pour le choix d'un type de processus »

A. Yagouti, I. Abi-Zeid, T. B.M.J. Ouarda et B. Bobée

Revue des sciences de l'eau / Journal of Water Science, vol. 14, n° 3, 2001, p. 323-361.

Pour citer cet article, utiliser l'information suivante :

URI: <http://id.erudit.org/iderudit/705423ar>

DOI: 10.7202/705423ar

Note : les règles d'écriture des références bibliographiques peuvent varier selon les différents domaines du savoir.

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter à l'URI <https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. Érudit offre des services d'édition numérique de documents scientifiques depuis 1998.

Pour communiquer avec les responsables d'Érudit : info@erudit.org

Revue de processus ponctuels et synthèse de tests statistiques pour le choix d'un type de processus

Review and classification of statistical tests applied to point processes

A. YAGOUTI^{1*}, I. ABI-ZEID², Taha B.M.J. OUARDA¹, B. BOBÉE¹

Reçu le 1^{er} avril 2000, accepté le 27 août 2001**.

SUMMARY

The design and management of hydraulic structures require a good knowledge of the characteristics of extreme hydrologic events such as floods and droughts, that may occur at the site of interest. Occurrences of such events may be modelled as temporal point processes. This modelling approach allows the derivation of various performance indices related to the design and operation of this infrastructure, as well as to the quantification and management of the associated risks. In this paper, we present statistical tests that may be applied for the modelling of a series of events by temporal point processes. A point process is defined as a stochastic process for which each realisation constitutes a series of points. Although a large body of literature dealt with temporal point processes, very few focused on the analysis of a series of events.

In the present paper we identify two types of series of events: the first represents a series of only one type of event, and the second represents a series of several types of events. The main objective of this research is to comprehensively review the statistical tests applied to the series of one or several types of events and to propose a classification of these tests. This comprehensive review of statistical tests applied to point processes is carried out with the ultimate objective of applying these tests to real case studies within the framework of risk analysis. For example, an extended low-flow event constitutes a risk that may place a water resources system in a state of failure. Thus, it's important to identify and quantify this risk in order to ensure the optimal management of water resources. The modelling of the observed series of

1. Chaire en hydrologie statistique, Institut national de la recherche scientifique, INRS-Eau, 2800, Rue Einstein, CP 7500, Sainte-Foy, Québec, G1V 4C7, Canada.

2. Centre de recherches pour la défense Valcartier, CRDV, BAT 24. 2459, Boulevard Pie XI Nord, Val Bélair, Québec, G3J 1X5, Canada.

* Correspondance. E-mail : yagoutab@inrs-eau.uquebec.ca

** Les commentaires seront reçus jusqu'au 14 août 2002.

events by point processes can provide some statistical results, such as the distribution of number of events or the shape of the intensity function. These results are useful in a risk analysis framework, which includes two steps: risk evaluation and risk management.

In the first part of the paper, a review and classification of the various temporal point processes are presented. These include the homogeneous and nonhomogeneous Poisson processes, the Negative Binomial process, the cluster point processes (such as the Neyman-Scott and the Bartlett-Lewis processes), the doubly stochastic Poisson processes, the self-exciting point processes, the homogeneous and nonhomogeneous renewal processes and the semi-markov processes. Also, we illustrate the various links and relationships that exist between these point processes. This classification is elaborated by considering the homogeneous Poisson process as the starting point. The simplicity and the wide use of this process in the statistical and hydrological literature justify this choice.

In the second part of the paper, statistical tests of a series of one type of event are identified. A series of events may be characterised by the number of events, the occurrence times of the events or by the duration of each event. These characteristics are considered as random variables that must be represented by suitable statistical distributions. A series of events may also be characterised by the intensity function, which represents the instantaneous average rate of occurrence of an event. Clearly, the choice of the statistical distribution to model the number of events in a series or the intensity function depends on the nature of the observed data. For example, a stationary series of events may be represented by a constant intensity function. Thus, it is necessary to conduct an analysis of the observed series of the events, such as graphical analysis and statistical testing in order to select and validate the hypothesis underlying the point process model. The hypotheses that may be verified include trend analysis, homogeneity analysis, periodicity analysis, independence of intervals between events, and the adequacy of a given distribution for the number of events and for the time intervals separating events.

In the third part, the applicability of the tests identified in the second part to the case of a series of two or more types of events is examined. In this part, our goal is to analyse the global point process (or the pooled output) obtained by the superposition of the p subsidiary point processes. The decomposition of the global process into p point processes necessitates an identification of each type of event, characterised generally by the number of occurrences and by the intervals between the successive events of the same type. We also examine the applicability of the statistical tests identified in the second part to the case where the global point process is characterised by the duration of each type of event. We investigate more specifically the case of two subsidiary point processes ($p=2$) where the two event types alternate in the time (an alternating point process). Finally, statistical tests identified in the second part are classified into four categories: tests based on graphical analysis; tests applied to the homogeneous and nonhomogeneous Poisson processes; tests applied to the homogeneous renewal process; and finally tests of discrimination between two specific processes. These tests of discrimination include the selection among the Poisson process and the renewal process, the Poisson process and the Binomial point process, and finally, the selection among these three point processes: Cox process, Neyman-Scott process and renewal process.

The results of this research indicate that, in the past, mostly tests for a series of one type of event were presented in the literature. These tests are only valid for the following point processes: a homogenous Poisson process or a homogenous renewal process. The application of these tests to a series of two or several types of events is possible as long as these events are only described by their number and time of occurrence i.e. the duration of each event can

not be taken into consideration. Otherwise, these tests are applicable to the alternating point process, which is characterised only by the number and the duration of the two types of events.

Key-words: *point process, alternating process, event, statistical test, classification, risk.*

RÉSUMÉ

Nous nous intéressons dans ce travail de recherche à la modélisation d'une série d'événements par la théorie des processus ponctuels temporels. Un processus ponctuel est défini comme étant un processus stochastique pour lequel chaque réalisation constitue une collection de points. Un grand nombre d'ouvrages traitent particulièrement de ces processus, cependant, il existe dans la littérature peu de travaux qui se préoccupent de l'analyse de séries d'événements. On identifie deux catégories de séries d'événements : une série d'un seul type d'événements et une série de plusieurs types d'événements.

L'objectif de ce travail est de mettre en évidence les différents tests statistiques appliqués aux séries d'un seul ou de plusieurs types d'événements et de proposer une classification de ces tests. Nous présentons d'abord une revue de littérature des processus ponctuels temporels, accompagnée d'une classification de ces modèles. Par la suite, nous identifions les tests statistiques de séries d'un seul type d'événements et nous examinons leur applicabilité pour une série de deux ou de plusieurs types d'événements. Les tests statistiques identifiés sont répartis en quatre classes : analyse graphique, tests appliqués au processus de Poisson homogène et non homogène, tests appliqués au processus de renouvellement homogène et les tests de discrimination entre deux processus ponctuels. Ce travail est réalisé avec l'idée d'une application ultérieure dans le cadre de l'analyse du risque.

Les résultats de cette recherche ont montré qu'il n'existe dans la littérature que des tests d'une série d'un seul type d'événements et ils sont, généralement, valables pour les processus ponctuels suivants : Poisson homogène et renouvellement homogène. L'application de ces tests aux séries de deux ou de plusieurs types d'événements est possible dans le cas où les événements sont définis par leurs nombres et leurs temps d'occurrence seulement, i.e. la durée de chaque événement n'est pas prise en considération.

Mots clés : *processus ponctuel, processus alterné, événement, stationnarité, indépendance des intervalles, classification, risque.*

1 – INTRODUCTION

La construction et la gestion des ouvrages hydrauliques nécessitent une bonne connaissance des événements naturels extrêmes qui peuvent être enregistrés dans ces sites d'intérêt. Toutefois, cette information n'est pas suffisante pour la détermination d'un certain nombre de paramètres qui assurent une durée de vie optimale de ces ouvrages. Ainsi, on a recours à une alternative plus rationnelle, il s'agit de la modélisation. Dans sa définition la plus large, un modèle est une représentation théorique simplifiée d'un phénomène réel. Pour mieux représenter cette réalité, on identifie, généralement, cinq principales étapes au cours du processus de la modélisation : i) définition du problème et

des variables d'intérêt, ii) formulation des hypothèses et des équations du modèle qui lient ces variables, iii) estimation des paramètres du modèle et son application aux données observées, iv) validation du modèle et v) simulation et analyse de sensibilité. Chaque étape est importante dans ce processus et comprend une série de mesures à réaliser. À titre d'illustration, on présente (*figure 1*) ces étapes lors de la modélisation d'une série d'événements par un processus ponctuel temporel. Dans cette étude, on s'intéresse particulièrement à ce type de modèle. Il est à noter qu'il existe des processus ponctuels spatiaux ou spatio-temporels. Un processus ponctuel est défini comme étant un processus stochastique pour lequel chaque réalisation constitue une collection de points. Ce modèle est de plus en plus utilisé dans plusieurs domaines de recherche, là où l'occurrence des événements n'est pas régulière (crues extrêmes, étiages, tremblements de terre, écrasement d'un avion, etc.). Plusieurs ouvrages traitent particulièrement de ces processus, (COX et LEWIS, 1966; SNYDER, 1975 ; COX et ISHAM, 1980), cependant, il existe dans la littérature peu de travaux qui s'intéressent à l'analyse de séries d'événements d'une façon opérationnelle. On définit une série d'événements par une série de données correspondant à une réalisation d'un processus ponctuel. L'analyse de séries d'événements permet de choisir un modèle (processus ponctuel) approprié pour le phénomène observé. On identifie deux catégories de séries d'événements : une série d'un seul type d'événements et une série de plusieurs types d'événements. Les données de la première catégorie proviennent de l'enregistrement des réalisations d'un événement unique, par exemple les enregistrements des tremblements de terre en un site donné. Pour la deuxième catégorie, la série de données est le résultat de la réalisation de plusieurs types d'événements. Par exemple, en partant d'une série des débits journaliers ou mensuels et d'un seuil fixé (niveau de troncation), on identifie deux types d'événements : l'étiage (valeurs inférieures au seuil) et la crue (les valeurs supérieures ou égales au seuil). Dans ce cas, la série d'événements provient de deux type d'événements distincts.

L'objectif de cette étude est de mettre en évidence les différents tests statistiques appliqués aux séries d'un seul ou de plusieurs types d'événements et de proposer une classification de ces tests. On vise par cette étude à fournir aux utilisateurs un outil d'aide pour l'analyse d'une série d'événements et pour le choix du processus ponctuel correspondant. Ainsi, on s'intéresse à la phase de validation du processus de la modélisation. Cette étape permet d'analyser la série d'événements et de vérifier l'adéquation des résultats théoriques aux données empiriques. Elle peut être effectuée en utilisant une nouvelle série de données ou par l'application des tests statistiques. Ces tests peuvent être subdivisés en deux catégories :

1. La première catégorie consiste à tester si une série d'événements peut être représentée par un processus ponctuel bien défini, i.e. discrimination entre deux ou plusieurs processus ponctuels. À titre d'exemple, nos hypothèses H_0 et H_1 peuvent être formulées de la façon suivante :

H_0 : la série d'événements provient d'un processus de Poisson homogène;

H_1 : la série d'événements ne provient pas d'un processus de Poisson homogène.

2. La deuxième catégorie de tests concerne la vérification des hypothèses d'un processus ponctuel, étant donné que c'est le modèle adéquat aux don-

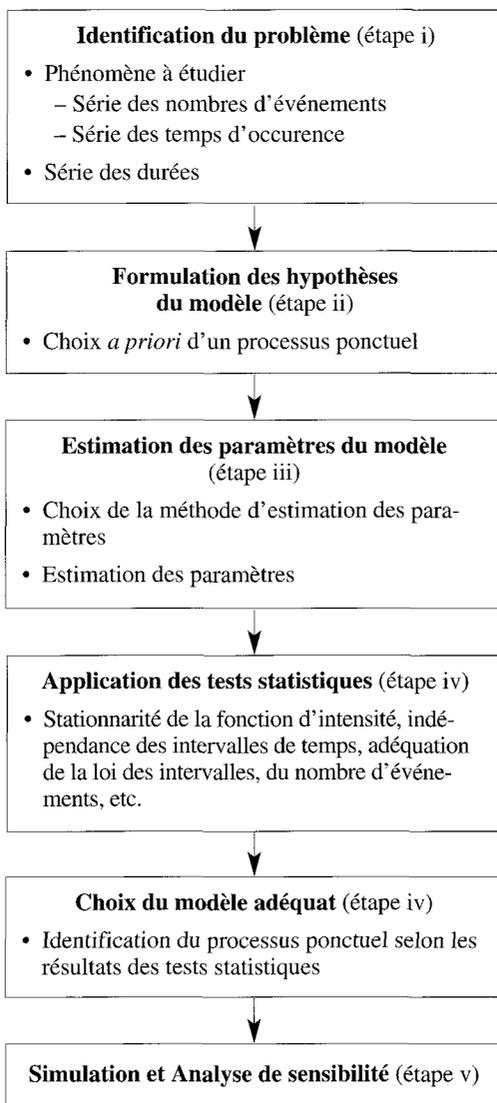


Figure 1 Démarche de la modélisation d'une série d'un seul type d'événements par un processus ponctuel.

nées. Parmi ces hypothèses, on cite l'ajustement de la loi statistique du nombre d'occurrence et des intervalles de temps, la stationnarité de la fonction d'intensité, etc.

Ce travail est réalisé avec l'idée d'une application ultérieure dans le cadre de l'analyse du risque. En effet, l'étude statistique de séries d'événements par les processus ponctuels peut se situer dans un cadre d'analyse du risque, qui comprend deux étapes : l'évaluation du risque et sa gestion. À titre d'exemple, un étiage constitue un risque qui peut mettre en défaillance le système de gestion de la ressource en eau. C'est pourquoi, il est nécessaire d'estimer l'intensité et la

durée de ce risque afin d'identifier les politiques de gestion appropriées. Plusieurs études ont montré qu'il est plus commode de définir le risque par des indices de risque, qui servent à juger et à comparer plusieurs politiques de gestion dans la prise de décision. Dans une étude d'analyse du risque appliquée aux étiages des cours d'eau, ABI-ZEID (1997) a considéré cinq indices du risque basés sur les durées des étiages. À titre d'exemple, on considère l'indice du risque suivant : la probabilité qu'un étiage dure n jours sachant qu'il a débuté à l'instant t .

Cet article est subdivisé en quatre principales sections. La première section comporte une présentation succincte de différents processus ponctuels utilisés dans la littérature. On s'intéresse ensuite, dans la deuxième section, à la description des tests statistiques de séries d'un seul type d'événements. La troisième section traite de l'identification des tests applicables à une série de deux ou de plusieurs types d'événements. Finalement, on termine par une classification des procédures d'analyse identifiées, et par des recommandations sur leurs utilisations.

2 – PROCESSUS PONCTUELS TEMPORELS

Un processus stochastique est une collection de variables aléatoires X_t , où le paramètre t est défini sur un espace paramétrique T . Quand le paramètre t représente le temps, il est défini sur l'échelle des réels non négatifs. Un processus ponctuel temporel est un processus stochastique dont la réalisation est une collection de points, chaque événement (point) e_i ayant une position t_i bien définie dans le temps tel qu'illustré sur la *figure 2* (ABI-ZEID et BOBÉE, 1999).

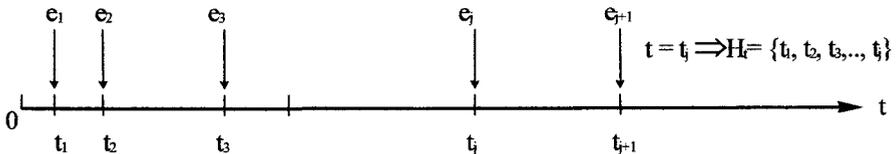


Figure 2 La réalisation d'un processus stochastique ponctuel (d'après ABI-ZEID et BOBÉE, 1999).

2.1 Processus de Poisson homogène, PPH

Le processus de Poisson homogène (PPH) est un choix classique en raison de sa simplicité et de son grand usage. Il est caractérisé par une fonction d'intensité (taux moyen instantané d'occurrence d'un événement, équation 1) constante :

$$\rho(t, H_t) = \lim_{\Delta t \rightarrow 0} \frac{\text{Prob}(N_{t+\Delta t} = 1 / H_t)}{\Delta t} = \lambda ; H_t = t_1, t_2, \dots, t_n \quad (1)$$

où :

Prob(.) représente la probabilité de la réalisation d'un événement dans le temps ;

λ est une constante ;

n est le nombre total d'événements sur une période $(0, t]$ (figure 2). La notation $(., .]$ représente un intervalle borné des deux bords. Cette notation est utilisée pour éviter toute confusion avec celle qui représente la partie entière d'un nombre réel ;

t : est la réalisation de la variable aléatoire temps T définie sur l'ensemble des entiers positifs ;

$N_{t, t+\Delta t}$ représente le nombre d'événements dans l'intervalle de $(t, t+\Delta t]$.

Ce processus est également caractérisé par les hypothèses de stationnarité, de non-multiplicité, et d'indépendance suivantes :

- a) la stationnarité implique que la loi du nombre d'événements enregistré dans un intervalle $(t, t+\Delta t]$ ne dépend que de la quantité Δt , i.e. la probabilité d'occurrence d'un événement est proportionnelle à la longueur d'un petit intervalle et indépendante du temps d'occurrence de l'événement lui-même ;
- b) la non-multiplicité implique que la probabilité d'occurrence de deux ou de plusieurs événements dans un petit intervalle $(t, t+\Delta t]$ est négligeable ;
- c) l'indépendance implique que le nombre d'événements dans un intervalle donné est indépendant du nombre d'événements dans d'autres intervalles disjoints. Ainsi, les probabilités au temps t ne dépendent pas de l'historique du processus, H_t , c'est-à-dire de tous les points dans l'intervalle $(0, t]$, et la probabilité de trouver un point dans l'intervalle $(t, t+\Delta t]$ ne dépend pas du nombre de points avant t , ni du fait qu'il y ait eu ou non un point à t .

Le nombre de d'événements dans un intervalle de longueur t , noté N_t , est distribué selon une loi de Poisson de moyenne λt :

$$\text{Prob}(N_t = n) = \exp(-\lambda t) \frac{(\lambda t)^n}{n!} \tag{2}$$

Dans ce travail, on note la variable aléatoire en majuscule et sa réalisation en minuscule.

Les intervalles, notés X_1, X_2, \dots, X_j , entre les points successifs sont indépendants et distribués selon une loi exponentielle de paramètre λ , GIRAULT (1965) :

$$f_x(x) = \lambda \exp(-\lambda x) \tag{3}$$

2.2 Processus de Poisson non homogène, PPNH

Le processus de Poisson non homogène (PPNH) est une extension du processus de Poisson homogène. Au lieu d'être constante (le cas du PPH, section 2.1), la fonction d'intensité est une fonction du temps, $\rho(t, H_t) = \lambda(t)$. Les hypothèses d'indépendance et de non-multiplicité des événements sont maintenues tandis que l'hypothèse de stationnarité ne l'est plus. Le nombre d'événements N_t est distribué selon une loi de Poisson de paramètre :

$$\tau(t) = \int_0^t \lambda(u) du \tag{4}$$

$\tau(t)$: représente la valeur moyenne du nombre d'observations dans l'intervalle $(0, t]$. Ainsi, l'indépendance des intervalles entre deux événements successifs est conditionnelle au temps d'occurrence du dernier événement.

Il est possible d'homogénéiser un PPNH d'intensité $\rho(t;H_t)$, i.e. passer d'un PPNH à un PPH, par la transformation suivante :

$$t' = \tau(t, H_t) = \int_0^t \rho(u, H_u) du \quad (5)$$

En effet, on sait que :

$$E(N_t) = \int_0^t \lambda(u) du \quad (6)$$

et que dans le cas d'un PPH et d'un PPNH, l'équation 6 s'écrit respectivement :

$$E(N_t) = \int_0^t \lambda du = \lambda t \quad (\text{la fonction d'intensité étant indépendante du temps}) \quad (7a)$$

$$E(N_t) = \int_0^t \lambda(u) du = \tau(t) \quad (7b)$$

En considérant le changement de variable suivant : $t' = \tau(t)$ et en supposant que la fonction d'intensité $\lambda(t)$ possède une forme explicite, on détermine l'expression de $E(N_{t'})$: nombre d'événements dans le nouveau intervalle de temps $(0, t']$. À partir des équations 5 et 7, on a :

$$E(N_{t'}) = \int_0^{t'} \lambda(u) du \quad (8)$$

$$E(N_{t'}) = \int_0^{t'} \frac{d\tau(u)}{du} du = \int_0^{t'} d\tau(u) = \int_0^{t'} dt = 1 \times t' \quad (9)$$

et en remplaçant l'équation 9 dans l'équation 8, on obtient :

$$\frac{d\tau(t)}{dt} = \lambda(t) \quad (10)$$

Le processus obtenu, défini par les temps d'occurrence t'_j , est alors un processus de Poisson homogène d'intensité unitaire. Toutefois, τ est une fonction déterministe du temps si et seulement si le processus ponctuel est un processus de Poisson non homogène. Dans les autres cas, $\lambda(t;H_t)$ est une fonction de H_t et la transformation est alors aléatoire puisqu'elle dépend de la réalisation particulière du processus (COX et ISHAM, 1980).

2.3 Processus Binomial négatif, PBN

Le processus Binomial négatif constitue une bonne alternative au processus de Poisson quand l'indice de dispersion du nombre d'événements (rapport entre la variance et la moyenne du nombre d'événements, équation 14) est supérieur à l'unité (LANG, 1999). Les caractéristiques du PBN sont :

- Le nombre d'événements N_t suit une loi Binomiale négative de paramètre γt :

$$\text{Prob}(N_t) = n = \binom{\gamma t + n - 1}{n} \delta^{\gamma t} (1 - \delta)^n \quad (11)$$

où :

δ est un paramètre compris entre 0 et 1.

À partir de l'équation 11, on obtient les expressions de la moyenne et de la variance du nombre d'événements (équations 12 et 13) :

$$E[N_t] = \frac{\gamma t(1-\delta)}{\delta} \tag{12}$$

$$Var[N_t] = \frac{\gamma t(1-\delta)}{\delta^2} \tag{13}$$

Par conséquent, l'indice de dispersion est égal à :

$$I_d = \frac{Var[N_t]}{E[N_t]} = \frac{1}{\delta} > 1 \tag{14}$$

La distribution des intervalles de temps suit une distribution exponentielle de paramètre $-\gamma \ln \delta$.

Contrairement au PPH, le processus Binomial négatif permet le regroupement des événements car la valeur de l'indice de dispersion est supérieure à l'unité (LANG, 1999).

2.4 Processus de renouvellement homogène, PRH et non homogène, PRNH

Un processus de Poisson dont les intervalles sont indépendants et distribués selon une loi exponentielle, est généralisé à un processus de renouvellement homogène (PRH) lorsque les intervalles $\{X_1, X_2, \dots, X_i\}$ sont indépendants et distribués selon une loi commune arbitraire de densité de probabilité f . L'intensité $\rho(t, H_t)$ d'un PRH dépend uniquement de l'instant d'occurrence du dernier événement, $\rho(t, H_t) = \gamma(t, t_{N_t}) = \rho(t_{N_t})$. Dans le cas où cette fonction dépend du temps, le processus devient un processus de renouvellement non homogène, PRNH.

BERNIER (1967) a présenté deux cas particulier du PRH. Le premier processus est basé sur l'hypothèse que la loi des intervalles de temps est une loi d'Erlang. Sa fonction densité de probabilité est donnée par :

$$f(x) = \frac{\lambda^i x^{i-1} \exp(-\lambda x)}{\Gamma(i)}, \quad i=1, 2, \dots \tag{15}$$

On constate que cette loi se réduit à une loi exponentielle pour une valeur de $i=0$. Ainsi, le processus obtenu est un processus de Poisson homogène, PPH.

Le deuxième processus est obtenu en supposant que la fonction densité de probabilité de X_1 est différente de celle des autres valeurs de X_i . Ce processus est appelé processus de renouvellement retardé. On rappelle que dans le cas d'un PRH les intervalles de temps sont distribués selon une loi commune de densité $f(x)$.

2.5 Processus semi-markovien, PSM (LEVY, 1954 ; SMITH, 1955)

La distribution des intervalles d'un processus stochastique est représentée par une chaîne de Markov. À titre d'exemple, soient k fonctions de répartition F_1, \dots, F_k et supposons que le processus est dans l'état j à un instant t avec une

fonction F_j . Alors, on suppose que la séquence des états (série des intervalles) du processus est déterminée par une chaîne de Markov avec une matrice des probabilités de transition, notée $P=(P_{ij})$, telle que si le processus est dans l'état i , la probabilité que le prochain état sera dans l'état j est P_{ij} (COX et ISHAM, 1980). Le processus ponctuel déterminé par ce type de séquences est appelé processus semi-markovien. Dans le cas d'un processus de renouvellement non homogène, la distribution de la séquence des intervalles x_i est supposée connue. Cependant, le processus semi-markovien représente la séquence des intervalles par une chaîne de Markov.

2.6 Processus de Poisson doublement stochastique, PPDS (COX, 1955)

Le processus de Poisson doublement stochastique (PPDS), appelé également processus de Cox, est caractérisé par une fonction d'intensité, $\Lambda(t, Y_t)$, aléatoire et non stationnaire, où $\{Y_t\}$ représente un processus d'information externe. Dans le cas où ce processus est connu, le PPDS devient un processus de Poisson non homogène conditionnel (PPNHC) de fonction d'intensité $\lambda(t, Y_t=y_t; t \geq 0)$, (ABI-ZEID et BOBÉE, 1999). En effet, le PPNHC est un PPNH qui tient compte de certaines variables explicatives (non aléatoire) et pouvant influencer ou expliquer le comportement du processus.

2.7 Processus de Poisson groupé, PPG (BARTLETT, 1963 ; LEWIS, 1964)

Il s'agit d'associer à chaque réalisation du processus de Poisson un nombre aléatoire de points formant un processus subsidiaire tout en respectant la condition de non multiplicité du processus initial. On obtient un processus de Poisson groupé en superposant les points de chaque processus subsidiaire (COX et ISHAM, 1980). Afin de spécifier un PPG, il est nécessaire de déterminer le processus qui génère les centres des groupes et les mécanismes qui gouvernent chaque groupe. Ainsi, le processus PPG doit vérifier deux conditions suivantes : les tailles des groupes doivent être indépendamment et identiquement distribuées (i.e. la stationnarité est exigée) et le processus des centres de chaque groupe est représenté par un processus de Poisson homogène. Deux cas spécifiques du PPG sont le processus de Neyman-Scott, où les points de chaque groupe sont distribués identiquement et indépendamment autour du centre et celui de Bartlett-Lewis où les intervalles de temps séparant deux points successifs de chaque groupe suivent une distribution statistique quelconque f , (i.e. Le processus subsidiaire est un processus de renouvellement, section 2.4). Ces deux modèles ont été utilisés dans plusieurs études (COX et ISHAM, 1980 ; KAVVAS, 1975 ; WAYMIRE et GUPTA, 1981).

2.8 Processus linéaire auto-excité, PLAÉ (SNYDER, 1975)

Le processus linéaire autoexcité (PLAÉ) est un processus plus général comparativement aux autres modèles. Il est obtenu à partir d'un processus de Poisson en permettant à l'intensité de dépendre du processus ponctuel lui-même. Ainsi, le nombre d'événements N_t passés et leurs temps d'occurrence $(t_1, t_2, \dots, t_{N_t})$ peuvent influencer le nombre d'événements futurs et leurs temps d'occurrence (Abi-Zeid et Bobée, 1999).

L'intensité d'un PLAÉ $\lambda(t, N_t; t_1, t_2, \dots, t_{N_t}), < t \leq t_{N_t+1}$ est un processus aléatoire ayant des chemins connus, étant donné l'historique du processus ponctuel H_t .

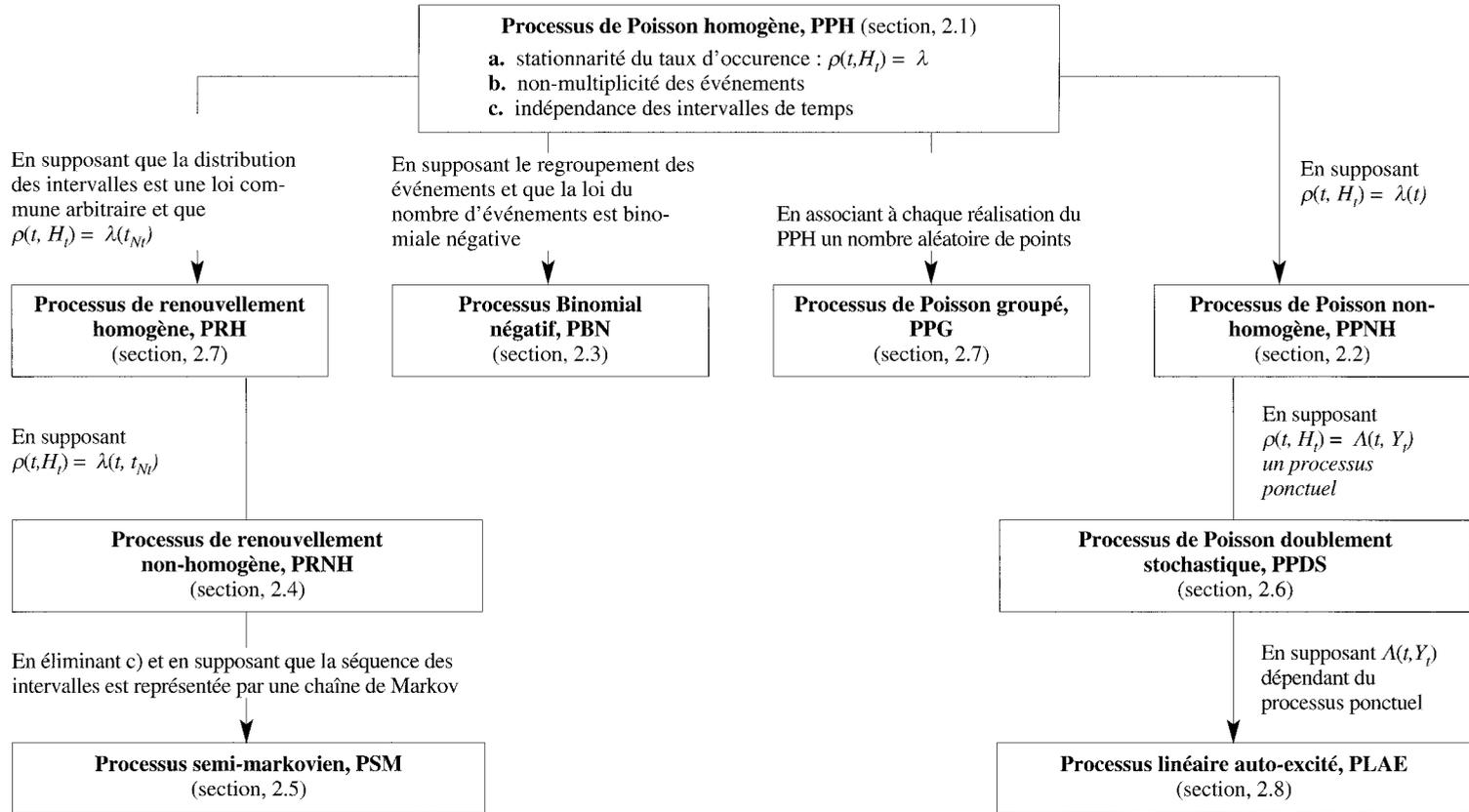


Figure 3 Illustration des liens entre les différents processus ponctuels.

La figure 3 illustre les liens entre les processus ponctuels présentés dans le cadre de ce chapitre.

2.9 Quelques applications des processus ponctuels

On présente quelques travaux qui ont été réalisés sur la base de la théorie des processus ponctuels. Dans le cadre de la modélisation des phénomènes naturels par les processus ponctuels, KAVVAS (1975) a modélisé les précipitations en utilisant les processus de Poisson non homogène et le processus ponctuel de Neyman-Scott, tout en vérifiant les hypothèses de chaque processus. SMITH (1980) a examiné la possibilité de modéliser les données de précipitations par les processus de Poisson, de Poisson groupé, de renouvellement, et de Cox. DIGGLE (1983) a employé la théorie des processus ponctuels spatiaux pour la modélisation de plusieurs variables en biologie : à titre d'exemple, l'auteur a modélisé le nombre des arbres dans une forêt afin de mieux connaître la répartition des oiseaux dans cette zone d'étude. KARR (1986) s'est intéressé aux inférences statistiques à partir des processus de Poisson, de renouvellement et de Cox. THOMPSON (1988) a également utilisé les processus ponctuels pour modéliser les arrivées des particules radioactives dans un compteur Geiger, les collisions successives d'une particule avec d'autres particules dans la théorie cinétique d'un gaz parfait, etc. Dans une autre étape, cet auteur a exploité ses résultats dans le domaine de la sécurité et de la fiabilité des systèmes. BRILLINGER (1994) a appliqué les processus ponctuels dans le domaine de l'hydrologie et de la sismologie. ONOF et WHEATER (1993) et KHALIQ et CUNNANE (1996) ont appliqué le modèle de Bartlett-Lewis pour des données de précipitations horaires. ABI-ZEID (1997) a effectué une évaluation des processus ponctuels pour modéliser les occurrences et les durées des étiages et des surplus afin de caractériser et de quantifier les risques qui en résultent, tout en tenant compte de la non stationnarité du phénomène.

Nous présentons dans le *tableau 1* quelques applications opérationnelles des processus ponctuels en l'hydrologie. On remarque que le PPG est souvent utilisé pour la modélisation des données de précipitations.

Tableau 1 Applications opérationnelles des processus ponctuels en hydrologie.

Processus Ponctuels	Type de données	Auteurs
PPNH ¹ , PPG ² (Neyman-Scott)	Données de précipitations	KAVVAS (1975)
PRNH ³ et PPDS ^{4*}	Données de précipitations	SMITH (1980)
PPNH et PRNH ⁵ alternés	Débits journaliers	NORTH (1981)
PPG	Données de précipitations et de débits de crue	CERVANTES et al. (1983)
PPG (Bartlett-Lewis)	Données de précipitations horaires	ONOF et WHEATHER (1993) ; KHALID et CUNNANE (1996)
PPNH, PRNH et PPNHC ⁶ alternés	Débits journaliers et données de précipitations et de température	ABI-ZEID (1997)

1 : Processus de Poisson non homogène, section 2.2. ; 2 : Processus de Poisson groupé, section 2.7. ; 3 : Processus de renouvellement homogène, section 2.4. ; 4 : Processus de Poisson doublement stochastique, section 2.6. ; 5 : Processus de renouvellement non homogène, section 2.4. ; 6 : Processus de Poisson non homogène conditionnel, section 2.6.

3 – TESTS STATISTIQUES D'UNE SÉRIE D'UN SEUL TYPE D'ÉVÉNEMENTS

Plusieurs travaux de recherche ont été effectués dans le cadre de l'identification des tests statistiques applicables à la vérification des hypothèses d'un modèle donné, telles que l'indépendance, la stationnarité, l'homogénéité des séries de données. Deux récentes revues bibliographiques, ONDO *et al.* (1997) et FAUCHER *et al.* (1997), ont identifié ces tests pour des séries chronologiques (dans ce cas-ci, les séries de données sont considérées à pas de temps constant : le jour, le mois, etc.). Toutefois, ces tests ne sont pas applicables aux séries d'événements à cause de la variabilité des intervalles de temps séparant deux événements consécutifs. Dans cette partie, nous nous intéressons aux procédures d'analyse et aux tests statistiques appliqués à une série d'un seul type d'événements. Ces tests seront répartis en quatre classes : analyse graphique, tests appliqués au processus de Poisson homogène et non homogène, tests appliqués au processus de renouvellement homogène et les tests de discrimination entre deux processus ponctuels. Les techniques d'analyse des propriétés des dénombrements et des intervalles présentées dans cette section sont basées en grande partie sur les travaux de COX et LEWIS (1966).

3.1 Analyse graphique

L'analyse graphique de séries de données est une étape exploratoire, mais nécessaire pour détecter toute forme de tendance, de dépendance ou de périodicité. KAVVAS (1975) a proposé cinq représentations graphiques pour analyser les données d'une série d'un seul type d'événements :

- Le nombre d'événements en fonction du temps d'occurrence

Ce graphique est utilisé pour la détection de la tendance dans la série d'événements. Dans le cas d'un processus stationnaire, la pente de la courbe obtenue doit être constante.

Ce test graphique peut être complété par le calcul de l'intervalle de tolérance du nombre d'événements N_t (LANG *et al.*, 1999). Dans le cas d'un PPH, l'intervalle de tolérance est estimé à partir de la distribution de probabilité conditionnelle suivante :

$$w_k(t) = \text{Prob}(N_t = k / N_{t_{fin}} = n) = \binom{n}{k} \left(\frac{t}{t_{fin}}\right)^k \left(1 - \frac{t}{t_{fin}}\right)^{n-k} \tag{16}$$

où : t_{fin} est le temps de l'horizon des observations et n représente le nombre d'événements enregistré dans l'intervalle $(0, t_{fin}]$.

On démontre que l'intervalle de tolérance de niveau $(1-\alpha)$ est obtenu à partir des deux égalités suivantes (LANG *et al.*, 1999) :

$$\sum_{k=0}^{N_{ij}(\alpha/2)} w_k(t_j) = \frac{\alpha}{2} \tag{17a}$$

$$\sum_{k=0}^{N_{ij}(1-\alpha/2)-1} w_k(t_j) = 1 - \frac{\alpha}{2} \tag{17b}$$

où : $N_{ij}(\alpha)$ représente le quantile de N_{ij} de probabilité α .

- Le taux moyen d'occurrence au temps t

$$\lambda_\tau = \frac{N_{t,t+\tau}}{\tau} \tag{18}$$

Pour cette analyse, on doit choisir une période de temps τ constante (un jour, une semaine, un mois, etc.). La variable $N_{t,t+\tau}$ est le nombre d'occurrence durant l'intervalle $(t, t+\tau]$, $\forall t$. Ce graphique permet de visualiser l'existence des fluctuations locales dans le taux moyen d'occurrence λ_τ , i.e. détection de la tendance. Dans le cas d'un processus ponctuel stationnaire, ce taux est représenté par une droite horizontale.

- La fonction d'intensité en fonction du temps

Sous l'hypothèse de stationnarité, la fonction intensité est donnée par :

$$m_f(\tau) = \frac{dM_f(\tau)}{d\tau} = \lim_{\Delta t \rightarrow 0} \frac{\text{Prob}(\text{événement} \in (t + \tau, t + \tau + \Delta t) / \text{événement à } t)}{\Delta t} \tag{19}$$

$M_f(\tau)$ est appelée la fonction densité de renouvellement. Dans le cas du processus de Poisson homogène, la fonction d'intensité est représentée par une droite horizontale ($m_f(\tau) = \lambda$).

- Le spectre du nombre d'événements pour l'analyse de la périodicité

L'application de l'analyse spectrale est nécessaire pour la détection de la périodicité dans la série de données. On définit le spectre des dénombrements par :

$$g_+(\omega) = \int_0^\infty \exp(i\omega t) \gamma(t) dt \tag{20}$$

où : $\gamma(t) = \lim_{s \rightarrow 0} \text{cov} \left(\frac{N_{t+u+s} - N_{u+1}}{s}, \frac{N_{u+s} - N_u}{s} \right)$

Le périodogramme est l'estimateur du spectre des dénombrements défini par (KAVVAS, 1975) :

$$\hat{g}_+(\omega_p) = \frac{1}{\pi T} I_{\omega_p}, \text{ où } I_{\omega_p} = [A_T^2(\omega_p) + B_T^2(\omega_p)] \tag{21}$$

$$A_T(\omega_p) = \sum_{j=1}^n \cos(t_j \omega_p), \text{ et } B_T(\omega_p) = \sum_{j=1}^n \sin(t_j \omega_p) \tag{22}$$

où $\omega_p = \frac{p^2 \pi}{T}$, $p = 1, 2, \dots, [n/2]$. Ici p, est dit l'indice de fréquence, T est l'horizon d'observation, et t_j est le temps d'arrivée de l'événement e_j .

Un spectre des dénombrements $g_+(\omega)$ non plat peut révéler des périodicités et une structure de dépendance des dénombrements. Toutefois cette théorie est basée sur les processus stationnaires. Une déviation d'un spectre plat (spectre d'un PPH) peut donc être une déviation d'un processus de Poisson. Le spectre d'un processus de Poisson d'intensité ρ , normé par π/ρ lissé sur k périodes, et multiplié par $2k$, est distribué selon une loi de χ^2 à $2k$ degrés de liberté.

- La fonction de la variance temporelle $V(t)$

L'analyse de la stationnarité par cette fonction n'est pas toujours recommandée, car les estimations de $V(t)$ pour différentes valeurs de t sont très corrélées entre elles (COX ET LEWIS, 1966).

ABI-ZEID, (1997) a utilisé une autre représentation graphique. Il s'agit de la fonction de survie logarithmique empirique, basée sur les intervalles classés en ordre croissant. Pour un processus de Poisson homogène, le graphique de

$\ln\left\{1 - \frac{i}{n_0 + 1}\right\}$ en fonction de x_i est représenté par une droite, n_0 étant le nombre total d'événements.

LANG (1999) a proposé un test graphique de l'adéquation de la loi exponentielle aux séries des intervalles de temps des processus suivant : PPH, PBN et le processus Binomial, PB. Le nombre d'événement de ce dernier processus suit une distribution Binomiale. Ainsi, il représente un cas particulier des processus PPH et PBN décrits, respectivement, dans les sections 2.1 et 2.3. Le test consiste à représenter graphiquement les valeurs théoriques et expérimentales de $\text{Log}[1 - \text{Prob}(X_i < x)]$ en fonction des valeurs de $X_i, i=1, \dots, n$. On démontre, dans le cas d'un PPH, que la courbe théorique est représentée par une droite dont la pente est égale à λ (paramètre de la loi de Poisson).

3.2 Tests relatifs au processus de Poisson

3.2.1 Analyse de tendance

3.2.1.1 Détection de tendance à partir du taux d'occurrence

i. Objet et hypothèses du test

Examiner l'existence d'une tendance dans le taux d'occurrence par la méthode de la régression linéaire. Les hypothèses du test sont :

H₀ : Il n'existe pas de tendance dans la série d'événements ;

H₁ : Il existe une tendance dans la série d'événements.

ii. Limite et portée du test

Ce test n'est valable que pour un processus ponctuel de Poisson non homogène et on suppose que la taille de la série d'événements est assez grande. La fonction d'intensité doit avoir une forme exponentielle dans le temps.

iii. Méthode

Supposons l'existence d'une certaine tendance dans le taux d'occurrence, représentée par un changement dans le temps :

$$\lambda(t) = \exp(\alpha + \beta t) \tag{23}$$

Au voisinage de $\beta = 0$, la tendance est linéaire.

Dans le cas d'un PPNH, on fait l'hypothèse que les intervalles de temps sont indépendants, étant donné le dernier événement, et que la probabilité d'occurrence d'un événement dans l'intervalle $(t_{i+1}, t_{i+1} + \Delta t]$ est donnée par :

$$\text{Prob}\left[N_{t_{i-1}, t_{i-1}+\Delta t} = 1\right] = \lambda(t_{i-1}) \exp\left\{-\int_{t_i}^{t_{i-1}} \lambda(u) du\right\} \Delta t + o(\Delta t) \quad (24)$$

Si la série est observée dans l'intervalle $(0, t_0]$ et si les événements résultent aux temps t_1, \dots, t_n , la fonction de vraisemblance est donnée par :

$$L = \prod_{i=1}^n \text{Prob}\left[N_{t_i, t_i+\Delta t} = 1\right] = \left\{ \prod_{i=1}^n \lambda(t_i) \right\} \exp\left\{-\int_0^{t_0} \lambda(u) du\right\} \quad (25)$$

En remplaçant $\lambda(u)$ par sa valeur (équation 23), et en divisant l'équation 25 par la probabilité marginale des n événements, on obtient une distribution conditionnelle $L(\beta)$ qui dépend uniquement du paramètre β . La transformée logarithmique de la fonction de vraisemblance conditionnelle est donnée par (COX et LEWIS, 1966) :

$$LL = \text{Log}[L(\beta)] = n \text{Log} \beta - n \text{Log}(e^{\beta t_0} - 1) + \beta \sum_{i=1}^n t_i + \text{Log} n! \quad (26)$$

Afin de tester l'hypothèse nulle $H_0: \beta = \beta_0$, on construit la statistique suivante :

$$S_L = \frac{LL'(\beta_0)}{\sqrt{I(\beta_0)}} \quad (27)$$

La fonction $LL'(\beta)$ représente la dérivée première par rapport à β de la fonction $LL(\beta)$ et la fonction $I(\beta)$, dite fonction d'information, est définie par :

$$I(\beta) = E\left\{-\frac{\partial^2 LL}{\partial \beta^2}(\beta)\right\} \quad (28)$$

Sous l'hypothèse nulle, cette statistique suit asymptotiquement une loi normale centrée réduite $N(0, 1)$.

iii. Cas particulier du test

Un cas particulier de ce test correspond à $\beta_0 = 0$, i.e. on teste l'hypothèse $H_0: \beta = 0$. Dans ce cas, COX et LEWIS (1966) ont considéré la statistique :

$$S = \sum_{i=1}^n T_i \quad (29)$$

Cette statistique suit une distribution rectangulaire de la somme de n variables aléatoires indépendantes. Ainsi, CRAMER (1946) propose la statistique U basée sur l'équation 27 :

$$U = \frac{S - \frac{1}{2}nt_0}{t_0 \sqrt{n/12}} \quad (30)$$

qui tend rapidement à la forme normale centrée réduite. Ce test est optimal quand l'hypothèse alternative est basée sur l'expression de la fonction d'intensité donnée par l'équation 23, où $\beta \neq 0$.

3.2.1.2 Analyse de régression des intervalles

i. Objet et hypothèses du test

Examiner l'existence d'une tendance dans la série des intervalles de temps séparant chaque deux événements. Les hypothèses du test sont :

H₀ : Il n'existe pas de tendance dans la série d'événements ;

H₁ : Il existe une tendance dans la série d'événements.

ii. Limite et portée du test

Ce test n'est valable que pour un processus ponctuel de Poisson non homogène et on suppose que la taille de la série d'événements est assez grande.

iii. Méthode

Afin d'assurer les hypothèses de la régression linéaire, on choisit un entier *l* de préférence inférieur à quatre pour ne pas avoir de changement dans le taux d'occurrence durant la réalisation des *l* événements. Soient *y*₁ le temps observé du début jusqu'au *l*^{ième} événement, *y*₂ le temps du *l*^{ième} jusqu'au *2l*^{ième} événement, etc. Ainsi, on a une série d'intervalles de temps *y*₁, ..., *y*_{*r*}. Si le processus est poissonnien, on peut écrire que (COX et LEWIS, 1966) :

$$E[\text{Log}(Y_i)] = -\text{Log}\lambda_i + e_i \tag{31}$$

$$\text{var}[\text{Log}(Y_i)] = v_i = \text{constante} \tag{32}$$

où *e*_{*i*} et *v*_{*i*} sont des constantes indépendantes de λ_i . COX et LEWIS (1966) ont démontré que :

$$v_i \approx \frac{1}{l - 1/2} \tag{33}$$

et que la distribution des Log *Y*_{*i*} est une loi Log Chi-deux à 2*n* degrés de liberté.

On suppose que les observations sont générées par un processus de Poisson avec les hypothèses de non-multiplicité et d'indépendance (hypothèses b et c, section 2.1) et que le taux d'occurrence du processus varie légèrement avec le temps. Alors, on construit le modèle suivant :

$$\text{Log}\lambda_i = \alpha + \beta z_i, \text{ d'où :} \tag{34}$$

$$\lambda_i = \exp(\alpha + \beta z_i) ; i=1, \dots, r \tag{35}$$

*z*_{*i*} est une variable indépendante et peut être représentée par (1) la valeur centrale de l'intervalle *y*_{*i*} ; (2) le numéro d'événement au centre de l'intervalle *y*_{*i*}. Ce cas se présente si λ est exprimé en fonction du nombre d'événements ; ou (3) la moyenne des valeurs comprises dans l'intervalle *y*_{*i*}.

En remplaçant la valeur de Log λ_i de l'équation 34 dans l'équation 31, on obtient :

$$E[\text{Log}(Y_i)] = -(\alpha' + \beta z_i) \tag{36}$$

Ainsi, on peut estimer les valeurs de α' et β par la méthode des moindres carrés; comparer les résidus quadratiques de la régression avec la valeur théorique *v*_{*i*} ; et tester l'hypothèse nulle $H_0 : \beta=0$ (pas de tendance). On peut établir,

par analogie, un modèle de régression en fonction du nombre d'événements n dans un intervalle de temps donné ou en fonction de ses valeurs transformées, COX et LEWIS (1966).

3.2.1.3 Détection de tendance à partir des intervalles

i. Objet et hypothèses du test

Examiner l'existence d'une tendance dans le taux d'occurrence par la méthode de la régression linéaire. Les hypothèses du test sont :

H_0 : Il n'existe aucune tendance dans la série d'événements ;

H_1 : Il existe une tendance dans la série d'événements.

ii. Limite et portée du test

Ce test n'est valable que pour un processus ponctuel de Poisson non homogène et on suppose que la taille de la série d'événements est assez grande.

iii. Méthode

Dans la section 3.2.1.1, nous avons supposé que le taux d'occurrence $\lambda(t)$ est une fonction exponentielle du temps. Dans ce cas, on considère la forme suivante du taux d'occurrence :

$$\lambda_i = \theta + \varphi z_i \quad (37)$$

où z_i est la variable explicative définie dans la section 3.2.1.2. Les paramètres θ et φ sont estimés par la méthode du maximum de vraisemblance.

La stationnarité dans ce cas est vérifiée en testant l'hypothèse suivante :

$$H_0 : \varphi = 0$$

COX et LEWIS (1966) ont considéré la statistique $\sum x_i z_i$ conditionnellement à $\sum x_i$ et ont montré que :

$$R' = \frac{\sum_i (z_i - \bar{z})(x_i - \bar{x})}{\left(\sum_i x_i \right) \left[\frac{\sum_i (z_i - \bar{z})^2}{n(n+1)} \right]^{1/2}} \quad (38)$$

suit une loi normale de moyenne zéro et de variance unitaire. Finalement, on mentionne qu'il est possible d'effectuer une analyse par la méthode du maximum de vraisemblance des nombres d'événements résultant dans différents intervalles.

3.2.1.4 Test d'uniformité conditionnelle

i. Objet et hypothèses du test

Détecter une tendance dans la série d'événements représentée par un processus de Poisson homogène. Les hypothèses du test sont :

H_0 : La série d'événement est représenté par un processus de Poisson homogène ;

H₁ : La série d'événement est représenté par un processus de Poisson non homogène.

ii. Méthode

Soit une série d'événements de taille n_0+1 dont les temps d'occurrence sont t_1, \dots, t_{n_0+1} . x_i représentent les temps séparant ces événements ($i=1, \dots, n_0+1$). L'hypothèse nulle peut être formulée de la façon suivante :

$$\mathbf{H}_0 : F(x_1, \dots, x_{n_0+1}) = \prod_{i=1}^{n_0+1} \left\{ 1 - \exp\left(-\frac{x_i}{E(X)}\right) \right\} \quad (\text{Processus de Poisson homogène}).$$

BARNARD (1953) a proposé un changement de variable $U_{(i)} = \frac{T_i}{T_{n_0+1}}$ conditionnellement à la valeur observée t_{n_0+1} de T_{n_0+1} . $U_{(i)}$ est une variable aléatoire (v.a) indépendamment distribuée selon une loi uniforme sur l'intervalle $(0,1]$, c'est pourquoi ce test est appelé test d'uniformité conditionnelle. Finalement, on peut tester l'hypothèse nulle par un test d'ajustement de la loi uniforme tel que le test de Kolmogorov-Smirnov ou de Cramér-von Mises (COX et LEWIS, 1966). Ces tests d'ajustement sont décrits en annexe. Il est à remarquer que la représentation graphique de la distribution empirique $F_{n_0}(u) = \frac{\text{nombre de } u_i \leq u}{n_0}$ en fonction du

temps est équivalente au graphique du nombre d'événements en fonction du temps (section, 3.1), d'où la sensibilité de ce test à l'hypothèse de stationnarité.

3.2.1.5 Détection de rupture pour des observations poissonniennes

GHORBANZADEH et LOUNES (1996) ont présenté deux méthodes pour la détection de rupture des observations poissonniennes : l'une basée sur le rapport du maximum de vraisemblance et la seconde basée sur la théorie bayésienne. Nous présentons dans le cadre de ce travail les principes de la seconde méthode, étant donné que la première nécessite une grande connaissance de théorie mathématique.

i. Objet et hypothèses du test

Détection d'une rupture dans une série d'observations poissonniennes (série des nombres d'événements). Cette rupture est liée à la non-stationnarité du taux d'occurrence du processus. Les hypothèses du test sont :

$$\mathbf{H}_0 : P(M_i = m_i) = \exp\left(-n_i \lambda_0 \frac{(n_i \lambda_0)^{m_i}}{m_i!}\right) \quad i = 1, \dots, n \quad (39a)$$

$$\mathbf{H}_1 : P(M_i = m_i) = \begin{cases} \exp\left(-n_i \lambda_1 \frac{(n_i \lambda_1)^{m_i}}{m_i!}\right) & \text{si } i = 1, \dots, k \\ \exp\left(-n_i \lambda_2 \frac{(n_i \lambda_2)^{m_i}}{m_i!}\right) & \text{si } i = k + 1, \dots, n \end{cases}, 1 \leq k \leq n - 1 \quad (39b)$$

où M_1, \dots, M_n peuvent être considérées comme une somme de n_i variables aléatoires et les paramètres λ_0 et (λ_1, λ_2) représentent, respectivement, les paramètres de la loi de Poisson sous les hypothèses H_0 et H_1 .

ii. Limite et portée du test

Cette méthode est relativement simple à mettre en œuvre, cependant, ses résultats restent fortement liés aux choix des lois a priori des paramètres λ_1 , λ_2 et du paramètre de rupture k .

iii. Méthode

GHORBANZADEH et LOUNES (1996) ont proposé le modèle suivant pour l'approche bayésienne :

$$X_i \sim \text{Poisson}(n_i \lambda_1) \text{ si } i = 1, \dots, k$$

$$X_i \sim \text{Poisson}(n_i \lambda_2) \text{ si } i = k + 1, \dots, n$$

Les hypothèses du test décrites par les équations 39a et 39b s'écrivent de la façon suivante :

$$H_0 : k=n \text{ (pas de rupture } \lambda_1 = \lambda_2 \text{)} ;$$

$$H_1 : 1 \leq k \leq n - 1 \text{ (rupture à } k \text{)}.$$

On suppose que les trois paramètres k , λ_1 et λ_2 sont indépendants.

La démarche de cette procédure est composée de trois grandes étapes : 1) définir les distributions a priori des paramètres, 2) construire une loi a posteriori du paramètre k , sachant qu'une rupture est possible pour des valeurs de k comprises entre 1 et $(n - 1)$, et finalement 3) identifier la valeur maximale de la probabilité a posteriori nous indiquant ainsi le point de la rupture dans la série de données.

3.2.2 Test de dispersion appliqué au nombre d'événements

i. Objet et hypothèses du test

Vérification de l'adéquation de la loi de Poisson à la série des nombres d'événements. Les hypothèses du test sont :

$$H_0 : n_i, i=1, \dots, k, \text{ proviennent d'une loi de Poisson de paramètre } \mu ;$$

$$H_1 : n_i \text{ ne proviennent pas d'une loi de Poisson.}$$

ii. Limite et portée du test

Exigence d'une grande taille d'échantillon.

iii. Méthode

Soient n_1, n_2, \dots, n_k k observations d'une v.a discrète N et soit $\bar{n} = \frac{n_1 + \dots + n_k}{k}$ la moyenne des k observations.

Ce test est appelé test de dispersion d'homogénéité et il est basé sur la statistique suivante :

$$d = \sum_{i=1}^k \frac{(n_i - \bar{n})^2}{\bar{n}} \quad (40)$$

Sous l'hypothèse nulle, d suit asymptotiquement une distribution Chi-deux à $(k - 1)$ degrés de liberté (RAO et CHAKRAVARTI, 1956). SUKHATME (1938) a montré que l'approximation de d est adéquate pour des valeurs de $k > 5$ si $\bar{n} > 1$ ou bien pour des valeurs $k > 15$ si $\bar{n} < 1$. CUNNANE (1979) a indiqué que l'approximation de la loi de la statistique d est valable pour des valeurs de $\bar{n} > 5$.

Lorsque l'hypothèse nulle n'est pas vérifiée, LANG (1999) a proposé de considérer les lois Binomiale ou Binomiale négative comme alternatives à la loi de Poisson. En effet, une valeur de l'indice de dispersion I_d (défini dans la section 2.3) inférieure à l'unité nous indique que le nombre d'événements N_t peut être représenté par une loi Binomiale. En revanche, une valeur de I_d supérieure à l'unité nous indique que N_t peut être représenté par une loi Binomiale négative.

KAVVAS (1975) a proposé le calcul du coefficient de variation du nombre d'événements (le rapport entre l'écart-type et la moyenne du nombre d'événements). Si ce coefficient est supérieur à l'unité, le processus observé est trop dispersé et s'éloigne du processus de Poisson homogène, nous indiquant ainsi de l'existence d'un regroupement des données. L'utilisation des processus PPG, dans ce cas, serait également une bonne alternative au processus de Poisson.

3.3 Tests relatifs au processus de renouvellement homogène

3.3.1 Tests du coefficient d'autocorrélation entre les intervalles

i. Objet et hypothèses du test

Examiner l'indépendance des intervalles de temps séparant deux événements successifs, autrement dit, examiner si la série d'événements peut être représentée par un processus de renouvellement. Les hypothèses du test sont :

H_0 : Il n'existe aucune dépendance entre les intervalles de temps;

H_1 : Il existe une dépendance entre les intervalles de temps.

ii. Limite et portée du test

Exigence d'une grande taille de la série d'observations. Les résultats de ce test doivent être interprétés avec précaution à cause de la variabilité du pas de temps entre les événements.

iii. Méthode

Le coefficient d'autocorrélation d'ordre k , ρ_k , est défini par la formule suivante :

$$\rho_k = \frac{n}{n-k} \frac{\sum_{j=1}^{n-k} (x_j - \bar{x})(x_{j+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \tag{41}$$

où x_i représente le temps entre deux événements successifs.

Les hypothèses du test se transforment sous la forme suivante :

H_0 : $\rho_k = 0$ (indépendance des intervalles) ;

H_1 : $\rho_k \neq 0$ (dépendance des intervalles).

La statistique du test est donnée par :

$$\hat{\rho}_k \sqrt{n-1} \tag{42}$$

où $\widehat{\rho}_k$ est l'estimation de ρ_k . Cette statistique suit une loi normale $N(0,1)$ à condition que le nombre d'observations soit assez grand.

3.3.2 Tests du coefficient d'autocorrélation des rangs

i. Objet et hypothèses du test

Examiner l'indépendance des intervalles de temps séparant chaque deux événements successifs. Les hypothèses du test :

H_0 : Il n'existe aucune dépendance entre les intervalles de temps ;

H_1 : Il existe une dépendance entre les intervalles de temps.

ii. Limite et portée du test

Exigence d'une grande taille de la série d'observations.

iii. Méthode

COX et LEWIS (1966) ont présenté la statistique suivante :

$$R_j = \sum_{i=1}^{n-j} x_{i+j} x_i \quad (43)$$

où x_i représente le temps entre deux événements successifs.

La procédure de ce test consiste à remplacer les valeurs de x_i par leur rang. Le rang de x_k est déterminé par la façon suivante : 1 + (nombres vérifiant $x_i < x_k$, $i \neq k$). À partir de la statistique définie par l'équation 43, on obtient :

$$E(R_1) = \frac{1}{12} (n-1)(n+1)(3n+2) \quad (44)$$

$$\text{var}(R_1) = \frac{5n^6 + 16n^5 - 14n^4 - 80n^3 - 35n^2 + 64n + 44}{720(n-1)} \quad (45)$$

La loi asymptotique de la statistique R_1 est une loi normale de moyenne et de variance définies respectivement par les équations 44 et 45, (WALD et WOLFOWITZ, 1943).

3.3.3 Test de Bartlett : homogénéité de la variance

i. Objet et hypothèses du test

Tester l'homogénéité de la variance des intervalles de temps séparant les événements. Ainsi, ce test permet de vérifier la stationnarité de ces intervalles. Les hypothèses du test sont :

H_0 : Homogénéité de la variance des intervalles ;

H_1 : Non homogénéité de la variance des intervalles.

ii. Limite et portée du test

Le test de Bartlett suppose que les intervalles proviennent de populations normales.

iii. Méthode

On groupe les intervalles entre les événements X_j en k groupes de taille n_i . σ_k^2 est la variance du $k^{\text{ème}}$ groupe d'intervalles. Les hypothèses du test peuvent s'écrire sous la forme suivante :

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

$$H_1 : \sigma_i^2 \neq \sigma_j^2, \text{ si } i \neq j$$

La statistique du test est donnée par :

$$K = \frac{-2 \log I^*}{1 + \frac{1}{3k-3} \left[\sum_{i=1}^k \frac{1}{v_i} - \frac{1}{v} \right]} \tag{46}$$

où :

$$v_i = n_i - 1, \quad -2 \log I^* = v \log s^2 - \sum_{i=1}^k v_i \log s_i^2$$

$$s_i^2 = \frac{1}{v_i} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2, \quad s^2 = \frac{1}{v} \sum_{i=1}^k v_i s_i^2, \quad v = \sum_{i=1}^k v_i$$

et : x_{ij} est le $j^{\text{ème}}$ intervalle du groupe i .

La statistique K suit approximativement une loi de Chi-deux à $(k-1)$ degrés de liberté et on rejette l'hypothèse nulle à un niveau de signification α si $K > \chi^2_{\alpha}(k-1)$.

3.3.4 Test de tendance dans le périodogramme

i. Objet et hypothèses du test

Tester l'hypothèse d'homogénéité du processus (absence de tendance dans le temps) en utilisant l'estimateur du spectre des intervalles. Les hypothèses du test sont :

H₀ : Il n'existe aucune tendance dans la série d'événements ;

H₁ : Il existe une tendance dans la série d'événements.

ii. Limite et portée du test

Ce test est basé sur celui de Bartlett, qui suppose que les intervalles proviennent de populations normales seulement.

iii. Méthode

On définit l'estimateur du spectre des intervalles par la fonction $I_n(w_p)$ (COX et LEWIS, 1966) :

$$I_n(w_p) = \frac{1}{2\pi n} \sum_{s=1}^n \sum_{u=1}^n X_s X_u e^{i(s-u)w_p} \tag{47}$$

où $0 < w_p = \frac{2\pi p}{n} \leq \pi$ et $p=1,2,\dots,[n/2]$. $[.]$: signifie la partie entière d'un nombre réel.

Dans le cas d'un processus de renouvellement, l'espérance de la quantité définie par l'équation 47 est égale à (COX et LEWIS, 1966) :

$$E[I_n(w_p)] = \frac{\sigma^2}{2\pi} = \sigma^2 f(w_p) \quad (48)$$

Cette quantité représente la puissance du spectre du processus de renouvellement. La forme générale de la fonction $f(w)$ est donnée par l'équation suivante :

$$f(w) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{k=1}^{\infty} \rho_k \cos(kw) \right\}, \quad -\pi \leq w \leq \pi \quad (49)$$

Les hypothèses du test peuvent être formulées de la façon suivante :

H₀ : L'estimateur du spectre $I_n(w_p)$ est indépendamment et asymptotiquement distribué selon une loi exponentielle de moyenne $\sigma^2 f(w)$ (Processus de renouvellement homogène);

H₁ : Existence d'une tendance dans le spectre.

Le rapport $\frac{I_n(w_p)}{f(w_p)}$ peut être considéré comme une estimation de la variance

à partir d'une population normale. On peut, alors, appliquer le test d'homogénéité des variances de Bartlett (BARTLETT, 1937) décrit dans la section 3.3.3.

3.3.5 Test d'indépendance des intervalles basé sur le périodogramme

i. Objet et hypothèses du test

Examiner l'indépendance des intervalles de temps séparant chaque deux événements successifs. Les hypothèses du test sont :

H₀ : Il n'existe aucune dépendance entre les intervalles de temps;

H₁ : Il existe une dépendance entre les intervalles de temps.

ii. Méthode

Ce test d'indépendance des intervalles est basé sur l'estimation du spectre et sur les tests d'ajustement indépendants de la distribution. Soit la quantité :

$$U_{(i)} = \frac{\sum_{p=1}^i \frac{I(w_p)}{f(w_p)}}{\sum_{p=1}^i \frac{I(w_p)}{f(w_p)}} \quad (50)$$

Si $U_{(i)}$ suit une loi uniforme alors l'hypothèse H_0 est vérifiée (indépendance des intervalles). La condition d'uniformité des variables $U_{(i)}$ est vérifiée par l'utilisation d'un test d'ajustement d'une distribution statistique. COX et LEWIS (1966) ont comparé les puissances des tests d'ajustement de Kolmogorov-Smirnov et de Cramér-von Mises et ils ont conclu que le second test est plus approprié pour la vérification de l'adéquation de la loi uniforme aux variables $U_{(i)}$.

3.4 Tests de discrimination entre deux processus ponctuels

3.4.1 Choix entre processus de Poisson et processus de renouvellement

i. Objet et hypothèses du test

On considère un test vérifiant l'hypothèse H_0 que le processus est poissonnien et en fixant comme hypothèse alternative H_1 que le processus est de renouvellement. Les deux hypothèses sont :

H_0 : La série d'événements est représentée par un processus de Poisson homogène ;

H_1 : La série d'événements est représentée par un processus de Renouvellement homogène.

ii. Méthode

Contrairement au processus de Poisson, dont la distribution des intervalles suit une loi exponentielle, la distribution des intervalles d'un processus de renouvellement suit une distribution statistique quelconque. Ainsi, nous avons considéré la distribution Gamma à deux paramètres (un paramètre d'échelle

$\alpha = \frac{\kappa}{\lambda}$ et un paramètre de forme κ). Ce choix est justifié par les liens existant entre les deux lois. En effet, la loi Gamma se transforme en une loi exponentielle en fixant le paramètre de forme κ égal à l'unité. Les deux hypothèses du test peuvent être formulées d'une façon équivalente :

H_0 : Les intervalles de temps entre les événements sont indépendants et suivent une distribution exponentielle ;

H_1 : Les intervalles entre les événements sont indépendants et suivent une distribution Gamma à deux paramètres.

On rappelle que la fonction densité de la loi Gamma à deux paramètres, (λ , κ) est donnée par :

$$f(x) = \left(\frac{\kappa}{\lambda}\right)^\kappa \frac{x^{\kappa-1} e^{-\kappa x/\lambda}}{\Gamma(\kappa)} \tag{51}$$

Selon la notation utilisée dans l'équation 51, les hypothèses du test sont :

H_0 : $\kappa=1$

H_1 : $\kappa \neq 1$

où le paramètre λ est non spécifié. L'estimation de κ peut être réalisé par la méthode du maximum de vraisemblance (COX et LEWIS, 1966). La statistique du test est donnée par MORAN (1951) :

$$I_n = \frac{-\text{Log}\bar{x} + \frac{1}{n} \sum_{i=1}^n \text{Log}x_i}{1 + \frac{n+1}{6n}} \tag{52}$$

Sous l'hypothèse nulle, I_n est approximativement distribué selon une loi de Chi-deux à $(n - 1)$ degrés de liberté.

3.4.2 Test de sélection entre trois processus ponctuels

SMITH et KARR (1985) ont développé une procédure de sélection entre plusieurs processus ponctuels représentant des données de précipitations. L'approche est basée sur les estimateurs du maximum de vraisemblance des paramètres de chaque modèle. SMITH et KARR (1985) ont considéré trois processus ponctuels pour modéliser les précipitations. Il s'agit des processus ponctuels suivants :

- processus de renouvellement homogène (PRH), section 2.4. ;
- processus de Cox à intensité markovienne (PCM), qui est un processus de Poisson non homogène et dont l'intensité est un processus de Markov à deux états 0 et un entier c ;
- processus de Neyman-Scott (PNS), section 2.7.

i. Objet et hypothèses du test

Développer une procédure de sélection entre ces processus ponctuels. Les hypothèses du test sont formulées de la façon suivante :

H_0^i : Le processus ponctuel P_i représente mieux la série des données, $i=1,2,3$;

H_1^j : Le processus ponctuel P_j représente mieux la série des données, $j=1,2,3$ et $j \neq i$.

où $P = (P_1, P_2, P_3) = (\text{PRH}, \text{PCM}, \text{PNS})$.

ii. Limite et portée du test

Cette procédure n'est applicable que dans le cas où il est possible de supposer la fonction d'intensité stochastique.

iii. Méthode

Le test est basé sur le rapport de vraisemblance :

$$\Lambda(n, s_1, \dots, s_n) = -\int_0^t (\lambda_0(u) - \lambda_1(u)) du + \int_0^t \log \left(\frac{\lambda_0(u)}{\lambda_1(u)} \right) dN(u) \quad (53)$$

où : λ_0 et λ_1 sont, respectivement, les intensités stochastiques sous les hypothèses H_0 et H_1 .

La procédure du choix entre le modèle exprimé par l'hypothèse nulle et celui exprimé par l'hypothèse alternative H_1 est constituée de trois étapes : (1) calculer les estimateurs de maximum de vraisemblance pour les modèles relatifs aux hypothèses H_0 et H_1 ; (2) construire le rapport de vraisemblance en utilisant les estimateurs de l'étape 1 ; et (3) le modèle de H_0 est adopté si le logarithme du rapport de vraisemblance est supérieur à zéro, sinon le modèle de H_1 est choisi.

4 – TESTS STATISTIQUES D'UNE SÉRIE DE DEUX OU DE PLUSIEURS TYPES D'ÉVÉNEMENTS

Dans cette section, on s'intéresse à un cas qui est fréquemment rencontré dans la pratique. Il s'agit de l'analyse d'un processus ponctuel global (*output pooled*) obtenu en superposant p processus ponctuels. Dans la littérature, il n'existe pas de tests statistiques ou de procédures d'analyse de ce type de données. Toutefois, nous allons examiner la possibilité d'appliquer les tests présentés dans la section 3 pour un processus ponctuel global.

4.1 Superposition de p processus ponctuels

Afin de mieux visualiser ce type de processus, nous représentons, *figure 4*, un schéma de superposition de deux processus ponctuels, cependant, ce schéma peut être généralisé pour $p > 2$ processus ponctuels.

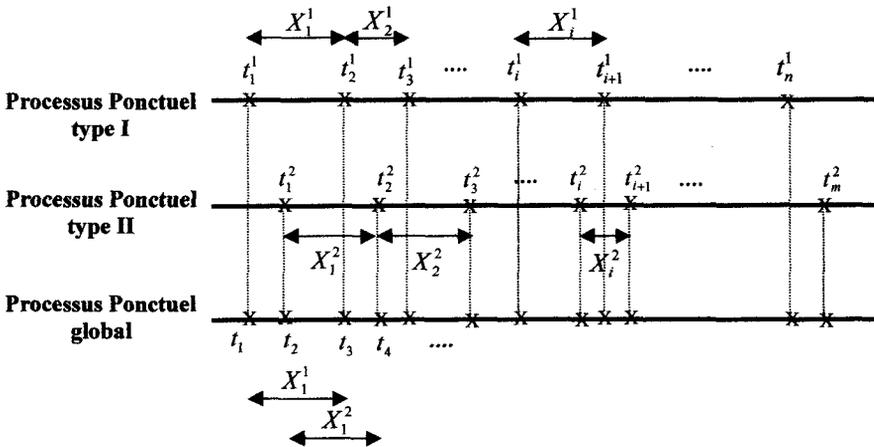


Figure 4 Superposition de deux processus ponctuels.

Soient :

t_k^i : représente le temps d'occurrence du $k^{\text{ième}}$ événement du processus ponctuel i , $i = 1, 2$;

X_k^i : représente l'intervalle de temps séparant deux événements successifs enregistrés aux temps t_k^i et t_{k+1}^i pour le processus ponctuel i , $i = 1, 2$.

On rappelle que l'hypothèse de non-multiplicité b) définie dans la section 2.1 est respectée dans le schéma de la *figure 4* et que le principe de superposition reste le même pour $p > 2$ processus ponctuels. La décomposition du processus global en p processus ponctuels nécessite une identification préalable de chaque type d'événements, caractérisé par les temps d'occurrence et les intervalles de temps séparant les événements consécutifs de même type. Dans ce cas, les temps de séjour (durées des événements) ne sont pas pris en considération. Ainsi, il est possible de modéliser les p processus par la théorie des processus ponctuels et par la suite, les tests statistiques identifiés dans la

section 3 restent valables à ce type de données. L'analyse du processus global exige une connaissance du type des processus ponctuels subsidiaires. Plusieurs travaux de recherche ont été réalisés sur la superposition des processus ponctuels. COX et SMITH (1954) ont réalisé le premier travail de recherche sur la superposition de p processus de renouvellement homogènes. Ils ont montré que le processus global est un processus de Poisson homogène quand p tend vers l'infini. COX et LEWIS (1966) ont montré que, sous les hypothèses de stationnarité et d'indépendance des p processus de Poisson, le processus global est également un processus de Poisson. D'autres résultats ont été démontrés par AMBARTZUMIAN (1965), CINLAR (1968), CINLAR et AGNEW (1968), LAWRENCE (1973), RIPLEY (1976), ITO (1977), CHANDRAMOHAN et LIANG (1985), THOMPSON (1988), et LAM et LEHOCZKY (1991). Cependant, on ne trouve pas dans ces travaux des tests statistiques spécifiques à une série de deux ou de plusieurs types d'événements. Les auteurs supposent généralement les deux hypothèses de stationnarité et d'indépendance des p processus ponctuels, qui sont représentés par un processus de Poisson ou un processus de renouvellement.

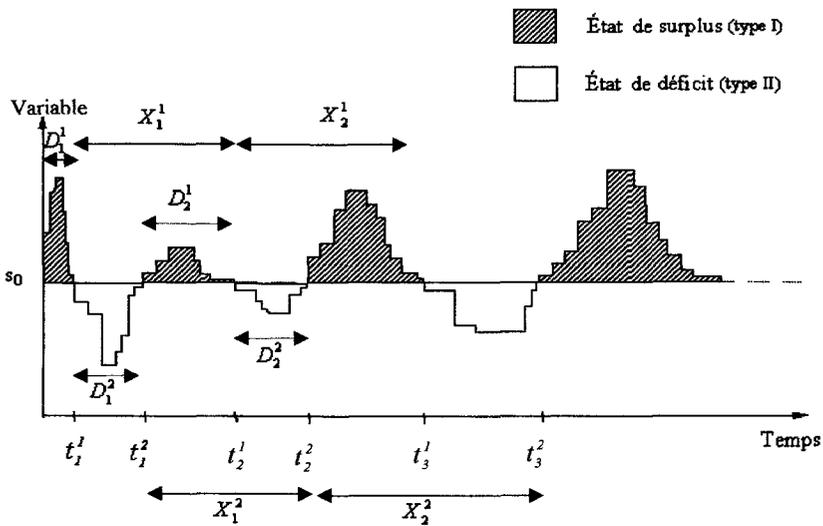


Figure 5 Schématisation du processus alterné.

4.2 Processus alterné

Un cas particulier de la superposition des processus ponctuels est le processus alterné (figure 5). Dans ce cas le processus global est constitué de deux types d'événements, type I et type II, qui sont alternés dans le temps. Les événements sont définis par leurs temps d'occurrence et de séjour (durée de l'événement). Les intervalles de temps de chaque processus peuvent être définis, cependant il n'est pas intéressant de modéliser ces séries de données par un processus ponctuel global car on perdrait la notion « durée » dans un état. Ceci est dû au fait que la variable X_k^i est la somme des deux durées D_k^i , $i=1,2$ (D_k^i est la durée du $k^{\text{ième}}$ l'événement de type i). En effet, ABI-ZEID (1997) a modélisé les séries des débits journaliers par ce type de processus. En fixant un niveau de

troncature s_0 , elle a obtenu deux séries d'événements. Une série des surplus construite à partir des débits supérieurs ou égaux au seuil s_0 et une série de déficits obtenue en considérant les débits inférieurs au seuil s_0 . Ainsi, le processus alterné est composé, d'une part, d'un processus de rupture des surplus, défini par les temps d'occurrence $t_1^1, t_2^1, t_3^1, \dots$, et par les durées de ces événements $D_1^1, D_2^1, D_3^1, \dots$, et d'autre part, d'un processus de rupture des déficits défini par les temps d'occurrence $t_1^2, t_2^2, t_3^2, \dots$, et par les durées de ces événements $D_1^2, D_2^2, D_3^2, \dots$. L'occurrence d'un événement, dans ce cas, est obtenue lors du passage de l'état de déficit à l'état de surplus et *vice versa*.

Pour la validation des hypothèses du processus ponctuel alterné, ABI-ZEID (1997) a appliqué quelques tests statistiques sur les durées (tableau 2). CHEB-CHOUB et al. (1999) les ont appliqués à des données du réseau hydrographique tunisien. Les tests identifiés dans la section 3, et qui sont valables pour les intervalles d'un processus ponctuel à un seul type d'événements, peuvent être appliqués aux durées des deux processus formant le processus alterné.

Tableau 2 Identification des tests statistiques appliqués aux processus alternés (d'après de ABI-ZEID, 1997).

Tests statistiques	Processus Ponctuel
Fonction de survie logarithmique (analyse graphique, section 3.1)	PPNH*
Le test du périodogramme appliqué aux durées homogénéisées de chaque processus : indépendance des durées (section 3.3.5)	PPNH, PRNH**, PNHC***
Adéquation de la loi exponentielle aux durées homogénéisées : statistique de Kolmogorov-Smirnov et Anderson (annexe)	PPNH, PNHC
Adéquation de la loi Gamma aux durées homogénéisées : statistiques de Kolmogorov-Smirnov et Anderson (annexe)	PRNH
Adéquation de la loi des durées non homogénéisées (NORTH, 1981)	PPNH

* : Processus de Poisson non homogène, section 2.2.

** : Processus de Renouveau non homogène, section 2.4.

*** : Processus de Poisson non homogène conditionnel, section 2.6 (ABI-ZEID, 1997).

5 – DISCUSSION

L'objectif de cette étude était d'identifier les tests statistiques appliqués à une série d'événements. Toutefois, nous avons visé à travers cette étude de présenter un article de synthèse qui servira comme référence pour la sélection d'un processus ponctuel et des procédures de validation des hypothèses de ce modèle. Il est clair que le choix d'un processus ponctuel reste une tâche assez délicate, qui nécessite une analyse a priori de la série d'événements et une expertise du chercheur dans le traitement de ce type de données. La figure 3 représente une classification des processus ponctuels décrits dans la section 2. Cette classification complète celle présentée par ABI-ZEID et BOBÉE

(1999) et considère le processus de Poisson homogène comme point de départ afin d'aboutir à des processus plus complexes. Ces processus sont obtenus en considérant leurs hypothèses appropriées. À titre d'illustration, le processus de Poisson non homogène est obtenu en supposant que l'hypothèse de stationnarité du processus de Poisson homogène ne tient plus. Il est à noter que d'autres difficultés peuvent survenir une fois le modèle choisi, il s'agit de la définition de la fonction d'intensité et de l'estimation des paramètres du modèle.

Comme on l'a déjà mentionné, la phase de validation du modèle est très importante dans le processus de la modélisation de séries d'événements. Ainsi, nous avons présenté dans cette étude plusieurs procédures d'analyse de ces séries. Dans la littérature, on n'a rencontré que des tests élaborés pour vérifier la validité des hypothèses d'un processus ponctuel modélisant une série d'un seul type d'événements. Cependant, nous avons tenté d'identifier les tests applicables à une série de deux ou de plusieurs types d'événements. Le problème se pose au niveau du choix du modèle, quand chaque type d'événements est défini par les intervalles de temps séparant deux événements consécutifs du même type et que les temps de séjour ne sont pas nuls. Ainsi, si les durées suivent une loi exponentielle, la série des intervalles suit une distribution qui est la somme de ces deux lois. Finalement, il est important de préciser, dans ce cas, que les tests appliqués aux intervalles, décrits dans la section 3, sont valables pour les durées seulement. Afin de faciliter l'usage de ces tests, nous présentons (*figure 6*) une classification des procédures d'analyse pour une série d'un seul type d'événements. Cette classification est élaborée pour la vérification des hypothèses suivante : la stationnarité de la série d'événements, l'indépendance des intervalles de temps séparant deux événements et l'adéquation des distributions du nombre d'événements et des intervalles de temps. Nous avons également présenté dans cette classification les tests de sélection d'un processus par élimination. Il est important de mentionner que les tests décrits dans la *figure 6* sont majoritairement applicables aux processus de Poisson et de renouvellement. Les procédures d'analyse de tendance sont basées sur le modèle de régression (sections, 3.2.1.1, 3.2.1.2, 3.1.2.3), et ne sont pas applicables quand la série d'événements présente une forme de périodicité. ABI-ZEID (1997) a modélisé les séries d'étiage en considérant une fonction d'intensité qui tient compte de l'aspect non stationnaire et périodique des données. Cet auteur a élaboré une méthode de sélection entre les processus de PPNH, PRNH et PPNHC (cas particulier du PPDS, section 2.6). Nous présentons dans la *figure 7* les étapes de la discrimination entre ces trois processus à partir des durées des événements.

Un autre objectif de cette étude était de présenter aux utilisateurs une méthodologie pour la modélisation des séries d'événements. Il est difficile de formuler une méthodologie standard, car l'application des processus ponctuels reste typique à chaque série de données. Cependant, nous proposons un organigramme de sélection d'un processus ponctuel (*figure 8*). Pour des fins de simplicité, nous supposons que les séries d'événements sont stationnaires, i.e. la fonction d'intensité est constante. Dans le cas contraire, on peut tenir compte de toute forme de tendance en choisissant un processus ponctuel non homogène. Ainsi, la fonction d'intensité de ce processus est définie en considérant le caractère non stationnaire et notamment périodique des événements. La *figure 8* comprend les quatre premières étapes, décrites dans la *figure 1*, de

STATIONNARITÉ

Analyse de tendance

- Analyse graphique :
 - Nombre d'événements en fonction du temps d'occurrence (section 3.1)
 - Fonction de la variance temporelle (section 3.1)
- Analyse de régression des intervalles (Processus de Poisson, section 3.2.1.2)
- Méthodes basées sur la fonction du maximum de vraisemblance (Processus de Poisson)
 - À partir du taux d'occurrence (section 3.2.1.1)
 - À partir de l'analyse des intervalles (section 3.2.1.3)
- Détection de rupture pour des observations poissonniennes
 - Test du rapport de vraisemblance (non présenté dans cet article)
 - Approche bayésienne (section 3.2.1.5)
- Test d'uniformité conditionnelle (Processus de Poisson, section 3.2.1.4)

Homogénéité : analyse graphique

- La moyenne du taux d'occurrence (section 3.1)
- Fonction d'intensité (section 3.1)

Analyse de périodicité

- Le spectre du nombre d'événements (section 3.1)

INDÉPENDANCE DES INTERVALLES

- Tests du coefficient d'autocorrélation entre les intervalles (Processus de renouvellement, section 3.3.1)
- Tests du coefficient d'autocorrélation des rangs (Processus de renouvellement, section 3.3.2)
- Test de Bartlett (Processus de renouvellement, section 3.3.3)
- Tests basés sur l'analyse spectrale (Processus de renouvellement, sections 3.3.4 et 3.3.5)

ADÉQUATION DE LA DISTRIBUTION DU NOMBRE D'ÉVÉNEMENTS ET DES INTERVALLES DE TEMPS

- Adéquation graphique de la loi exponentielle des intervalles de temps (Processus de Poisson, section 3.1)
- Test de dispersion (Processus de Poisson, section 3.2.2)
- Test d'ajustement de Kolmogorov-Smirnov et de Cramèr-Von Mises des distributions du nombre d'événements et des intervalles de temps (Annexe)

TESTS DE SÉLECTION D'UN PROCESSUS PONCTUEL PAR ÉLIMINATION

- Sélection entre le processus PPH, PBN et PB basée sur l'indice de dispersion (section 3.2.2)
- Tests entre processus de Poisson et processus de renouvellement (section 3.4.1)
- Procédure de sélection entre les processus de Cox, de Neyman-Scott et de renouvellement (section 3.4.2)

Figure 6 Récapitulatif des différentes analyses pour une série d'un seul type d'événements.

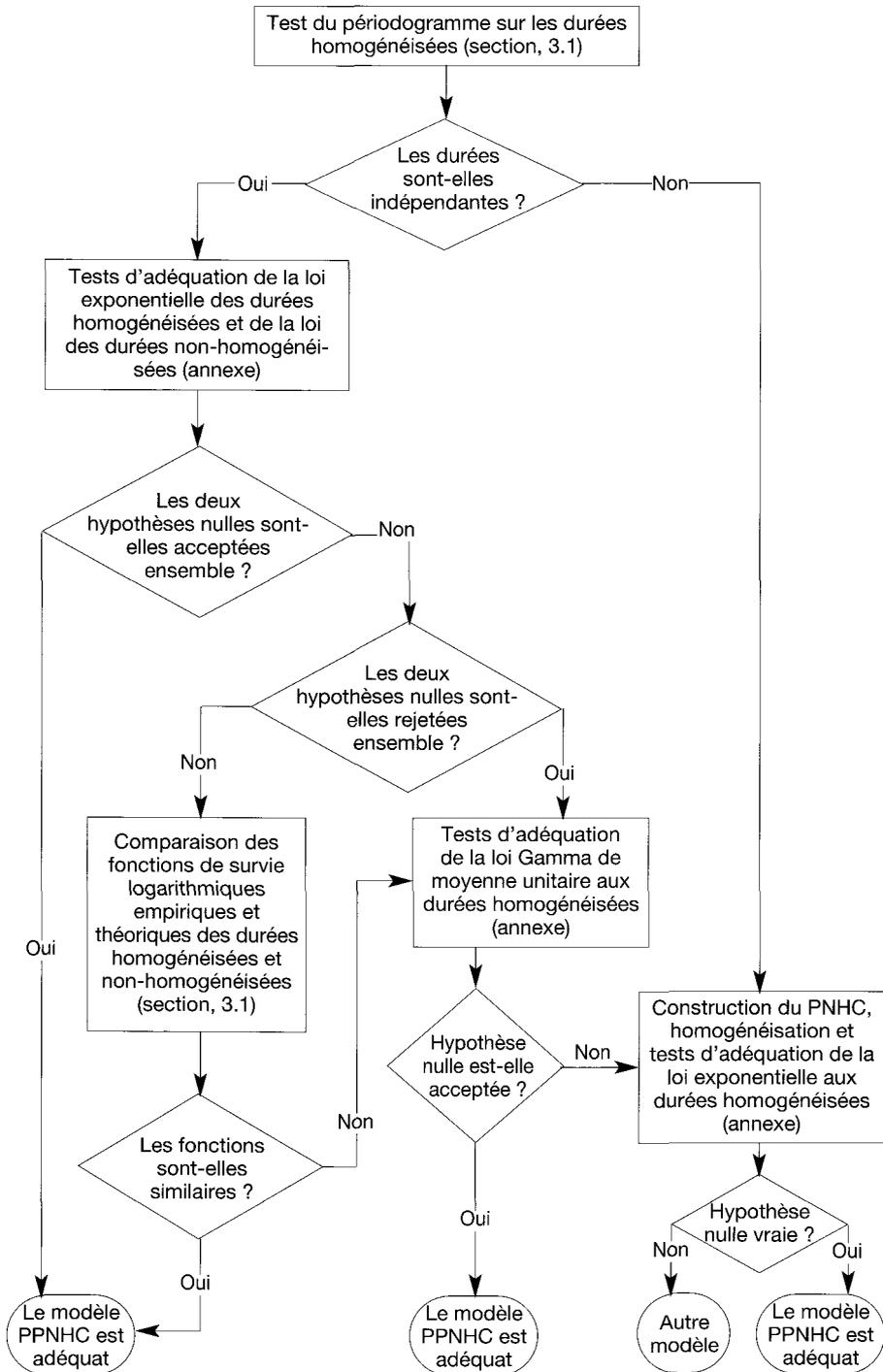


Figure 7 Discrimination entre les processus PPNH, PRNH et PPNHC (d'après ABI-ZEID, 1997).

la modélisation d'une série d'événements par un processus ponctuel. Après la préparation des données, nous proposons de calculer l'indice de dispersion du nombre d'événements. Selon la valeur de cet indice, nous pouvons effectuer une première discrimination entre les processus PPH, PBN et PB. Il faut noter que le calcul du coefficient de variation est également utile pour la discrimination entre les processus PPH et PPG, (KAVVAS, 1975). Dans le cas où les processus PPH, PBN et PB ne sont retenus par les tests statistiques, nous proposons de modéliser la série d'événements par un processus de renouvellement en considérant que la loi des intervalles n'est pas exponentielle. Si le processus PRH n'est pas retenu, alors on peut utiliser la procédure de sélection entre deux ou plusieurs processus ponctuels proposée par SMITH et KARR (1985), section 3.4.2. Cependant cette procédure ne constitue pas un moyen exhaustif pour la sélection du processus adéquat.

6 – CONCLUSIONS

Cette étude a été menée dans le but d'identifier les tests statistiques d'une série d'un seul ou de plusieurs types d'événements pour l'analyse du risque. Les modèles adoptés pour la modélisation de ce type de données sont les processus ponctuels temporels. Ainsi, nous avons réalisé une revue exhaustive de la littérature sur les tests statistiques appliqués aux séries d'événements. Cette étude a permis de montrer que la grande majorité des travaux de recherche utilisent des processus ponctuels basés sur les hypothèses de stationnarité et d'indépendance pour modéliser les séries d'événements (processus de Poisson homogène, processus de renouvellement homogène). Cependant, les données disponibles dans la réalité ne sont pas toujours conformes à ces deux hypothèses.

Nous avons identifié plusieurs méthodes pour l'analyse des séries d'un seul type d'événements, en partant de l'étape exploratoire qui est l'analyse graphique des données jusqu'aux tests statistiques basés sur des théories assez complexes. Les tests identifiés visent essentiellement l'analyse de tendance dans la série de données, l'indépendance et l'adéquation de la distribution statistique des intervalles et finalement la discrimination entre deux ou plusieurs processus ponctuels. Afin de faciliter leur utilisation, nous avons proposé une méthodologie qui nous permet de choisir le processus ponctuel adéquat en utilisant les tests statistiques approprié à ce processus (*figure 8*). Cependant, cette représentation ne constitue pas un moyen standard pour la sélection d'un processus ponctuel et des tests qui le valident.

Nous avons examiné la possibilité d'appliquer ces tests statistiques à une série de deux ou de plusieurs types d'événements. Dans le cas où l'événement est défini par son temps d'occurrence seulement, i.e. sa durée est négligeable, on peut appliquer les tests identifiés dans section 3 pour chaque série d'événements. L'étude des propriétés du processus global a montré que la considération de la série superposée comme une série d'un seul type d'événements reste liée à la nature des p processus subsidiaires. Ainsi, on démontre que la superposition de p processus ponctuels de Poisson est un processus de Pois-

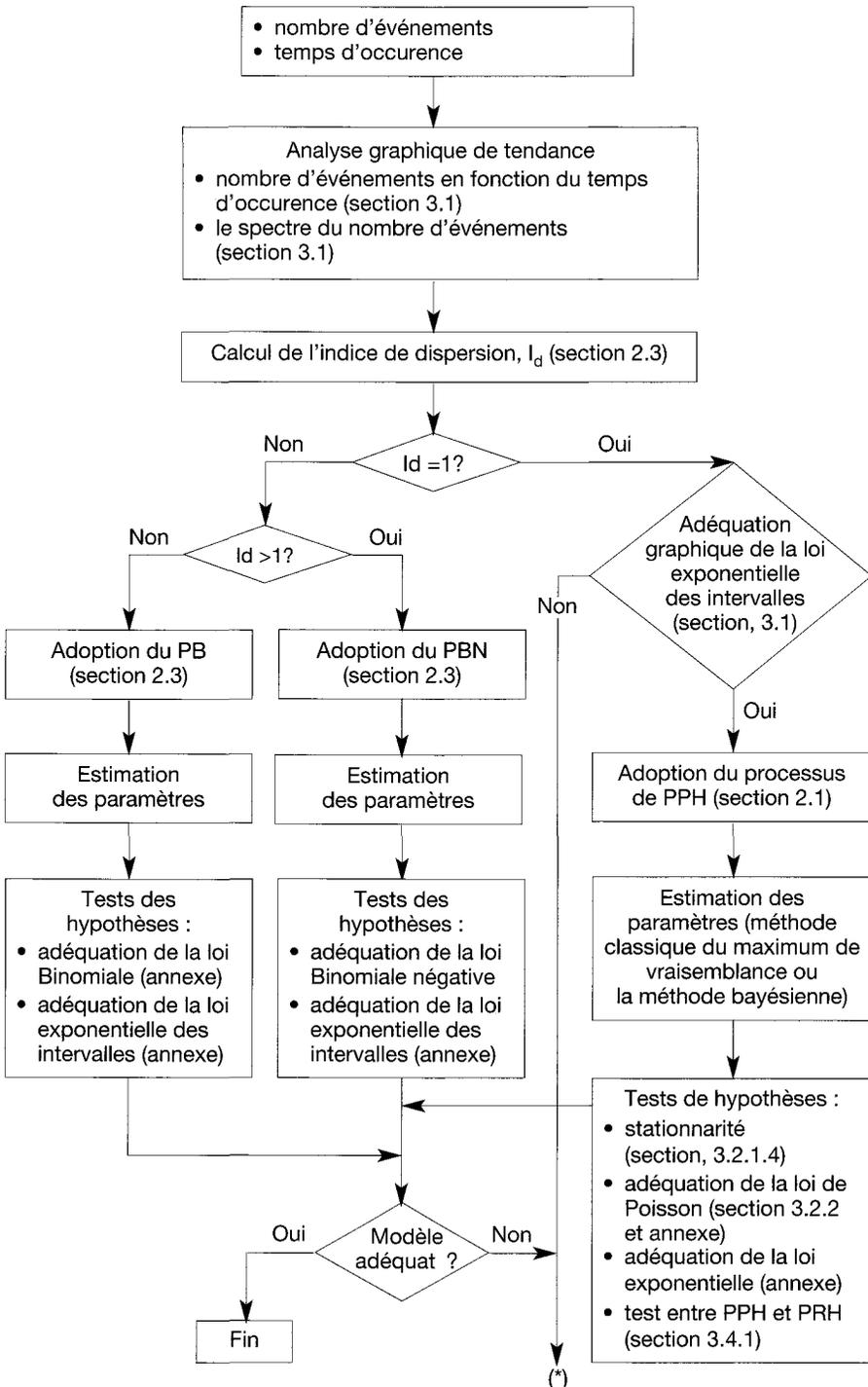


Figure 8 Proposition des étapes de la modélisation d'une série d'événements par un processus ponctuel stationnaire.

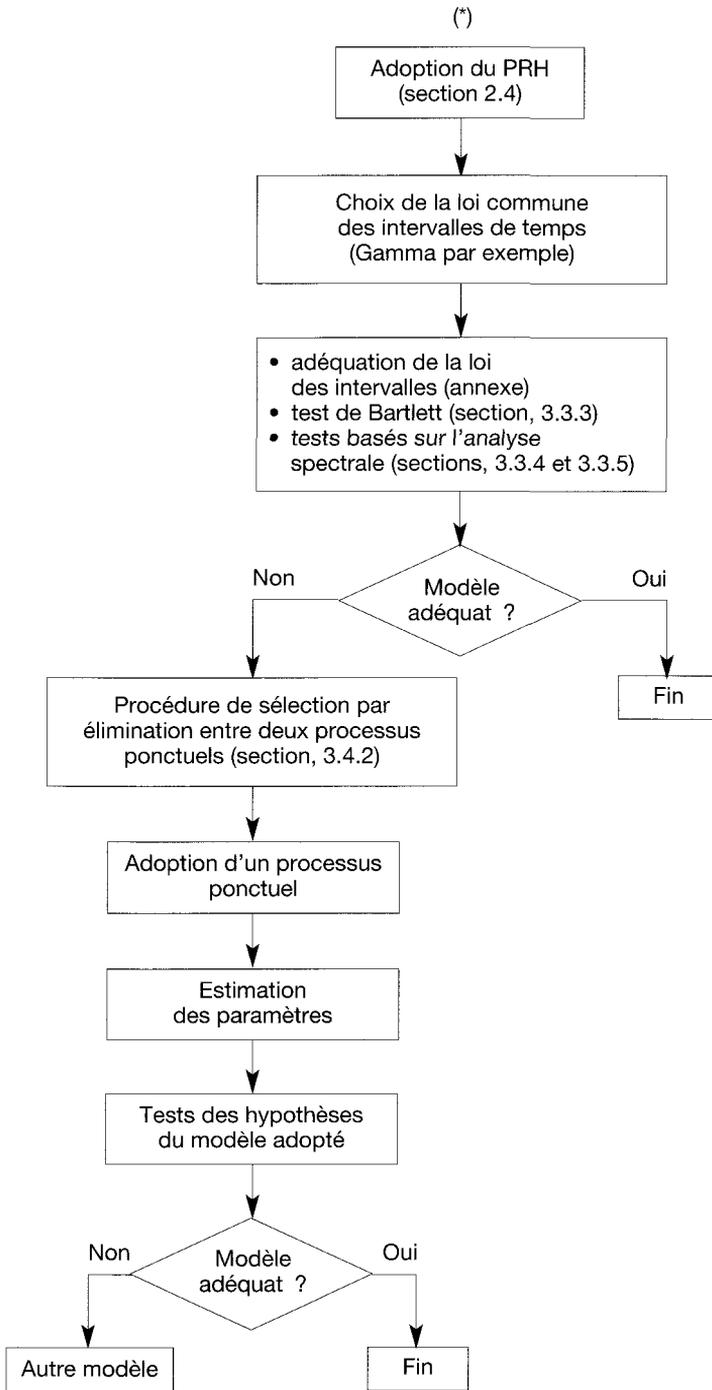


Figure 8 (Suite).

son quand le paramètre p tend vers l'infini. Il est à noter que la superposition des processus ponctuels utilisés dans la littérature est basée principalement sur le processus de Poisson homogène et le processus de renouvellement homogène. Dans le cas où les événements sont définis par leurs temps d'occurrence et de séjour (durées), les procédures d'analyse seront appliquées aux durées seulement. C'est le cas du processus alterné défini dans la section 4.2.

ANNEXE

A. Tests d'ajustement indépendants de la distribution

« *distribution-free tests* »

Ces tests sont appliqués dans le but de trouver la distribution adéquate pour les intervalles de temps séparant deux événements consécutifs. Cependant, ils peuvent être utilisés dans d'autres approches concernant l'analyse de séries d'événements. On suppose que l'on dispose d'une série de variables aléatoires représentée par une fonction de répartition inconnue $F_X(x)$. Soit $F_0(x)$ une distribution spécifiée telle que les hypothèses suivantes sont respectées :

$$H_0 : F_X = F_0 ;$$

$$H_1 : F_X \neq F_0, \text{ (test bilatéral).}$$

où la variable aléatoire X peut représenter le nombre d'événements, les intervalles de temps entre les événements, etc.

Les statistiques utilisées sont basées sur la distance entre $F_0(x)$ et la distribution empirique $F_n(x)$, où : $F_n(x) = \frac{\text{nombre de } x_i \leq x}{n}$. Il existe dans la littérature plusieurs tests d'ajustement indépendant de la distribution, on cite les tests de Kolmogorov-Smirnov (COX et LEWIS, 1966), Cramér-von Mises (COX et LEWIS, 1966), ces deux tests sont décrits en appendice, Stépens (GOMES et al., 1975), et Sherman (GOMES et al., 1975). Il est important de souligner que les statistiques de ces tests ne dépendent pas de l'égalité $F_0(x) = F_X(x)$ et qu'elles n'exigent pas la normalité des données.

A.1 Tests bilatéral et unilatéral de Kolmogorov-Smirnov

On rappelle les hypothèses du test :

$$H_0 : F_X = F_0 ;$$

$$H_1 : F_X \neq F_0, \text{ (test bilatéral).}$$

On définit les statistiques du test de Kolmogorov-Smirnov pour les cas suivants :

- test unilatéral à droite :

$$D_n^+ = \sup_{-\infty < x < +\infty} \{F_n(x) - F_0(x)\} = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - F_0(x_{(i)}) \right\} \quad (A1)$$

- test unilatéral à gauche :

$$D_n^- = \sup_{-\infty < x < +\infty} \{F_0(x) - F_n(x)\} = \max_{1 \leq i \leq n} \left\{ F_0(x_{(i)}) - \frac{i-1}{n} \right\} \quad (\text{A2})$$

- test bilatéral :

$$D_n = \sup_{-\infty < x < +\infty} |F_n(x) - F_0(x)| = \max_{1 \leq i \leq n} (D_n^+, D_n^-) \quad (\text{A3})$$

Il faut noter que les distributions asymptotiques ne suivent pas une loi normale. Le test bilatéral rejette l'hypothèse H_0 à un niveau de signification α ($\text{Prob}(D_n > d_{n,\alpha}) = \alpha$) si la valeur de D_n observée est supérieure à $d_{n,\alpha}$. Ce test est consistant, mais il est biaisé (MASSEY, 1950), cependant les tests unilatéraux sont consistants et non biaisés (COX et LEWIS, 1966).

Pour une distribution empirique donnée, le test basé sur la statistique D_n acceptera H_0 à un niveau de signification α , $\forall x$, si :

$$F_n(x) - d_{n,\alpha} \leq F_0(x) \leq F_n(x) + d_{n,\alpha} \quad (\text{A4})$$

A.2 Test de Cramér-von Mises

La statistique du test est définie par la quantité suivante :

$$W_n^2 = \int_{-\infty}^{+\infty} \{F_n(x) - F_0(x)\}^2 dF_0(x) \quad (\text{A5})$$

Cette quantité mesure l'erreur quadratique moyenne ou la distance entre $F_n(x)$ et $F_0(x)$. La distribution asymptotique est donnée par :

$$W_n^2 = \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^n \left\{ F_0(x_{(i)}) - \frac{2i-1}{2n} \right\}^2 \quad (\text{A6})$$

Cette distribution, qui ne suit pas une loi normale, a été obtenue par Smirnov (1949).

Pour les échantillons de petite taille, W_n^2 ne converge pas toujours rapidement. Cependant, pour $n > 20$, la distribution asymptotique peut être utilisée.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Abi-Zeid I., 1997. La modélisation stochastique des étages et de leurs durées en vue de l'analyse du risque. *Thèse de Doctorat*, INRS-Eau, Sainte-Foy, Canada, 278 p.
- Abi-Zeid I., Bobée B., 1999. La modélisation stochastique des étages : une revue bibliographique. *Revue des Sciences de l'Eau*, 5, 459-484.
- Ambartzumian R.V., 1965. Two inverse problems concerning the superposition of recurrent point processes, *Journal of Applied Probability*, 2, 449-454.
- Barnard G.A., 1953. Time intervals between accidents, *Biometrika*, 40, 212-213.
- Bartlett M.S., 1937. Properties of sufficiency and statistical tests, *Proc. Roy. Soc. A*, 160, 268-282.
- Bartlett M.S., 1963. The spectral analysis of point processes, *Journal of the Royal Statistical Society, Serie B*, 25, 264-296.

- Bernier J., 1967. Sur la théorie de renouvellement et son application en hydrologie. *Notes de Electricité de France*, HYD. 67, N. 10, 22 p.
- Brillinger D.R., 1994. Trend analysis: Time series and point process problems. *Environmetrics*, 5, 1-19.
- Chandramohan J., Liang L.K., 1985. Bernoulli, multinomial and Markov chain thinning of some point processes and some results about the superposition of dependent renewal processes. *Journal of Applied Probability*, 22, 828-835.
- Chebchoub A., Bargaoui Z., Abi-Zeid I., Bobée B., 1999. Modélisation de l'occurrence des étiages à partir des débits moyens journaliers à l'aide des processus ponctuels. IAHS Publ. 255, 155-163.
- Cinlar E., 1968. On the superposition of m-dimensional point processes. *Journal of Applied Probability*, 5, 169-176.
- Cinlar E., Agnew R.A., 1968. On the superposition of point processes. *Journal of the Royal Statistical Society, Serie, B.30*, 576-581.
- Cerventes J.E., Kavvas M.L., Delleur J.W., 1983. A cluster Model for flood analysis. *Water Resources Research*, 19 (1), 209-224.
- Cox D.R., 1955. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society, Serie, B. 17*, 129-164.
- Cox D.R., Isham V., 1980. Point process, Chapman et Hall, 188 p.
- Cox D.R., Lewis P.A.W., 1966. The statistical analysis of series of events, Methuen and Co LTD. 285 p.
- Cox D.R., Smith W.L., 1954. On the superposition of renewal processes. *Biometrika*, 41, 91-99.
- Cramer H., 1946. Mathematical methods of statistics, Princeton University Press.
- Cunnane C., 1979. A Note on the Poisson assumption partial duration series models. *Water Resources Research*, 15 (2), 489-494.
- Diggle P.J., 1983. Statistical analysis of spatial point patterns, Academic Press, 148 p.
- Faucher D., Ouarda T.B.M.J., Bobée B., 1997. Revue bibliographique des tests de stationnarité, Rapport de Recherche R-499, *Chaire en Hydrologie Statistique*, INRS-Eau, 65 p.
- Ghorbanzadeh D., Lounes R., 1996. Deux procédures de rupture pour des observations poissonniennes groupés, *Revue de Statistique Appliquée*, XLIV(3), 47-61.
- Girault M., 1965. Processus aléatoires, Edition Dunod, 150 p.
- Gomes M.I, Baroso H.M., Amaral M.A., 1975. Étude expérimentale de tests d'ajustements, *Revue de Statistique Appliquée*, XXIII (2), 5-18.
- Ito Y., 1977. Superposition of distinguishable point processes, *Journal of Applied Probability*, 14, 200-204.
- Karr A.F., 1986. Point processes and their statistical inference, Marcel Dekker Inc, 490 p.
- Kavvas M.L., 1975. Point stochastic analysis of daily rainfall with application to Indiana, Thèse de Doctorat à l'Université de Purdue, Indiana Napolis, USA, 185 p.
- Khaliq M.N., Cunnane C., 1996. Modelling point rainfall occurrences with the modified Bartlett-lewis rectangular pulses model. 180, 109-138.
- Lam C.T., Lehoczy J.P., 1991. Superposition of renewal processes, *Advances in Applied Probability*, 23, 64-85.
- Lang M., 1999. Theoretical discussion and Monte-Carlo simulations for a Negative Binomial process paradox, *Stochastic Environmental Research and Risk Assessment*, 13, 183-200.
- Lang M., Ouarda T.B.M.J., Bobée B., 1999. Towards operational guidelines for over-threshold modeling, *Journal of Hydrology*, 225, 103-117.
- Levy P., 1954. Processus semi-markoviens, Proc. Int. Cong. Math. (Amsterdam), 3, 416-426.
- Lawrance A.J., 1973. Dependency of intervals between events in superposition processes. *Journal of the Royal Statistical Society, Serie, B. 35*, 306-315.
- Lewis P.A.W., 1964. A branching Poisson process model for the analysis of computer failure patterns. *Journal of the Royal Statistical Society, Serie, B. 26*, 398-456.
- Massey F.J., 1950. A note on the power of a non-parametric test. *Ann. Math. Statist.*, 20, 440-442.

- Moran P.A.P., 1951. The random division of an interval, Part II. *Journal of the Royal Statistical Society, Serie, B.* 13, 147-150.
- North M., 1981. Processus hydrologique intermittents, Étude de modélisations stochastiques et applications. Thèse de Doctorat, No. 419, École Polytechnique Fédérale de Lausanne, Lausanne, Suisse, 194 p.
- Ondo J.C, Ouarda T.B.M.J., Bobée B., 1997. Revue bibliographique des tests d'homogénéité et d'indépendance. Rapport de recherche N° R-500, *Chaire en Hydrologie Statistique*, INRS-Eau, Sainte-Foy, Canada, 78 p.
- Onof C., Wheater H.S., 1993. Modelling of British rainfall using random parameter Bartlett-Lewis rectangular pulse model. *Journal of Hydrology.* 149, 67-95.
- Rao C.R., Chakravarti I.M., 1956. Some small sample tests of significance for a Poisson distribution, *Biometrika*, 12, 264-282.
- Ripley B.D., 1976. On stationarity and superposition of point processes. *The Annals of Probability*, 4(6) 999-1005.
- Smirnov N.V., 1949. On the Cramer-von Mises criterion, *Uspehi, Matem. Nauk.* (NS), 4, 196-197.
- Smith W.L., 1955. Regenerative stochastic processes, *Proc. Roy. Soc. Serie, A.* 232, 6-31.
- Smith J.A., 1980. Point process models of rainfall. Thèse de Doctorat, Université Johns Hopkins, Baltimore, Maryland, USA, 181 p.
- Smith J.A., Karr A.F., 1985. Statistical inference for point process models of rainfall, *Water Resources Research*, 21(1), 73-79.
- Snyder D.L., 1975. Random point processes, John Wiley & sons, Inc, 485 p.
- Sukhatme P.V., 1938. On the distribution of χ_2 in samples of the Poisson series. *Journal of the Royal Statistical Society*, 5(1), 75-79.
- Thompson W.A., 1988. Point process models with application to safety and reliability, Chapman et Hall, 146 p.
- Wald A., Wolfowitz J., 1943. An exact test of randomness in the nonparametric case based on serial correlation. *The Annals of Mathematical Statistics*, 14, 378-388.
- Waymire E., Gupta V.K., 1981. The mathematical structure of rainfall representations: 2, A review of the theory of point processes. *Water Resources Research*, 17(5), 1273-1285.