Procesos espacio-temporales de tipo self-exciting: Extensiones recientes y aplicaciones a datos criminológicos

Álvaro Briz Redón

Departamento de Estadística e Investigación Operativa
Universitat de València

alvaro.briz@uv.es

6 de mayo de 2025, Bogotá, D.C.
IV Congreso Colombiano de Estadística

Vniversitat de València

Contenidos

¿Qué es un proceso puntual?

- Es un proceso estocástico que describe la ocurrencia de puntos (*eventos*) sobre un espacio $W \subset \mathbb{R}^d$.
- Una realización de un proceso puntual es un conjunto de puntos $\{x_i\}_{i=1}^n$, donde $x_i \in W$, el espacio de estudio.
- Cada realización del proceso puntual se denomina *patrón puntual* sobre W.

Proceso puntual de Poisson homogéneo

Empezamos por el proceso puntual más simple: el proceso Poisson homogéneo. Presenta las siguientes características:

- Definido por un único parámetro, que representa que la *función de intensidad* es constante: $\lambda > 0$.
- Propiedades:
  - El número de puntos en una región $A \subset \mathbb{R}^d$ se distribuye según la siguiente distribución de Poisson:

$$N(A) \sim \text{Poisson}(\lambda|A|),$$

    donde $|A|$ representa el volumen de A.
  - Independencia: Si A_1 y A_2 son regiones disjuntas ($A_1 \cap A_2 = \varnothing$), entonces $N(A_1)$ y $N(A_2)$ son independientes.

Función de intensidad

Hemos dicho que la función de intensidad en el proceso Poisson homogéneo es constante e igual a λ. **¿Qué es exactamente la función de intensidad?**

Sobre un espacio $W \subset \mathbb{R}^d$, la función de intensidad se define como:

$$\lambda(x) = \lim_{|\mathrm{d}x| \to 0} \frac{\mathbb{E}[N(\mathrm{d}x)]}{|\mathrm{d}x|}, \forall x \in W$$

En palabras, esto significa que el número de puntos esperados en una región infinitesimal alrededor de $x \in W$. Por tanto, si $\lambda(x) = \lambda$:

$$E(N(A)) = \lambda |A|, \forall A \subset W$$

Proceso puntual de Poisson inhomogéneo

En general, es poco realista pensar que la intensidad de un proceso es constante sobre todo el espacio W. El proceso Poisson inhomogéneo mantiene las dos propiedades básicas del Poisson homogéneo, pero ahora:

$$E(N(A)) = \int_A \lambda(x)dx$$

Si asumimos que la intensidad $\lambda(x)$ es paramétrica y dependiente de un vector de parámetros, θ, la función de verosimilitud es (Daley and Vere-Jones, 2002):

$$L(\theta) = \exp\left(-\int_W \lambda(u)\,du\right)\prod_{i=1}^{n}\lambda(x_i)$$

Y la de log-verosimilitud:

$$\log L(\theta) = \sum_{i=1}^{n}\log\lambda(x_i) - \int_W \lambda(u)\,du$$

Procesos self-exciting: Motivación

- Los procesos puntuales de tipo self-exciting (auto-excitables) son herramientas fundamentales para modelar eventos **recurrentes**.
- Se usan ampliamente en criminología, sismología, economía y epidemiología.
- En estos modelos, se asume que los eventos pasados incrementan la probabilidad de eventos futuros **cercanos en espacio y tiempo**.
- Permiten capturar la **interacción espacio-temporal** del proceso subyacente.

Función de intensidad condicional

- Sea $\{(x_i, t_i)\}_{i=1}^n \subset W \times T \subset \mathbb{R}^2 \times \mathbb{R}$ una realización de un proceso puntual espacio-temporal, donde x_i indica posición espacial y t_i temporal.
- Asumimos que las localizaciones temporales están ordenadas y que no presentan empates: $t_1 < t_2 < \cdots < t_n$ (*orderly pattern*).
- La intensidad condicional define la probabilidad instantánea de un nuevo evento dadas las observaciones previas:

$$\lambda(x, t|\mathcal{H}_t) = \lim_{|dx|,|dt| \to 0} \frac{\mathbb{E}[N(dx \times dt)|\mathcal{H}_t]}{|dx \times dt|},$$

donde $\mathcal{H}_t = \{(x_i, t_i) : t_i < t\}$ representa *la historia del proceso* en tiempo t.

Modelo self-exciting estándar

El modelo self-exciting estándar presenta la siguiente función de intensidad condicional (con sus ligeras variantes):

$$\lambda(x, t|\mathcal{H}_t) = \lambda_0 + \sum_{j:t_j<t} \theta e^{-\frac{|t-t_j|}{\varphi_T}} e^{-\frac{\|x-x_j\|}{\varphi_S}}$$

Descripción de parámetros:

- λ_0: Parámetro de intensidad de fondo (homogénea). Puede reemplazarse por una intensidad inhomogénea, dependiente de covariables, por ejemplo: $\exp(\beta_1 X_1(x, t) + ... + \beta_p X_p(x, t))$.
- θ: Parámetro de productividad. Representa la contribución máxima de un evento previo a la intensidad condicional.
- φ_T, φ_S: Parámetros que controlan el nivel de decaída del efecto en el tiempo y el espacio.

Modelo self-exciting estándar: Limitaciones

El modelo self-exciting estándar presenta algunas limitaciones conocidas:

- Se asume que la influencia de cada evento previo en la intensidad es **separable en espacio y tiempo** (el término $\theta e^{-\frac{|t-t_j|}{\varphi_T}} e^{-\frac{\|x-x_j\|}{\varphi_S}}$ es separable en espacio y tiempo).
- Se asume que el **parámetro de productividad, θ, es constante**, no depende ni del evento ni de la localización del evento. **Recientemente, nos hemos centrado en esta cuestión** (Briz-Redón and Mateu, 2025).

Modelos self-exciting con productividad variable

Ya existen propuestas sobre modelos self-exciting con productividad variable. Por ejemplo:

- El modelo Epidemic-Type Aftershock Sequence (ETAS) ampliamente utilizado en sismología (Ogata, 1988), asume que la influencia de cada terremoto previamente observado en la intensidad condicional **depende de la magnitud de dicho terremoto**.

  En concreto, a partir de la ley de Omori (Utsu et al., 1995), la cual describe cómo decae la intensidad sísmica tras un terremoto previo, se emplea una intensidad (temporal) del tipo:

$$\lambda(t) = \lambda_0 + \sum_{j:t_j<t} \frac{K_0}{(t-t_j+c)^p} e^{\alpha(M_j-M_0)}$$

Modelos self-exciting con productividad variable

Ya existen propuestas sobre modelos self-exciting con productividad variable. Por ejemplo:

- En el ámbito de la criminología, se han propuesto modelos self-exciting con productividad variable **según el tipo de evento delictivo** (considerando patrones marcados). Por tanto, la productividad en este caso toma valores en un conjunto finito (Mohler, 2014; Mohler et al., 2018).
- En el ámbito de la epidemiología, algunos autores han vinculado el parámetro de productividad a **covariables estáticas o variables en el tiempo** para analizar la propagación del COVID-19 (Chiang et al., 2022), mientras que otros han propuesto que el parámetro de productividad está **inversamente relacionado con la intensidad condicional** debida a intervenciones humanas (Schoenberg et al., 2019).

Modelos self-exciting con productividad variable

Ya existen propuestas sobre modelos self-exciting con productividad variable. Por ejemplo:

- Además, se han empleado métodos **no paramétricos** para estimar el parámetro de productividad **en función de las características del evento**, como la magnitud del terremoto en el contexto de los estudios sismológicos (Fox et al., 2016; Marsan and Lengline, 2008; Schoenberg, 2022), o en función del tiempo (Lewis and Mohler, 2011; Phillips and Schoenberg, 2024).
- Más recientemente, Kwon et al. (2023) introdujeron un modelo espacio-temporal flexible de autoexcitación para el análisis de terremotos. Su modelo incluye un término adicional que permite que la productividad varíe en el espacio, estimada de forma **no paramétrica en función de la ubicación espacial del evento**.

Modelos self-exciting con productividad variable

Algunas posibles limitaciones de estas propuestas:

- Requieren **suposiciones fuertes** sobre el parámetro de productividad. Habitualmente, no disponemos de tanta certeza sobre el problema.
- No suelen tener en cuenta la **localización espacial** del evento. En caso de hacerlo, solo encontramos una propuesta no paramétrica (Kwon et al., 2023).
- No tienen en cuenta la localización **tanto en espacio como en tiempo**.

Modelo propuesto con productividad variable

Para tratar de mejorar con respecto de algunas de estas limitaciones, planteamos la siguiente propuesta de modelo self-exciting:

$$\lambda(x, t|\mathcal{H}_t) = \lambda_0 + \sum_{j:t_j<t} b(x)b(t)\theta e^{-\frac{|t-t_j|}{\varphi_T}} e^{-\frac{\|x-x_j\|}{\varphi_S}}$$

En este nuevo modelo tenemos:

- $b(x), b(t) \in [0, 1]$ actúan como *factores de corrección espacial y temporal*, respectivamente.
- Por tanto, el nuevo parámetro de productividad es $b(x)b(t)\theta$, el cual varía en espacio y tiempo.
- Los factores de corrección se definen mediante funciones spline:

$$\text{logit}(b(x)) = \sum_k \alpha_k f_k(x), \quad \text{logit}(b(t)) = \sum_\ell \beta_\ell f_\ell(t)$$

Modelo propuesto con productividad variable

Potenciales ventajas del modelo propuesto:

- Permite considerar que la productividad es **variable en espacio y tiempo** de forma **flexible**, **sin necesidad** de hacer **suposiciones fuertes** sobre el mecanismo que hace variar la productividad.
- El uso de funciones de tipo spline garantiza que la variación en espacio y tiempo de la productividad sea **suave**. Esto puede facilitar la **identificación** de patrones de variación.

Potenciales desventajas del modelo propuesto:

- Incrementa considerablemente la **complejidad computacional** del modelo.
- Requiere de la **elección de una base de splines**. Si escogemos una base muy densa (con muchas funciones implicadas), el coste computacional podría resultar excesivo.

Inferencia e implementación

- Se realiza inferencia Bayesiana, trabajando con la función de log-verosimilitud asociada a un proceso puntual, donde Θ representa el conjunto de parámetros implicados:

$$\ell(\Theta) = \sum_{i=1}^{n} \log(\lambda(x_i, t_i|\mathcal{H}_{t_i})) - \int_T \int_W \lambda(x, t|\mathcal{H}_t)\mathrm{d}x\mathrm{d}t$$

- Se imponen las siguientes distribuciones previas:

$$\begin{aligned}
\lambda_0 &\sim \mathit{Gamma}(10\frac{n}{|W||T|}, 10) \\
\alpha_k &\sim \mathit{Normal}(0, 10), \beta_\ell \sim \mathit{Normal}(0, 10) \\
\theta &\sim \mathit{Gamma}(\mathit{mean} = \widehat{\mathbb{E}(\theta_{std})}, \mathit{sd} = \widehat{\mathrm{SD}(\theta_{std})}) \\
\varphi_T &\sim \mathit{Gamma}(\mathit{mean} = \widehat{\mathbb{E}(\varphi_{T_{std}})}, \mathit{sd} = \widehat{\mathrm{SD}(\varphi_{T_{std}})}) \\
\varphi_S &\sim \mathit{Gamma}(\mathit{mean} = \widehat{\mathbb{E}(\varphi_{S_{std}})}, \mathit{sd} = \widehat{\mathrm{SD}(\varphi_{S_{std}})}),
\end{aligned} \tag{1}$$

donde se hacen uso de las estimaciones proporcionadas por el modelo self-exciting estándar para elaborar las previas sobre θ, φ_T y φ_S.

Inferencia e implementación

- El modelo se ha implementado en R con el paquete `nimble`, el cual permite realizar inferencia Bayesiana con métodos de tipo Markov chain Monte Carlo (MCMC).
- Se usa el **zeros-ones trick** (Ntzoufras, 2011) para trabajar con la función de log-verosimilitud.
- Requiere definir una cuadrícula espacio-temporal razonablemente fina sobre la que se aproxima la integral implicada en la función de log-verosimilitud.
- En el siguiente repositorio de GitHub puede encontrarse la implementación:

  `https://github.com/albrizre/SmoothProdSelfExc`

Estudio de simulación: Objetivos

- Evaluar la capacidad del modelo propuesto para capturar variaciones espacio-temporales en el parámetro de productividad.
- Comparar desempeño frente al modelo estándar.
- Comprobar si el proceso de inferencia es robusto.
- Comprobar si el modelo es consistente entre réplicas.

Configuración del estudio de simulación

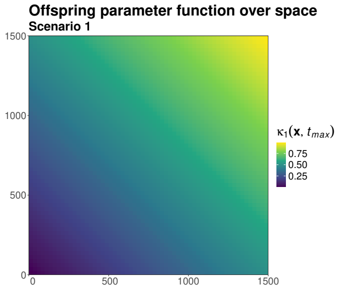
- Ventana espacial: $W = [0, 1500] \times [0, 1500]$ m.
- Ventana temporal: $T = [0, 365]$ días.
- Generación de patrones basada en estructura de ramificación (*branching structure*) del proceso self-exciting (Zhuang et al., 2004).
- Se simulan 10 patrones por escenario (no son muchas, pero el modelo es computacionalmente costoso).
- Se ajustan ambos modelos (estándar y propuesto) a cada patrón simulado.
- Se emplea una base de B-splines para la parte temporal y una base de thin-plate splines para la espacial.

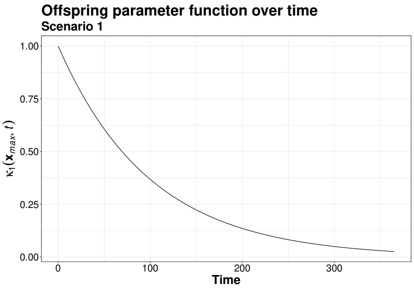
Algoritmo de simulación de patrones (resumen)

1. Generar eventos de fondo (*parents*) mediante un proceso de Poisson homogéneo con intensidad prefijada, λ_0.
2. Para cada evento, generar un número de descendientes (*offsprings*) con una $Po(\kappa(x,t))$, siguiendo la estructura de ramificación del proceso.
   La función $\kappa(x,t)$ representa cómo varía la generación de descendientes en espacio y tiempo.
3. Para cada evento, generar las localizaciones (x,t) de los descendientes a través de una Normal bivariante para la dispersión espacial y una exponencial para la temporal.
4. Repetir el procedimiento recursivamente con los nuevos eventos generados, hasta que no haya más descendientes.

Escenario 1

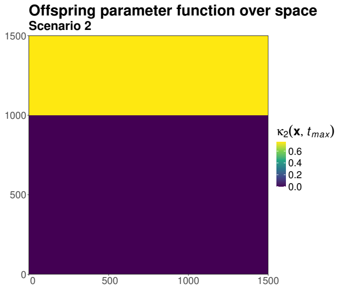
- $\kappa_1(x,t) = \dfrac{x_1 + x_2}{3000} e^{-t/100}$
- Mayor productividad hacia la esquina superior derecha de la ventana
- Disminución progresiva con el tiempo

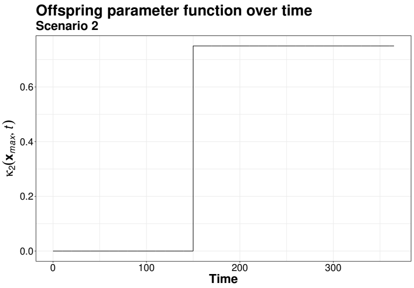




Escenario 2

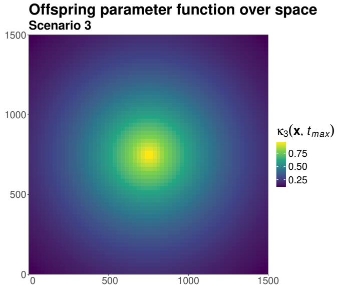
- $\kappa_2(x, t) = 0.75 \cdot \mathbf{1}_{[1000,1500]}(x_2) \cdot \mathbf{1}_{[150,365]}(t)$
- Productividad localizada únicamente en una zona específica de W y T
- El resto del espacio-tiempo se comporta como un proceso de Poisson homogéneo

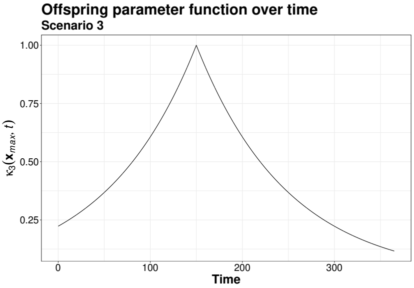




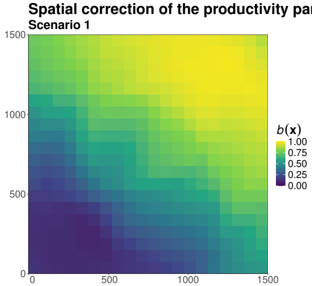
Escenario 3

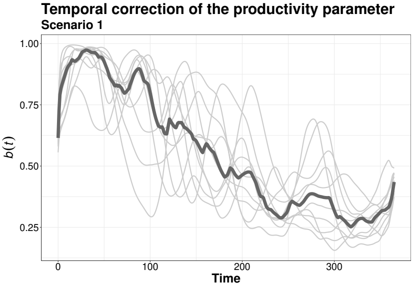
- $\kappa_3(x, t) = e^{-\frac{\|x-(750,750)\|}{500}} e^{-\frac{|t-150|}{100}}$
- Máximo de productividad en el centro de la ventana y en $t = 150$
- Decrece suavemente en espacio y tiempo



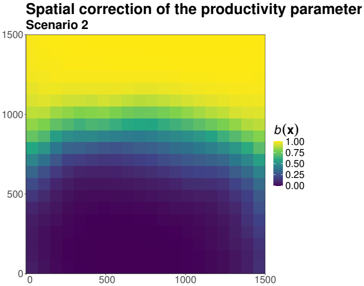


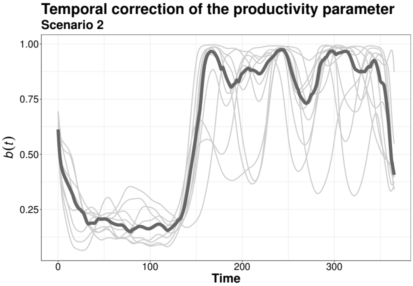
Resultados: Escenario 1



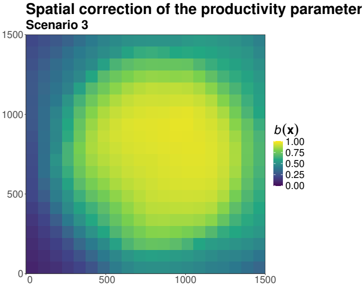


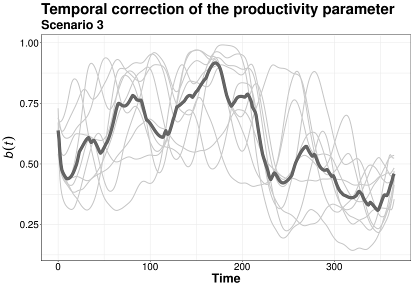
Resultados: Escenario 2





Resultados: Escenario 3





Conclusiones del estudio de simulación

- En general, el modelo propuesto, a través de los términos $b(x)$ y $b(t)$, recupera adecuadamente las variaciones de $\kappa(x,t)$.
- Mejor desempeño en escenarios 1 y 2 que en el 3 (emplear una base de splines más densa podría ayudar a mejorar los resultados).
- En general, la variabilidad de la estimaciones entre réplicas no es excesiva. Hay que tener en cuenta que los patrones simulados no son tan grandes (entre 180 y 350 eventos, aproximadamente).

Estudio de caso: Descripción general de los datos

- **Ventana espacial**: Ciudad de Valencia, España, en sus Distritos 1 a 6 (centro de la ciudad y alrededores)
- **Ventana temporal**: Año 2017
- **Eventos registrados**: 641 robos
- **Fuente**: Llamadas al 112 (número de emergencias)

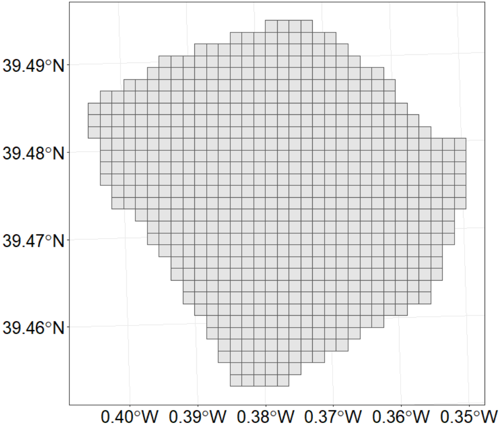
Área de estudio: Ventana espacial

Esta imagen contiene la zona de Valencia que consideraremos:

Figura: Imagen extraída de *OpenStreetMap*

Área de estudio: Ventana espacial

Trabajaremos sobre la siguiente malla, que denotamos G, la cual contiene 647 celdas de 150×150 m^2 (de área 22500 m^2):



Análisis exploratorio

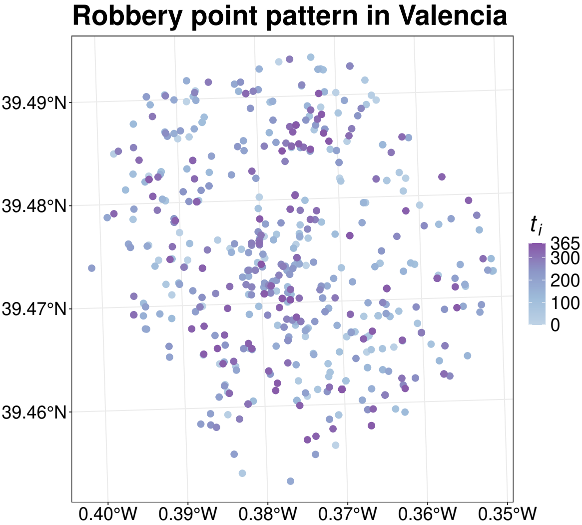


Figura: Patrón espacio-temporal observado

Análisis exploratorio

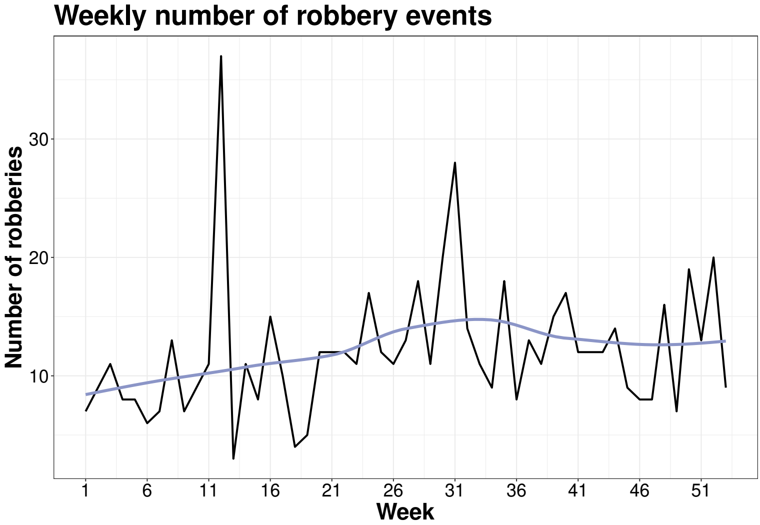
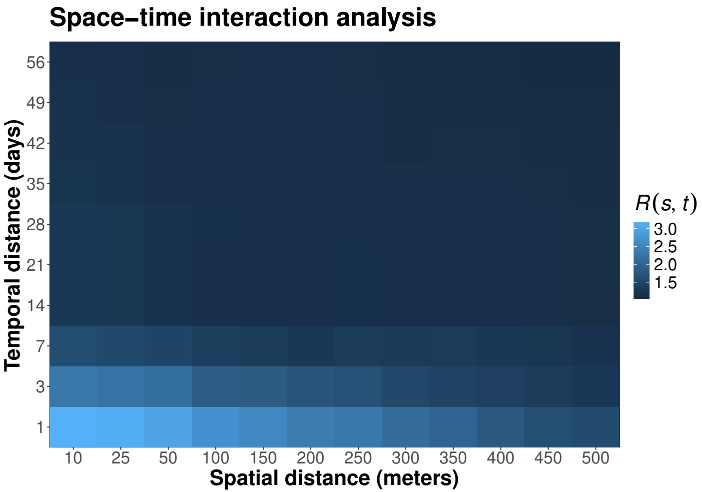


Figura: Número de eventos observados por semana

Interacción espacio-temporal

Se emplea el estadístico $R(s,t) = \frac{K_{ST}(s,t)}{K_S(s)K_T(t)}$, basado en la K-función, para detectar la existencia de interacción espacio-temporal en el patrón.



Estimación de los parámetros compartidos

El análisis exploratorio sugiere que existe interacción espacio-temporal, por lo que tiene sentido ajustar un modelo self-exciting. Pasamos a ver los resultados.

Model	Parameter	Mean	Lower	Upper
Standard	λ_0	$9.32 \cdot 10^{-8}$	$7.83 \cdot 10^{-8}$	$1.11 \cdot 10^{-7}$
	θ	$3.67 \cdot 10^{-7}$	$1.54 \cdot 10^{-7}$	$9.25 \cdot 10^{-7}$
	φ_S	$2.89 \cdot 10^{2}$	$6.95 \cdot 10^{1}$	$5.20 \cdot 10^{2}$
	φ_T	$1.14 \cdot 10^{0}$	$1.00 \cdot 10^{0}$	$1.51 \cdot 10^{0}$
Proposed	λ_0	$7.83 \cdot 10^{-8}$	$6.76 \cdot 10^{-8}$	$8.83 \cdot 10^{-8}$
	θ	$7.60 \cdot 10^{-7}$	$4.59 \cdot 10^{-7}$	$1.17 \cdot 10^{-6}$
	φ_S	$8.32 \cdot 10^{2}$	$5.95 \cdot 10^{2}$	$1.13 \cdot 10^{3}$
	φ_T	$1.12 \cdot 10^{0}$	$8.96 \cdot 10^{-1}$	$1.37 \cdot 10^{0}$

Cuadro: Media a posteriori e intervalo de credibilidad al 95 % para los parámetros compartidos por los dos modelos

Estimación de $b(x)$

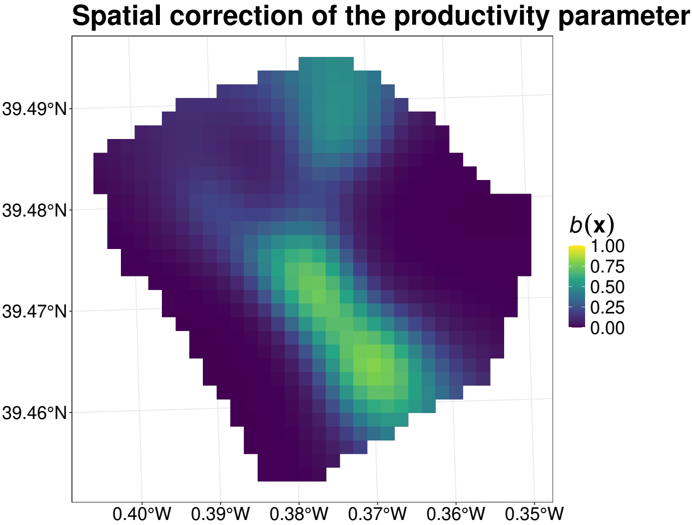


Figura: Estimación del factor $b(x)$ sobre el área de estudio

Estimación de $b(x)$: *Exceedance probabilities*

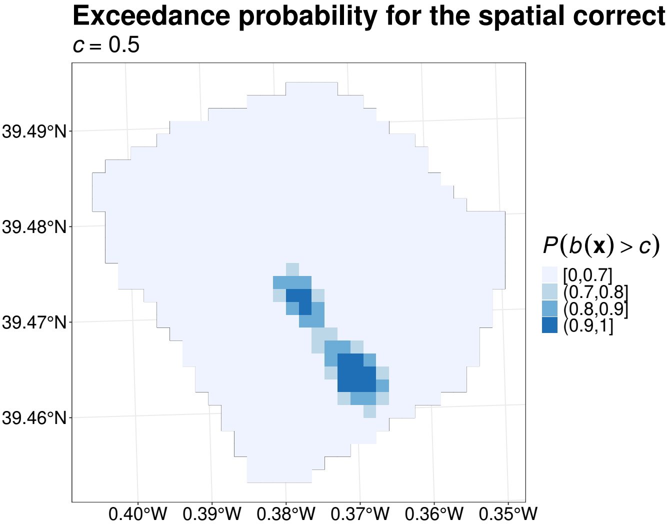


Figura: Exceedance probability calculada como $P(b(x) > 0.5)$

Estimación de $b(t)$

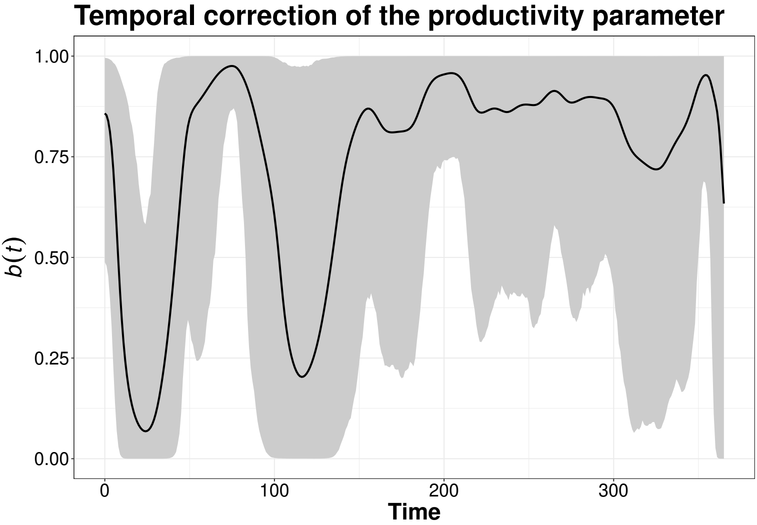


Figura: Estimación del factor $b(t)$ sobre el área de estudio

Estimación de $b(t)$: *Exceedance probabilities*

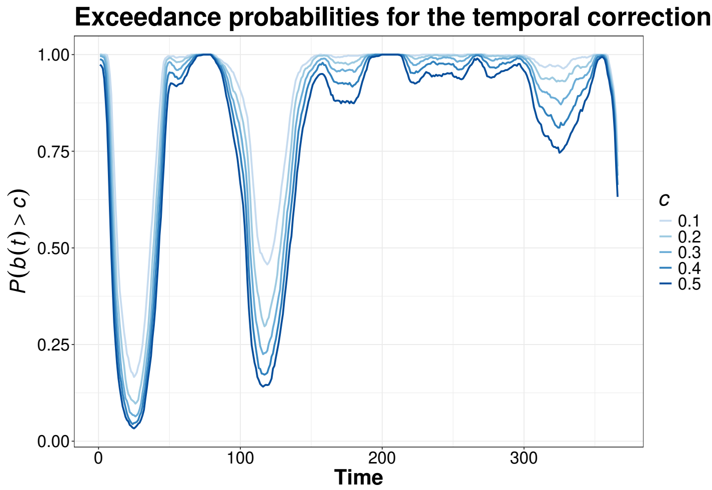
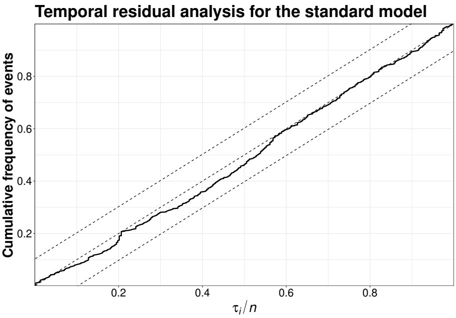


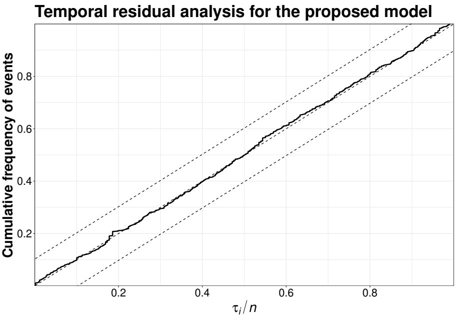
Figura: Exceedance probability calculada como $P(b(t) > c)$, para diversos valores de c

Análisis residual temporal

Se comprueba si los siguientes tiempos transformados se corresponden con un proceso proceso Poisson unitario ($\lambda(t) = 1$):

$$t_i \to \tau_i = \int_0^{t_i} \int_W \hat{\lambda}_\Theta(x, t|\mathcal{H}_t)\mathrm{d}x\mathrm{d}t$$



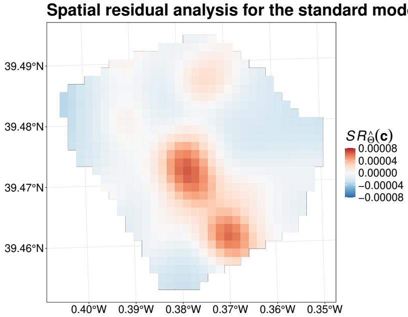


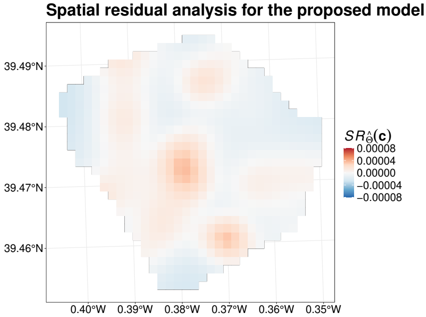
Análisis residual espacial

Propuesta de Baddeley et al. (2005) como residuo de un proceso puntual espacial:

$$SR_{\hat{\Theta}}(c) = \lambda^*(c) - \lambda^\dagger(c)$$

donde $\lambda^*(c)$ es una estimación no paramétrica de la función de intensidad, y $\lambda^\dagger(c)$ una versión suavizada del número estimado de evento en c, de acuerdo con el modelo ajustado.





Futura extensión: Introducción del concepto de *fairness*

- En aplicaciones **sensibles** (como la criminalidad), los modelos pueden amplificar sesgos espaciales o socioeconómicos preexistentes.
- Puede suceder que sobreestimemos la función de intensidad en localizaciones donde ciertas características poblacionales son más prevalentes.
- El concepto de *fairness* lleva tiempo estudiándose en diversas aplicaciones, incluyendo la criminología (Wu et al., 2023), con el objetivo de garantizar que no existan grupos poblacionales "perjudicados".
- Podemos tratar de controlar el nivel de *fairness* mediante la inclusión de variables sociodemográficas potencialmente susceptibles de ser perjudicadas y **regularizar** el efecto asociado a estas covariables (con métodos de tipo Lasso).

Conclusiones

- El modelo propuesto permite capturar heterogeneidad espacio-temporal en la productividad.
- Aporta interpretabilidad e identifica zonas y periodos con patrones repetitivos.
- El modelo es computacionalmente costoso, por lo que cualquier mejora en este sentido sería beneficiosa.
- El modelo podría extenderse para tener en cuenta el *fairness*.
- Queda también por estudiar el posible problema de identificación entre la tasa de fondo y el parámetro de productividad.

Bibliografía I

Baddeley, A., Turner, R., Møller, J., and Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(5):617–666.

Briz-Redón, Á. and Mateu, J. (2025). A self-exciting spatio-temporal model with a smooth space-time varying productivity parameter. Submitted for publication.

Chiang, W.-H., Liu, X., and Mohler, G. (2022). Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. *International Journal of Forecasting*, 38(2):505–520.

Daley, D. J. and Vere-Jones, D. (2002). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Springer New York.

Bibliografía II

Fox, E. W., Schoenberg, F. P., and Gordon, J. S. (2016). Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences. *The Annals of Applied Statistics*, pages 1725–1756.

Kwon, J., Zheng, Y., and Jun, M. (2023). Flexible spatio-temporal Hawkes process models for earthquake occurrences. *Spatial Statistics*, 54:100728.

Lewis, E. and Mohler, G. (2011). A nonparametric EM algorithm for multiscale Hawkes processes. *Journal of Nonparametric Statistics*, 1(1):1–20.

Marsan, D. and Lengline, O. (2008). Extending earthquakes' reach through cascading. *Science*, 319(5866):1076–1079.

Bibliografía III

Mohler, G. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting*, 30(3):491–497.

Mohler, G., Carter, J., and Raje, R. (2018). Improving social harm indices with a modulated Hawkes process. *International Journal of Forecasting*, 34(3):431–439.

Ntzoufras, I. (2011). *Bayesian modeling using WinBUGS*. John Wiley & Sons.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27.

Phillips, S. and Schoenberg, F. (2024). Efficient non-parametric estimation of variable productivity Hawkes processes. *Journal of Applied Statistics*, pages 1–18.

Bibliografía IV

Schoenberg, F. P. (2022). Nonparametric estimation of variable productivity Hawkes processes. *Environmetrics*, 33(6):e2747.

Schoenberg, F. P., Hoffmann, M., and Harrigan, R. J. (2019). A recursive point process model for infectious diseases. *Annals of the Institute of Statistical Mathematics*, 71:1271–1287.

Utsu, T., Ogata, Y., et al. (1995). The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth*, 43(1):1–33.

Wu, J., Abrar, S. M., Awasthi, N., and Frías-Martínez, V. (2023). Auditing the fairness of place-based crime prediction models implemented with deep learning approaches. *Computers, Environment and Urban Systems*, 102:101967.

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth*, 109(B5).

¡Muchas gracias por vuestra atención!