



HAL
open science

Processus ponctuels pour l'étude de l'apparition de fuites sur un réseau de distribution d'eau

Nicolas Dante, Bérengère Sixta-Dumoulin, Radu S. Stoica

► **To cite this version:**

Nicolas Dante, Bérengère Sixta-Dumoulin, Radu S. Stoica. Processus ponctuels pour l'étude de l'apparition de fuites sur un réseau de distribution d'eau. Journées de statistique 2019, Jun 2019, Nancy, France. hal-02187889v2

HAL Id: hal-02187889

<https://hal.science/hal-02187889v2>

Submitted on 13 Aug 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PROCESSUS PONCTUELS POUR L'ÉTUDE DE L'APPARITION DE FUITES SUR UN RÉSEAU DE DISTRIBUTION D'EAU.

Nicolas DANTE ¹ & Bérengère SIXTA DUMOULIN ² & Radu S. STOICA ¹

¹ *Université de Lorraine, CNRS, IECL, F-54000 Nancy, France*

[nicolas.dante@univ-lorraine.fr,radu-stefan.stoica@univ-lorraine.fr]

² *SEDIF, 120 Bd Saint Germain - 75006 Paris, France [b.sixta-dumoulin@sedif.com]*

Résumé. Cet article présente les processus ponctuels sur réseaux linéiques pour la modélisation spatiale des fuites sur un réseau de distribution d'eau. Cette modélisation se fait sachant la position des capteurs de détection de fuites sur le réseau. Des statistiques exploratoires permettant la comparaison des observations avec la réalisation d'un processus poissonien stationnaire sont présentées. Ensuite un processus ponctuel de Poisson inhomogène est testé sur des données réelles. A la fin, des conclusions et des perspectives sont formulées.

Mots-clés. processus ponctuels sur réseau, détection de fuites, analyse exploratoire, inférence bayésienne de type ABC.

Abstract. This paper presents point processes on linear networks for the spatial modeling of water leaks on a water distribution network. The water leaks modeling is done, conditionally to the position of water leaks detectors on the network. Summary statistics that allow the comparison with a stationary Poissonian reference are then presented. Then, a non-homogeneous Poisson point process is tested on real data. Conclusions and perspectives are finally depicted.

Keywords. point processes on linear networks, water leaks detection, summary statistics, ABC Bayesian inference.

1 Présentation des Données

Les données fournies par le Syndicat des Eaux d'Île-de-France consistent en un réseau de canalisations sur lequel sont localisées des fuites et des capteurs. Chaque capteur et chaque fuite sont accompagnées d'un grand nombre d'informations, respectivement : type de capteur, type de fuite, date de la fuite, date du déclenchement du capteur, pression, par exemple.

La Figure 1 montre la répartition spatiale des fuites et des capteurs sur un réseau dans une petite ville en France. Voilà quelques questions que l'on se pose :

- est-ce que les fuites sont réparties uniformément sur le réseau ?
- est-ce que la distribution des fuites est complètement indépendante de la position des capteurs ?

- si un telle dépendance peut être mise en évidence, est-ce que l'on pourrait quantifier à l'aide d'un modèle statistique?



FIGURE 1 – *Les Fuites (en noir) et capteurs (en rouge) sur les tronçons du réseau de canalisations*

Cet article apporte des éléments de réponses à ces questions en utilisant les processus ponctuels sur réseaux [1, 2].

2 Processus Ponctuels sur un réseau linéique

Un segment linéaire d'extrémités u et v est donné par :

$$l = [u, v] = \{tu + (1 - t)v : 0 \leq t \leq 1\}$$

La longueur d'un segment est : $|l| = \|v - u\|$ où $\|\cdot\|$ est la norme euclidienne.

Alors un réseau linéaire est défini comme une union finie de segments linéiques :

$$L = \bigcup_{i=1}^n l_i$$

La théorie des processus ponctuels peut être étendue aux réseaux linéiques. Il faut pour cela introduire une nouvelle distance ("plus court chemin") et une nouvelle intégrale sur réseau. Cependant la définition d'un processus ponctuel sur un réseau linéique est analogue à celle sur \mathbb{R}^d .

Considérons maintenant un processus ponctuel simple, donc qui ne contient pas deux fois le même point presque sûrement, sur un réseau linéique.

Soit X un processus de Poisson sur L . Le processus a une fonction d'intensité $\lambda(\cdot)$ qui si pour chaque segment linéique $B \subset L$:

$$\mathbb{E}[n(X \cap B)] = \int_B \lambda(u) du$$

où $n(X \cap B)$ est le nombre de points dans $X \cap B$.

Pour les moments de second ordre, le processus X a une intensité λ_2 si pour deux segments disjoints A et B on a :

$$\mathbb{E}[n(X \cap A)n(X \cap B)] = \int_A \int_B \lambda_2(u, v) dudv.$$

La fonction K d'un processus ponctuel stationnaire sur \mathbb{R}^d est proportionnelle au nombre moyen de voisins d'un point du processus. Par analogie, on construit la fonction K pour un processus ponctuel sur réseau. Dans ce cas, son estimateur corrigé qui utilise la distance du chemin le plus court sur le réseau est :

$$\hat{K}_{net}(r) = \frac{|L|}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{1[d_L(x_i, x_j) \leq r]}{m(x_i, d_L(x_i, x_j))}$$

où $d_L(\cdot, \cdot)$ est la distance entre deux points sur le réseau et $m(x, d)$ est le nombre de points du réseau exactement à distance d de x .

Il est tout à fait possible de construire la fonction de corrélation par paire de points, qui est proportionnelles à la dérivée du premier ordre de la fonction K . Son estimateur se construit d'une manière analogue au cas précédent, en utilisant la distance du chemin le plus court, sur un réseau. Pour un processus ponctuel homogène on a :

$$\hat{g}(r) = \frac{|L|}{n(n-1)} \sum_{i=1}^n \sum_{i \neq j} \frac{k(d_L(x_i, x_j) - r)}{m(x_i, d_L(x_i, x_j))}$$

alors que pour le cas inhomogène, nous obtenons

$$\hat{g}(r) = \frac{1}{\sum_i \frac{1}{\hat{\lambda}(x_i)}} \sum_{i=1}^n \sum_{i \neq j} \frac{k(d_L(x_i, x_j) - r)}{\hat{\lambda}(x_i)\hat{\lambda}(x_j)m(x_i, d_L(x_i, x_j))},$$

où $k(\cdot)$ est un noyau sur \mathbb{R} .

3 Application

Sous l'hypothèse poissonnienne, il est possible d'obtenir des expressions analytiques simples pour $K_{net}(r)$ et $g(r)$, respectivement. Ainsi, il est possible de construire des outils

d'exploration basés sur ces statistiques descriptives pour rejeter l'hypothèse nulle, ici le caractère poissonien des observations.

La Figure 3 montre, pour un fragment du réseau de distribution d'eau dans la ville considérée, les valeurs obtenues pour les estimateurs de K et ρ , respectivement. Pour plus de robustesse, un test d'enveloppes a été également effectué. Le niveau du test pour chaque rayon r a été 0.02. Ainsi, les résultats obtenus indiquent que les données ne puissent provenir d'un processus de Poisson stationnaire est rejetée. La réponse à notre première question serait donc, que ces données suggèrent que la répartition spatiale des fuites d'eau n'est pas uniforme sur le réseau.

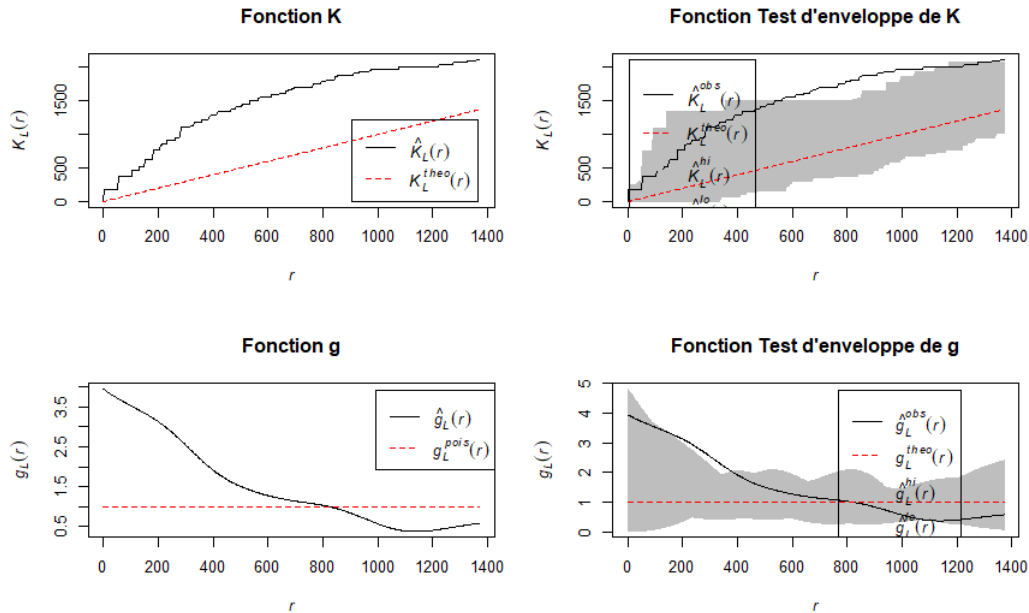


FIGURE 2 – Fonction K et de corrélation par paire des fuites sur un fragment de réseau de distribution d'eau : estimation ponctuelle et tests d'enveloppes. En rouge : les valeurs des fonctions sur l'hypothèse d'un processus de Poisson stationnaire.

Il est possible de construire des fonctions K et g pour vérifier s'il y a une relation de dépendance entre deux configuration de points distinctes [1, 2]. Nous utilisons ces fonctions pour vérifier s'il y a une relation entre la répartition spatiale des fuites et celle des capteurs. Les résultats sont montrés dans la Figure 3. Les estimateurs ponctuels de K et ρ montrent des différences entre ceux-là, et leur valeurs théoriques respectives. Cependant, le test d'enveloppes ne rejette pas l'indépendance entre ces deux distributions.

Le résultat de ce dernier test n'est pas contradictoire, bien au contraire. Il indique qu'un modèle spatial pour la répartition de fuites sachant la position du capteur devrait prendre en compte aussi toutes les informations disponibles relatives au temps d'apparition d'une fuite et de sa détection par un capteur.

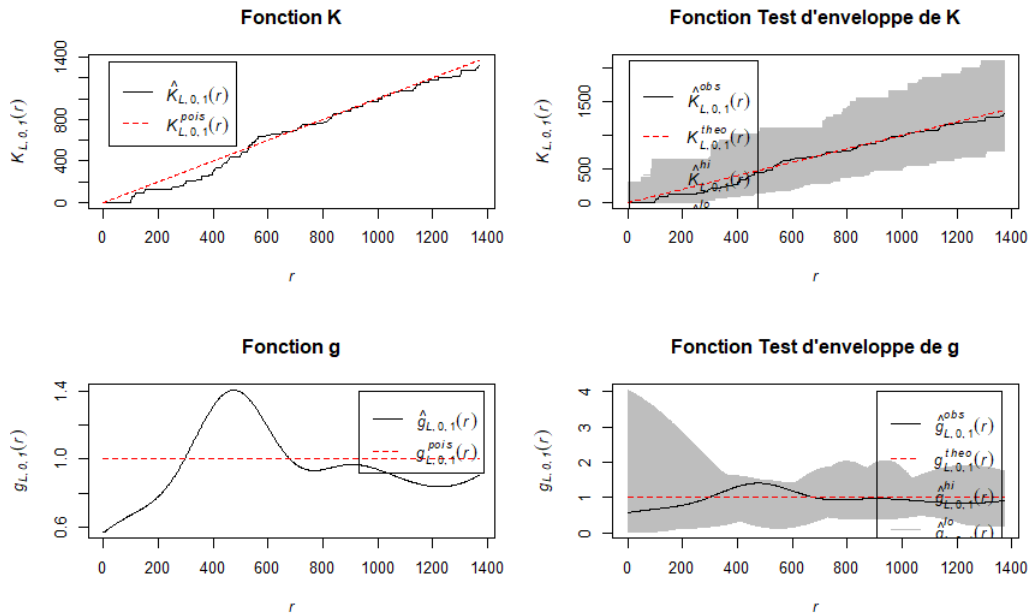


FIGURE 3 – *Fonction croisées (fuites et capteurs) K et de corrélation par paire sur un fragment de réseau de distribution d'eau. estimation ponctuelle et tests d'enveloppes. En rouge : les valeurs des fonctions sur l'hypothèse d'indépendance entre fuites et capteurs.*

Cette analyse exploratoire n'indique pas si la non-uniformité spatiale observée pour les fuites sur le réseau peut être expliquée par certaines co-variables, par une interaction entre les fuites, ou bien les deux aspects pris ensemble. Un modèle statistique pour la répartition spatiale des fuites devrait prendre en compte ses aspects. La construction d'un tel modèle est notre objectif mais ceci n'est pas une tâche triviale [4].

Face à ce challenge, nous avons décidé procéder par étapes. D'autant plus, que l'estimation des paramètres des processus ponctuels avec interactions reste encore un problème ouvert [3].

Dans ce contexte, nous proposons comme tout premier modèle, un processus ponctuel de Poisson inhomogène. Son intensité est donnée par

$$\lambda(u) = \exp(\theta_0 + \theta_1 \psi(u, \mathbf{c})) \tag{1}$$

avec θ_0 et θ_1 les paramètres du modèle à estimer, \mathbf{c} l'ensemble des capteurs et leur informations associées qui est connu et la fonction ψ qui est une fonction indicatrice qui vaut 1 si un capteur s'active au maximum 7 jours après l'apparition d'une fuite située à moins de 500m de distance.

En observant les fuites et les capteurs dans une région de la ville, une procédure d'inférence bayésienne basée sur l'algorithme ABC Shadow [3] a été lancée. La Figure 3 montre l'estimation de la loi a posteriori du modèle (1). D'après ces résultats partiels,

les deux paramètres sont significativement non-nuls. Le *Mean Posterior Mode* du vecteur (θ_0, θ_1) obtenu a été $(-7.87, 5.5)$. A partir de ces valeurs nous avons estimé des erreurs asymptotiques de l'estimateur de maximum de vraisemblance Monte Carlo. La différence entre les vraies valeurs du modèles et celui du maximum de vraisemblance est donnée par $(1.29, 0.41)$. Alors que la différence entre le maximum de vraisemblance et son équivalent Monte Carlo est donnée par $(0.12, 0.013)$.

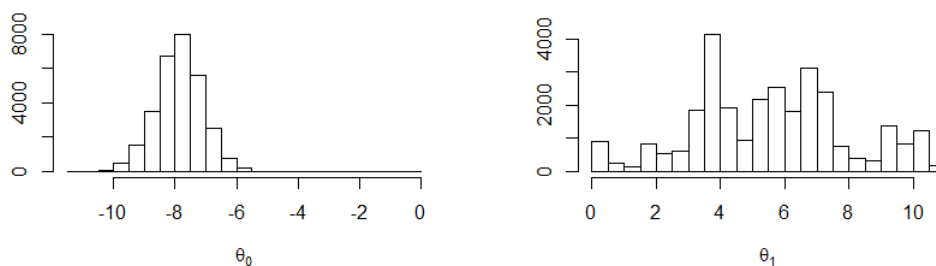


FIGURE 4 – Trajet de l’algorithme ABC Shadow , histogramme de la baseline et du paramètre du modèle.

4 Conclusion et perspectives

Les résultats obtenus sont encourageants. Le projet continue avec la construction d’un modèle permettant la prise en compte des différentes possibles interactions montrées par l’analyse exploratoire des données.

Références

- [1] Qi Wei Ang, Adrian Baddeley, and Gopalan Nair. Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scandinavian Journal of Statistics vol. 39*, pages 591–617, 2012.
- [2] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns Methodology and Applications with R*. CRC Press, 2016.
- [3] Radu S. Stoica, Anne Philippe, Pablo Gregori, and Jorge Mateu. Abc shadow algorithm : a tool for statistical analysis of spatial patterns. *Statistics and Computing*, 27(5) :1225–1238, 2017.
- [4] Marie-Colette van Lieshout. Nearest-neighbour markov point processes on graphs with euclidean edges. *Advanced in Applied Probability*, 50(4) :1275–1293, 2018.