



HAL
open science

Processus de Hawkes pour évaluer le risque de tavelure du pommier

Katarzyna Adamczyk-Chauvat, Madalina Deaconu

► **To cite this version:**

Katarzyna Adamczyk-Chauvat, Madalina Deaconu. Processus de Hawkes pour évaluer le risque de tavelure du pommier. 56es Journées de Statistique de la SFDS, SFDS, Jun 2025, Marseille, France. ⟨hal-05405094⟩

HAL Id: hal-05405094

<https://hal.inrae.fr/hal-05405094v1>

Submitted on 8 Dec 2025

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.



HAL Authorization

PROCESSUS DE HAWKES POUR ÉVALUER LE RISQUE DE TAVELURE DU POMMIER

Katarzyna Adamczyk-Chauvat¹ & Madalina Deaconu²

¹ *Université Paris-Saclay, INRAE, UR MaIAGE, France, Katarzyna.Adamczyk@inrae.fr*

² *Université de Lorraine, CNRS, Inria, IECL, France, Madalina.Deaconu@inria.fr*

Résumé. Le processus ponctuel de Hawkes modélise des phénomènes où l’occurrence d’un événement déclenche l’arrivée d’autres événements de même type. L’application présentée dans ce document concerne la tavelure, une maladie du pommier causée par un champignon pathogène *Venturia Inaequalis*. Le modèle de Hawkes proposé pour l’émission des spores du champignon tient compte de covariables météorologiques qui peuvent la déclencher. Les résultats montrent que l’estimateur de l’intensité conditionnelle du processus peut être considéré comme un indicateur pertinent du risque de contamination des pommiers.

Mots-clés. Processus de Hawkes, épidémiologie végétale, tavelure du pommier, covariables météorologiques.

Abstract. The Hawkes process models phenomena where the occurrence of an event triggers the arrival of other events of the same type. The application presented in this document concerns apple scab, a disease of apple trees caused by the pathogenic fungus *Venturia Inaequalis*. The Hawkes model proposed for the release of fungal spores takes into account meteorological covariates that may trigger it. The results show that the estimator of the conditional intensity of the process can be considered a relevant indicator of the risk of contamination of apple trees.

Keywords. Hawkes process, plant epidemiology, apple scab, meteorological covariates.

1 Introduction

Le processus ponctuel de Hawkes modélise des phénomènes où l’occurrence d’un événement déclenche l’arrivée d’autres événements de même type. Introduit par Alan G. Hawkes ([Hawkes, 1971], [Hawkes and Oakes, 1974]), le modèle a été appliqué dans divers domaines : la modélisation des séismes, l’analyse financière ou encore l’étude de réseaux sociaux. Nous présentons un exemple d’application du processus de Hawkes en épidémiologie végétale, pour décrire la propagation de la tavelure, une maladie du pommier. La maladie est causée par un champignon pathogène, *Venturia Inaequalis*, qui se développe sur les feuilles infectées, tombées aux pieds des arbres en automne. Arrivant à maturité au début de printemps, le champignon libère des spores qui se dispersent et infectent de nouvelles feuilles et fruits. La lutte contre la maladie repose sur un traitement fongique qui doit être synchronisé avec l’émission des spores. Le modèle de Hawkes a pour objectif de prédire les instants d’émission des spores en fonction de conditions météorologiques.

2 Processus de Hawkes

On considère un processus ponctuel simple $(N(t))_{t \geq 0}$ sur une demi-droite. L’histoire du processus en temps t , notée \mathcal{H}_t , est définie comme l’ensemble d’événements qui se sont produits avant l’instant

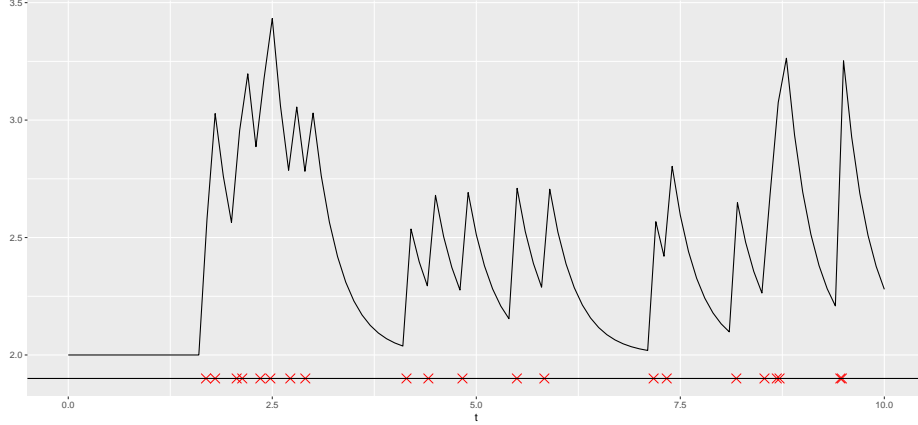


FIGURE 1 – Tracé de la fonction d’intensité conditionnelle $\lambda^*(t)$ pour le processus de Hawkes avec $\lambda(t) \equiv 2$ et $\mu(t) = 0.6 \exp(-3t)$ sur l’intervalle $]0, 10]$. Les croix représentent une trajectoire simulée du processus.

t . Le processus est caractérisé par la fonction d’intensité conditionnelle :

$$\lambda^*(t) = \lim_{dt \rightarrow 0^+} \frac{\mathbb{E}(N(t+dt) - N(t) | \mathcal{H}_t)}{dt} \quad (1)$$

La valeur de $\lambda^*(t)$ peut s’interpréter comme un risque instantané moyen de survenue d’un évènement à l’instant t , conditionnellement à l’histoire de processus avant t . On peut montrer que la fonction d’intensité conditionnelle (1) détermine la distribution du processus ponctuel de manière unique ([Daley and Vere-Jones, 2003], Proposition 7.2.IV).

Definition 2.1 (Processus de Hawkes). Le processus de Hawkes est caractérisé par la fonction d’intensité conditionnelle :

$$\lambda^*(t) = \lambda(t) + \sum_{t_i \in \mathcal{H}_t} \mu(t - t_i).$$

La fonction $\lambda(t) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ est appelé l’intensité de base et ne dépend pas de l’histoire du processus. La contribution des points du passé est donnée par la fonction d’excitation $\mu(t) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ telle que $\int_0^\infty \mu(s) ds < 1$.

La Figure 1 présente l’intensité conditionnelle de processus de Hawkes avec $\lambda(t) \equiv \lambda$ et le noyau exponentiel $\mu(t) = \alpha \exp(-\beta t)$, $0 < \alpha < \beta$. Dans cet exemple, les évènements les plus récents ont une plus grande influence sur le processus. L’arrivée d’un nouveau évènement augmente l’intensité $\lambda^*(t)$ de α et l’influence de cet évènement diminue avec le temps au taux β .

3 Inférence statistique

Diverses méthodes statistiques ont été développées pour inférer les paramètres de processus de Hawkes : ici nous nous focalisons sur l’estimation du maximum de vraisemblance. On considère une famille de processus ponctuels caractérisés par l’intensité $\lambda_\theta^*(t)$ ou $\theta \in \mathbb{R}^d$ est un vecteur des paramètres. Soit $n(T)$ le nombre d’évènements observés dans $]0, T]$ et soit $0 < t_1, \dots < t_{n(T)}$ une

trajectoire observée de processus. La fonction de log-vraisemblance du processus est donnée par :

$$\log L(\theta; \{t_i\}) = \sum_{i=1}^{n(T)} \log \lambda_{\theta}^*(t_i) - \Lambda(T) \quad (2)$$

où $\Lambda(T) = \int_0^T \lambda_{\theta}^*(t) dt$ est une fonction d'intensité cumulée. Ce résultat a été démontré dans [Daley and Vere-Jones, 2003]. Le comportement asymptotique de l'estimateur du maximum de vraisemblance $\hat{\theta}$ a été étudié dans [Ogata, 1978] pour les processus ponctuels stationnaires. La validation du modèle est basée sur le théorème (7.4.IV) de [Daley and Vere-Jones, 2003] :

Théorème 1 (Analyse résiduelle). Soit $\{t_1, t_2, \dots\}$ une trajectoire d'un processus ponctuel sur $(0, \infty)$ et soit $\Lambda(\cdot)$ la fonction d'intensité cumulée, supposée monotone, continue et telle que $\lim_{t \rightarrow \infty} \Lambda(t) = \infty$ p.s. La trajectoire transformée $\{t_1^*, t_2^*, \dots\} = \{\Lambda(t_1), \Lambda(t_2), \dots\}$ est une réalisation de processus de Poisson d'intensité 1 si et seulement si la trajectoire originale $\{t_1, t_2, \dots\}$ est une réalisation de processus ponctuel défini par $\Lambda(\cdot)$.

Il s'ensuit que pour valider le modèle, il suffit de vérifier si la trajectoire transformée $\{t_1^*, t_2^*, \dots\}$, restreinte à l'intervalle d'observations $]0, T]$, est une trajectoire de processus de Poisson unitaire. Pour ce faire, le test d'enveloppes globales basé sur des simulations de processus de Poisson unitaire peut être utilisé [Myllymäki et al., 2017].

4 Application aux données de la tavelure

Notre objectif est ici de prédire les instants d'émission de spores de *Venturia inaequalis*, le champignon à l'origine de la tavelure. L'émission des spores est déclenchée par les conditions météorologiques et continue pendant quelques heures qui suivent. L'expérience de monitoring des spores menée dans le verger expérimental de l'Institut National de Recherche en Horticulture à Skierniewice a fourni les données pour ajuster le modèle.

4.1 Capture des spores et données météorologiques

Un piège à spores a été placé sur le sol du verger expérimental recouvert de feuilles de la variété McIntosh fortement infectées par *V. Inaequalis* pendant la saison précédente [Jankowski and Masny, 2020]. L'appareil enregistre l'information de présence-absence de spores à une fréquence horaire pendant deux mois suivant la date de détection des premières spores. En parallèle, la station météorologique placée sur le site prenait des mesures de variables météorologiques qui pouvaient déclencher l'émission des spores. La température (temp) et l'humidité de l'air (hAir), la pression atmosphérique (pres), la hauteur des précipitations (prec), l'humidité foliaire (hFol), la vitesse et la direction du vent (vitV, dirV), la luminosité (lumi) et l'intensité lumineuse (intL) font partie des variables collectées dans l'expérience. Ces variables, mesurées initialement à une échelle temporelle plus fine, ont été agrégées par heure pour correspondre aux données de spores.

4.2 Modèle d'émission des spores

Le processus de Hawkes proposé pour modéliser l'émission des spores est donné par l'intensité conditionnelle :

$$\lambda^*(t) = \lambda(t) + \alpha \sum_{t_i \in \mathcal{H}_t} \exp(-\beta(t - t_i)) \quad (3)$$

$$\log(\lambda(t)) = \sum_{k=0}^K \gamma_k x_k(t)$$

où $(x_1(t), \dots, x_K(t))^T$ est un vecteur de K variables météorologiques mesurées au temps t et $x_0(t) \equiv 1$. L'intensité de base log-linéaire modélise l'impact des conditions climatiques sur le déclenchement de l'émission des spores et la partie auto-excitante tient compte de sa durée. Compte tenu des mesures expérimentales, l'intensité de base peut être approchée par la fonction constante par morceaux :

$$\log(\lambda(t)) = \sum_{k=0}^K \gamma_k \sum_{m=0}^{M-1} x_{k,m} \mathbb{1}_{(m, m+1]}(t)$$

où $]0, T] = \bigcup_{m=0}^{T-1}]m, m+1]$ et $x_{k,m}$ est la valeur de la k -ème variable sur l'intervalle de temps $]m, m+1]$. On assume par la suite que le processus a été observé jusqu'à la n -ème arrivée, on a alors $T = t_n$. La fonction de log-vraisemblance (2) pour le modèle (3) prend la forme :

$$\begin{aligned} \log L(\theta; \{t_i\}) &= \sum_{i=1}^n \log \left(\exp \left(\sum_{k=0}^K \gamma_k x_{k,m(i)} \right) + \alpha \sum_{j=1}^{i-1} \exp(-\beta(t_i - t_j)) \right) \\ &- \sum_{m=0}^{t_n-1} \exp \left(\sum_{k=0}^K \gamma_k x_{k,m} \right) - \frac{\alpha}{\beta} \sum_{i=1}^{n-1} (1 - \exp(-\beta(t_n - t_i))) \end{aligned}$$

où $x_{k,m(i)}$ est la valeur de la k -ème variable observée au temps t_i , $i = 1, \dots, n$, $k = 0, \dots, K$. L'estimateur du maximum de vraisemblance de θ a été calculé à partir de données collectées en 2011. Cette année on a relevé 215 événements d'émission des spores pendant 1416 heures de suivi. Les résultats d'estimation figurent dans la Table (1). Afin de comparer l'importance relative des covariables nous

cste	temp	hAir	pres	prec	hFol	vitV	dirV	lumi	intL		
$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$	$\hat{\gamma}_8$	$\hat{\gamma}_9$	$\hat{\alpha}$	$\hat{\beta}$
-4.26	0.40	1.06	-0.49	0.03	0.91	-0.11	0.06	0.59	-0.08	0.21	0.36

TABLE 1 – Estimation du maximum de vraisemblance pour le modèle 3. Les paramètres γ_i correspondent aux variables météorologiques, les paramètres α et β apparaissent dans la fonction d'excitation du modèle.

avons utilisé l'information sur l'ordre de leur introduction dans le modèle. Nous avons appliqué la stratégie ascendante, en ajoutant à chaque étape la variable qui améliore le plus ou détériore le moins l'AIC (Figure 2). La variable qui se place en tête du classement c'est l'humidité foliaire. Elle est suivie par la luminosité, l'humidité de l'air, la pression atmosphérique et la température : ces variables font cependant décroître l'AIC moins rapidement. L'ajout d'autres variables météorologiques fait augmenter le critère.

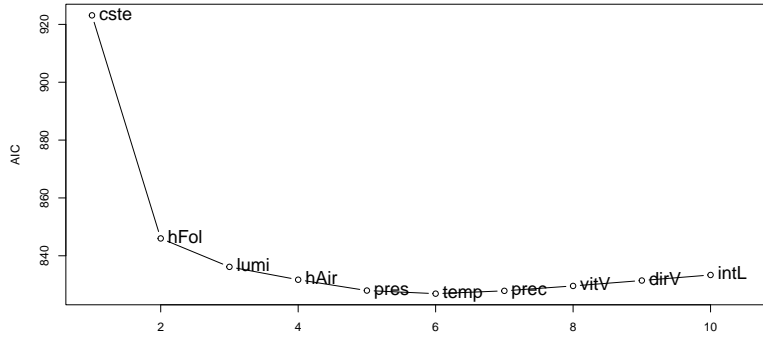


FIGURE 2 – L'évolution de l'AIC lors de l'ajout des variables météorologiques au modèle. A chaque étape on choisit la variable qui améliore le plus ou détériore le moins l'AIC, en commençant par le modèle avec l'intensité de base constante.

Le test d'enveloppes globales basé sur 500 simulations de processus de Poisson d'intensité 1 ne rejette pas l'hypothèse nulle de compatibilité entre modèle et données avec la p-valeur de 0.63. La Figure 3 présente l'interprétation graphique du résultat de test : le comptage associé à la trajectoire transformée $\{t_1^*, t_2^*, \dots, t_n^*\}$ reste à l'intérieur de la région délimitée par la procédure de test. Par conséquent la trajectoire originale t_1, \dots, t_n peut être modélisée comme une réalisation du modèle 3.

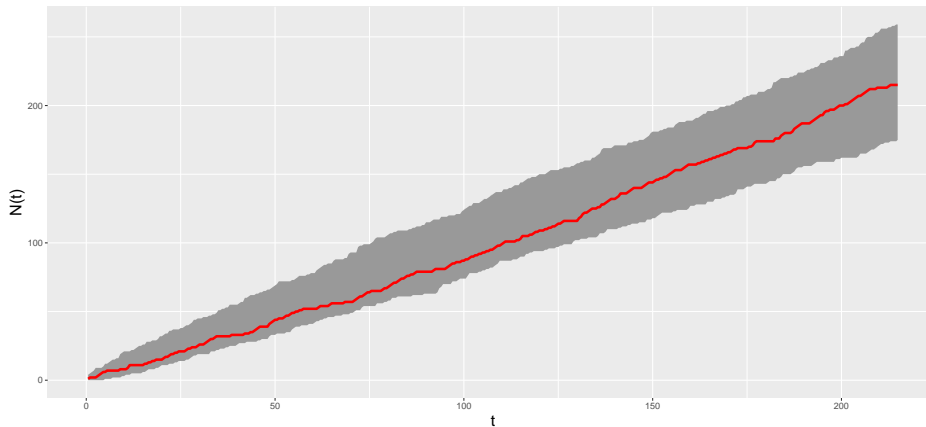


FIGURE 3 – Test d'enveloppes globales pour l'analyse résiduelle du modèle de Hawkes 3. La courbe rouge représente le comptage associé à la trajectoire de processus transformé, $\{\Lambda_{\hat{\theta}}(t_1), \Lambda_{\hat{\theta}}(t_2), \dots, \Lambda_{\hat{\theta}}(t_n)\}$. La courbe est située à l'intérieur de la région délimitée par le test, marquée en gris.

5 Conclusions

Ce document présente l'application de processus de Hawkes à la modélisation de l'émission des spores de *V. Inaequalis*. L'intensité conditionnelle du modèle proposé a été construite en utilisant la fonction d'excitation à décroissance exponentielle et l'intensité de base dépendant de covariables météorologiques. La méthode du maximum de vraisemblance a été utilisée pour estimer les paramètres du modèle. La comparaison de variables météorologiques a mis en évidence le rôle de variables liées à l'humidité dans le déclenchement de l'émission des spores. Le test d'enveloppes globales utilisant l'analyse résiduelle a été appliqué pour vérifier la qualité d'ajustement du modèle. Les résultats confirment que l'estimateur de l'intensité conditionnelle du modèle peut être utilisé comme un indicateur de risque de contamination par le champignon.

Références

- [Daley and Vere-Jones, 2003] Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Second Edition*. Springer, New York.
- [Hawkes, 1971] Hawkes, A. (1971). Spectra of some self-exciting and mutually-exciting point processes. *Biometrika*, 58 :83–90.
- [Hawkes and Oakes, 1974] Hawkes, A. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11 :493–503.
- [Jankowski and Masny, 2020] Jankowski, P. and Masny, S. (2020). Influence of moisture on maturation rate of the *Venturia inaequalis* (cooke) wint. ascospores in central poland. *Journal of Plant Diseases and Protection*, 127 :155–163.
- [Myllymäki et al., 2017] Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 79 :381–404.
- [Ogata, 1978] Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Ann. Inst. Statist. Math.*, 30 :243–261.
- [Pierrat, 2022] Pierrat, P. (2022). Processus de Hawkes : application en épidémiologie végétale. Technical report, INRAE - Département Mathématiques et Numérique, Inria - Équipe PASTA.