SORT 49 (2) July-December 2025, 1-30 DOI:

A stochastic partial differential equation for Bayesian spatio-temporal modeling of crime

Julia Calatayud[1], Marc Jornet[2], Javier Platero[3] and Jorge Mateu[4]

Abstract

We propose a stochastic partial differential equation to model geo-referenced data in the plane, with spatially correlated noise and a temporal log-normal evolution. Discretization in space permits us to develop the model in a finite-dimensional framework, reducing it to a set of stochastic differential equations coupled by correlated Wiener processes. The correlations are considered time-varying and stochastic, with a transformed log-normal distribution. The final model is framed within a hierarchical structure, and parameter inference is conducted jointly using Bayesian methods. The statistical methodology is illustrated by analyzing crime activity in the city of Valencia, Spain.

MSC: *35R60, 60H10, 62F15, 62M30, 62P25.*

Keywords: *Bayesian inference, crime time series, lattice data, space-time correlation, space-time intensity, stochastic log-Gaussian model, stochastic partial differential equation.*

1. Introduction

The use of differential equations in mathematical criminology started with the study of dynamic properties of urban crime hotspots, with the seminal paper of Short et al. (2008). The degree of attractiveness of each site influences the movement of burglars. The proposed model consists of two coupled reaction-diffusion partial differential equations describing the spatio-temporal evolution of density and attractiveness, giving rise to crime pattern formation. Building on this paper, extensions and further investigations have then been conducted on the dynamics of the reaction-diffusion system; see, for instance, the contributions by Short, Bertozzi and Brantingham (2010a; 2010b); Lloyd

[1] Department of Mathematics, University Jaume I, 12071 Castellón, Spain. calatayj@uji.es
[2] Escuela Superior de Ingeniería y Tecnología, Universidad Internacional de La Rioja, Av. La Paz 137, 26006 Logroño, Spain. marc.jornet@unir.net
[3] Department of Mathematics, University Jaume I, 12071 Castellón, Spain. platero@uji.es
[4] Department of Mathematics, University Jaume I, 12071 Castellón, Spain. mateu@uji.es
Received: May 2024
Accepted: April 2025

and O'Farrell (2013); Kolokolnikov, Ward and and Wei (2014); Tse and Ward (2016); Wang, Wang and Feng (2020); Rodriguez and Winkler (2022) and references therein.

Spatially homogeneous models of crime, based on temporal compartmental systems of ordinary differential equations, seem to begin with McMillon, Simon and and Morenoff (2014). The authors took a population-based approach, rooted in models that have been developed mainly to model the spread of diseases (Brauer, 2008; Acedo et al., 2010). The influence of peer pressure on criminal behavior (Burgess and Akers, 1966; Esiri, 2016; Harkins, Williams and Burger, 2017) is considered in a way reminiscent of a contagious disease; mathematically, the nonlinear terms in the compartments model complex interactions. Some papers on this type of models for crime evolution are given by Abbas, Tripathi and Neha (2017); González-Parra, Chen-Charpentier and Kojouharov (2018); Srivastav, Athithan and Ghosh (2020). Other disciplines employ the same approach, see Song et al. (2006); White and Comiskey (2007); Santonja et al. (2010) for drug or alcohol consumption, and Cervelló et al. (2014) for telecommunications, as good examples. Short reviews on this topic are published by Sooknanan and Comissiong (2017); Koss (2019). Recently, based on spatial epidemic models (van den Driessche, 2008; Wu, 2008; Schiesser, 2019), we extended this kind of compartmental model for crime to the case of spatial heterogeneity, by considering ordinary differential equations structured in patches and reaction-diffusion partial differential equations (Calatayud, Jornet and Mateu, 2025a).

These models contribute to the theory of mathematical criminology and the understanding of dynamical properties of criminality. However, because of the mechanisms involved in the models, it is certainly difficult to apply them when fitting to real data within a jurisdiction. The availability of adequate records, the lack of empirical values for model's parameters, or the divergence of the optimization procedure to calibrate the model's parameters (due to high dimensionality, stagnation at local minima, or unidentifiable values) are common issues when addressing the problem of data fitting. Therefore, the models are less useful for making specific predictions of crime events Kolokolnikov, Lloyd and Short (2019).

In the literature, works that aim to fit crime data with differential equations are more recent and scarce. Deterministic systems of ordinary differential equations were proposed by Lacey and Tsardakas (2016); Jane White et al. (2021). In Lacey and Tsardakas (2016), the authors studied serious and minor criminal events and built three coupled equations based on the attractiveness of the place. The monthly data on crime in Manchester were fitted by least-squares optimization. The authors highlighted that the inverse problem contained unidentifiable parameters. In Jane White et al. (2021), the authors used two coupled ordinary differential equations, based on population fluxes, to fit annual crime data in South Africa. The study region was divided into high- and low-conflicting areas. Bayesian inference was applied, with deterministic values reported. On the other hand, to accommodate fluctuations in crime time series and seek greater flexibility in modeling, stochastic differential equation models (Evans, 2012; Allen, 2007; Mao, 2007; Rackauckas, 2014) were investigated and calibrated in the works by Cao et al. (2013); Calatayud, Jornet and Mateu (2025b; 2023a; 2023b). In the report by Cao et al. (2013), the authors considered daily burglary data from Los Angeles (California) and Houston (Texas), and trend time series were modeled (for each city independently) with a stochastic Lotka-Volterra system, with independent Brownian noises. Least-squares fitting and maximum-likelihood estimation were carried out. In Calatayud et al. (2025b), criminality in Spain was fitted with a compartmental system of three ordinary differential equations (susceptible individuals, offenders in liberty, and offenders imprisoned), considering social influence with a nonlinear term, deterministic estimates with least-squares minimization, and nonlinear regression for generating confidence bands. An analysis of the basic reproduction number, conceived as the force of criminality in the region, was carried out. In Calatayud et al. (2023a), the events of aggression, theft, and woman alarm in the city of Valencia (Spain), were modeled with the stochastic log-normal model, by correlating two Brownian motions. Multidimensional correlations, beyond two Brownian motions, or spatial effects were not studied. This motivated the contribution by Calatayud et al. (2023b), who proposed stochastic differential equations of spatio-temporal type to investigate time-independent correlations of criminality between the twenty-six zip codes of the city of Valencia. The corresponding stochastic log-normal models were related by correlating Brownian processes.

The present paper continues the research of our previous contribution (Calatayud et al., 2023b). In particular, we aim at:

- Building a stochastic partial differential equation model for spatially referenced crime data with a random-field intensity, based on stochastic log-normal models;
- Studying spatio-temporal correlations of crime between regions, with transformed stochastic log-normal models;
- Framing all pieces of our model in a Bayesian hierarchical structure that provides an adequate framework for inference and forecasting;
- Illustrating the results on the real crime data of Valencia, with inference of the model's parameters.

Concerning the first point, we provide an adequate framework to Calatayud et al. (2023b), based on stochastic partial differential equations. The first paper that applied stochastic partial differential equations to real-data fitting in a spatial framework was proposed by Duan, Gelfand and Sirmans (2009), where urban housing development was modeled through a logistic-growth intensity function and Gaussian-process disturbances with a Matérn spatial covariance function (Gelfand et al., 2010) (but with no crime processes involved). Bayesian inference was used for inverse parameter estimation (Banerjee, Carlin and Gelfand, 2014; Lesaffre and Lawson, 2012). Such a mechanistic form for the intensity function was due to the clear sigmoid growth of the data. In our proposal, we focus on crime data instead, for the first time, and adopt a stochastic log-normal model, based on the financial literature (Lamberton and Lapeyre, 2011; Voit, 2010). The correlation is alternatively set by relating Brownian motions in a discretized space. Parameter estimation is computationally tractable within the Bayesian paradigm.

Regarding the second point, we extend the constant correlations in Calatayud et al. (2023b) to time-varying stochastic correlations. Some papers in finance treated stochastic correlations (van Emmerich, 2006; Teng, Ehrhardt and Günther, 2016) (the contribution by van Emmerich (2006) seems to be the pioneer). Parameter inference was conducted by fitting a stationary density to the empirical density of the historical-correlation time series. Nonetheless, the model was not tested on historical asset prices or correlations, but it was employed for option pricing. In our context of crime, we estimate the parameters for the transformed log-normal model instead, by using the Bayesian approach; these are the third and fourth points of the research. We are not aware of previous contributions that combine such methodology and techniques, and we think that it could be a good starting point for the analysis of several spatially distributed types of data. Here we particularize to the field of mathematical and statistical criminology, due to the significant importance of the topic in society.

The plan of the paper is the following. The methodological development comes in Section 2, where we present and develop the mathematical model, describe the corresponding differential equations, and frame the model into a hierarchical structure embedded in a Bayesian paradigm. Then, Section 3 presents the Bayesian implementation and the numerical and statistical results for the crime data of Valencia. The paper ends with some final considerations in Section 4, where a discussion and possible future extensions are given.

2. Formulation of the stochastic model

This section is devoted essentially to the design of the model. In Section 2.1, we start with a stochastic partial differential equation for the intensity function, with a spatially correlated noise and a temporal log-normal evolution. For geostatistical and lattice data, the model becomes finite dimensional and reduces to a set of stochastic differential equations, coupled by correlated Wiener processes. In Section 2.2, we further extend our proposal with a stochastic model for time-varying correlations. Hence, the intensity is in this case doubly stochastic and is framed within a Bayesian hierarchical structure.

2.1. Stochastic partial differential equations and spatio-temporal stochastic differential equations

An ordinary differential equation subject to an initial condition takes the form (Murray, 2002)

$$u'(t) = f(t, u(t), \Theta), \quad u(0) = u_0, \tag{1}$$

where $u(t)$ is the state variable with derivative $u'(t)$, f is a function that defines the model, Θ is a set of real parameters, and u_0 is the initial condition. This mathematical formula may be of use to model phenomena described by smooth curves, that are generally viewed as limits of difference equations. Although smoothness is rarely encountered in nature, an ordinary differential equation may be a useful tool to fit to average dynamics, especially when fluctuations in data are of small magnitude.

When fluctuations cannot be ignored, randomness can be incorporated into (1) through an irregular stochastic process, often a Gaussian white noise $B'(t)$ in $\mathbb{R}$ (Allen, 2007; Mao, 2007; Smith, 2013). The notation $B'(t)$ stands for a generalized derivative, since the formal differentiation of a standard Brownian-motion process $B(t)$ (Wiener process) yields a Gaussian white noise. The system (1) is then modified to (Evans, 2012; Allen, 2007; Mao, 2007; Rackauckas, 2014)

$$u'(t) = f(t,u(t),\Xi) + g(t,u(t),\Xi)B'(t), \quad u(0) = u_0,$$

where Ξ is a set of real parameters larger than or equal to Θ, g is a function that represents the intensity of the noise B', u is the stochastic state variable, and u_0 is the deterministic initial condition. Rigorously, the equation is interpreted in integral form under the theory of Itô calculus; therefore, a differential notation using d is employed, to give

$$\mathrm{d}u(t) = f(t,u(t),\Xi)\mathrm{d}t + g(t,u(t),\Xi)\mathrm{d}B(t), \quad u(0) = u_0. \tag{2}$$

Physically, and for modeling purposes, one may view the differential d as a very small increment. This pragmatic interpretation, which will be used extensively throughout the paper, agrees with the approximation of Itô stochastic differential equations by stochastic difference equations of Euler type (Euler-Maruyama scheme on a finite time mesh), in contrast to the Stratonovich formulation (Braumann, 2007). The increment $\mathrm{d}B(t) = B'(t)\mathrm{d}t$ is an uncorrelated Gaussian process with zero mean and variance $\mathrm{d}t$.

For modeling financial data (Lamberton and Lapeyre, 2011); Voit, 2010), where a single time series characterized by fluctuations is considered, an important model is usually employed: the log-normal equation or geometric Brownian-motion process, given by

$$\mathrm{d}u(t) = \mu u(t)\mathrm{d}t + \sigma u(t)\mathrm{d}B(t). \tag{3}$$

In the previous notation, $\Theta = \mu$, $\Xi = (\mu,\sigma)$, $f(t,u(t),\Xi) = \mu u(t)$ and $g(t,u(t),\Xi) = \sigma u(t)$. Parameter $\mu \in \mathbb{R}$, called drift, captures the growth rate, and $\sigma > 0$, called volatility, captures the magnitude of the fluctuations and the uncertainty on the future values. Essentially, the customary deterministic model for the infinitesimal growth rate, $(u(t+\mathrm{d}t) - u(t))/u(t) = \mu\mathrm{d}t$, is extended to a random setting as

$$\frac{u(t+\mathrm{d}t) - u(t)}{u(t)} \sim \mathrm{N}(\mu\mathrm{d}t, \sigma^2\mathrm{d}t), \tag{4}$$

with mutually independent perturbations, where N is the normal distribution [1]. By Itô's lemma, which extends the classical chain-rule theorem for non-differentiable processes,

[1] The randomization in (4) is the most consistent model. Indeed, consider a general model

$$\frac{u(t+\mathrm{d}t) - u(t)}{u(t)} = \mu\mathrm{d}t + \sigma(\mathrm{d}t)^{\gamma/2}\zeta_t,$$

where $\gamma \geq 0$ is a constant and ζ_t is an uncorrelated process with mean zero and variance one:

- The Gaussian behavior for ζ_t corresponds to the maximum-entropy distribution (Dorini and Samthe solution to (3) is given by the stochastic process

$$u(t) = u_0 \mathrm{e}^{(\mu - \frac{1}{2}\sigma^2)t + \sigma B(t)}.$$

The expected value of $u(t)$, $\mathbb{E}[u(t)]$, is the solution to the deterministic counterpart $\mathrm{d}\mathbb{E}[u(t)] = \mu \mathbb{E}[u(t)]\mathrm{d}t$,

$$\mathbb{E}[u(t)] = u_0 \mathrm{e}^{\mu t}. \tag{5}$$

Hence (3) can be interpreted as "mean" + "residual", for the differential. Since the probability distribution of $u(t)$ is log-normal, it is possible to compute a probabilistic interval at level $1-\alpha$,

$$\left[u_0 \mathrm{e}^{(\mu - \frac{1}{2}\sigma^2)t - \sigma \cdot \sqrt{t} \cdot q_{\alpha/2}}, u_0 \mathrm{e}^{(\mu - \frac{1}{2}\sigma^2)t + \sigma \cdot \sqrt{t} \cdot q_{1-\alpha/2}} \right], \tag{6}$$

where q is the quantile function of a $\mathrm{N}(0,1)$. The paths of u are continuous, autocorrelated, and nowhere differentiable. If $u_0 > 0$, then the paths of u are positive.

This log-normal model is not restricted to financial data; it can indeed be used whenever the observed fluctuations match the dynamics of (3). In fact, for crime time series, our prototype is (3). Intuitively, and despite the limitations, the evolution of crime incidence can be captured by the log-normal model, where there is a rate for the growth of criminality and a volatility for the random fluctuations. Albeit simple, the exponential model, rooted in birth-death environments, reflects imitative and social criminality (Burgess and Akers, 1966; Esiri, 2016; Harkins et al., 2017).

Until now, spatial effects have been omitted. To extend (3) to space, we consider a fixed compact region $D \subseteq \mathbb{R}^2$. Two settings may be considered: geostatistical data, where fixed (non-random) locations $s_1, \ldots, s_n \in D$ are studied, or lattice data, where D is partitioned into fixed (non-random) disjoint regions $D_1, \ldots, D_n$. In the context of criminology, the set D could represent a city of interest, for example. Lattice models divide D, for instance, into districts or streets. Geostatistical models deal with points of reference in the city, where we can have measurements of the quantity of interest. Strictly, these spatial formulations are not the same, but points of reference could also define a partition $\tilde{D} = \{s_1, \ldots, s_n\} = \{s_1\} \cup \ldots \cup \{s_n\}$ or $\tilde{D} = U_1 \cup \ldots \cup U_n$, where each U_i is a neighborhood of the representative s_i.

---

paio, 2012) (i.e. it has the minimum amount of information). Given an unbiased error, it is reasonable to presume a symmetric shape for its probability density. The normal distribution is mathematically convenient and simple.

- For $\gamma > 2$, the model is actually deterministic: $u' = \mu u$. For $\gamma \leq 2$, denote $\mathrm{d}G(t) = (\mathrm{d}t)^{\gamma/2}\zeta_t$. The correlation function of the noise, $\frac{\mathrm{d}G(t)}{\mathrm{d}t} = \frac{\zeta_t}{(\mathrm{d}t)^{1-\gamma/2}}$, is $C(t,\tau) = \mathbb{E}[\frac{\mathrm{d}G(t)}{\mathrm{d}t}, \frac{\mathrm{d}G(\tau)}{\mathrm{d}\tau}]$. For $t \neq \tau$, $C(t,\tau)\mathrm{d}\tau = 0$, because $\mathbb{E}[\zeta_t \zeta_\tau] = \mathbb{E}[\zeta_t]\mathbb{E}[\zeta_\tau] = 0$. For $t = \tau$, $C(t,\tau)\mathrm{d}\tau = \frac{1}{(\mathrm{d}\tau)^{1-\gamma}}$. The correlation only has an adequate structure when $\gamma = 1$, since in such a case, $C(t,\tau)\mathrm{d}\tau$ acts as the Dirac measure $\mathrm{d}\delta_t(\tau)$ and $C(t,\tau)$ acts as the Dirac-delta generalized function $\delta(t-\tau)$. Then, $\mathrm{d}G(t) = \mathrm{d}B(t)$. For $\gamma < 1$, $C(t,\tau) = \infty$ in the sense of distributions. For $\gamma > 1$, $C(t,\tau) = 0$ in the sense of distributions.

The process of interest in this paper is the (surface) intensity of crimes, $\Lambda(s,t)$, which is stochastic in both space and time. Conceptually, it is the expected number of crimes per unit area at coordinate s on the temporal interval $(t-1,t]$, for $s \in D$ and $t \in [0,T]$. The time horizon $0 < T < \infty$ comes up because data are always collected within a bounded period. This surface intensity on unit-time intervals is the integral of the surface-time intensity $\Gamma(s,\tilde{t})$:

$$\Lambda(s,t) = \int_{t-1}^{t} \Gamma(s,\tilde{t})\mathrm{d}\tilde{t}.$$

The intensities measure the risk of crime, hence an increase in the intensity could indicate a need of police intervention.

Based on (3), a stochastic partial differential equation model for the random field Λ is given by

$$\frac{\partial \Lambda(s,t)}{\partial t} = \mu(s)\Lambda(s,t) + \sigma(s)\Lambda(s,t)\xi(s,t), \tag{7}$$

for each point s within D, where $\mu(s) \in \mathbb{R}$ and $\sigma(s) \in (0,\infty)$ are spatially heterogeneous non-random functions, and ξ is a certain unbiased spatio-temporal noise independent of Λ, without further specification for now. The notation ∂ means a partial derivative, with respect to t here. In (7), the temporal evolution is given by (3), whereas the spatial distribution is determined by the noise ξ, which exhibits a certain spatial correlation. The system is subject to an initial state $\Lambda(s,0) = \int_{-1}^{0} \Gamma(s,\tilde{t})\mathrm{d}\tilde{t}$, $s \in D$. No boundary conditions are needed.

It is interesting to observe that (7) is equivalent to a stochastic integro-difference equation, like that proposed by Zammit-Mangion et al. (2012). These models describe the conditional dependence between the spatial field at a future-time point and the field at the present-time point through an integral operator, which is typically assumed to be linear (Zammit-Mangion and Wikle, 2020). Given time instants $t_k < t_{k+1}$ with $\Delta t = t_{k+1} - t_k$, one has

$$\begin{aligned} \Lambda(s,t_{k+1}) &\approx \Lambda(s,t_k) + \mu(s)\Lambda(s,t_k)\Delta t + \sigma(s)\Lambda(s,t_k)\xi(s,t_k)\Delta t \\ &= \int_D K(s,\tilde{s})F(\tilde{s},\Lambda(\tilde{s},t_k))\mathrm{d}\tilde{s} + e_k(s), \end{aligned} \tag{8}$$

where $K(s,\tilde{s}) = \delta(s-\tilde{s})$ is a Dirac-delta kernel, $F(\tilde{s},\Lambda(\tilde{s},t_k)) = \Lambda(\tilde{s},t_k) + \mu(\tilde{s})\Lambda(\tilde{s},t_k)\Delta t$ distorts the field in the sedentary stage, and $e_k(s) = \sigma(s)\Lambda(s,t_k)\xi(s,t_k)\Delta t$ is a zero-mean, correlated spatial disturbance. The δ-kernel corresponds to negligible spatial interactions through the drift and the volatility; spatial dependencies arise from the stochastic noise e_k and its covariance structure. The stochastic integro-difference formulation will not be used in the subsequent development; we will base our arguments on the differential form (7) instead.

The stochastic partial differential equation (7) (infinite-dimensional model) can actually be reduced to a set of n stochastic differential equations (finite-dimensional model), by taking the spatial discretization into account. For geostatistical data with locations

$s_1, \ldots, s_n$,

$$\frac{\partial \Lambda(s_i,t)}{\partial t} = \mu(s_i)\Lambda(s_i,t) + \sigma(s_i)\Lambda(s_i,t)\xi(s_i,t). \tag{9}$$

For lattice data with grid cells $D_1, \ldots, D_n$, if $\Lambda(D_i,t) = \int_{D_i} \Lambda(s,t)\mathrm{d}s$ is the aggregated intensity, then

$$\frac{\partial \Lambda(D_i,t)}{\partial t} = \int_{D_i} \mu(s)\Lambda(s,t)\mathrm{d}s + \int_{D_i} \sigma(s)\Lambda(s,t)\xi(s,t)\mathrm{d}s.$$

If we assume that $\mu(s)$, $\sigma(s)$ and the noise $\xi(s,t)$ are spatially homogeneous within D_i, given by $\mu(D_i)$, $\sigma(D_i)$ and $\xi(D_i,t)$, respectively (i.e., intra-region homogeneity and between-regions heterogeneity), then

$$\int_{D_i} \mu(s)\Lambda(s,t)\mathrm{d}s = \mu(D_i)\int_{D_i} \Lambda(s,t)\mathrm{d}s = \mu(D_i)\Lambda(D_i,t)$$

and

$$\int_{D_i} \sigma(s)\Lambda(s,t)\xi(s,t)\mathrm{d}s = \sigma(D_i)\xi(D_i,t)\int_{D_i} \Lambda(s,t)\mathrm{d}s = \sigma(D_i)\Lambda(D_i,t)\xi(D_i,t).$$

In consequence, for lattice data, we have

$$\frac{\partial \Lambda(D_i,t)}{\partial t} = \mu(D_i)\Lambda(D_i,t) + \sigma(D_i)\Lambda(D_i,t)\xi(D_i,t). \tag{10}$$

Both equations (9) and (10) can be combined into a single model, with the appropriate interpretations for Λ_i, μ_i, σ_i and ξ_i [2]:

$$\frac{\partial \Lambda_i(t)}{\partial t} = \mu_i\Lambda_i(t) + \sigma_i\Lambda_i(t)\xi_i(t). \tag{11}$$

[2] We have checked that the stochastic partial differential equation model for the intensity $\Lambda(s,t)$ implies temporal stochastic differential equations for aggregated intensities. The converse is also true. Suppose that any aggregated intensity

$$\Lambda(E,t) = \int_E \Lambda(\tilde{s},t)\mathrm{d}\tilde{s}$$

satisfies the stochastic differential equation

$$\frac{\partial \Lambda(E,t)}{\partial t} = \mu(E)\Lambda(E,t) + \sigma(E)\Lambda(E,t)\xi(E,t),$$

for small neighborhoods $E \subseteq D$ of $s \in D$ with area $|E|$. Then the local model

$$\frac{\partial \Lambda(s,t)}{\partial t} = \mu(s)\Lambda(s,t) + \sigma(s)\Lambda(s,t)\xi(s,t)$$

is retrieved, since

$$\Lambda(s,t) = \lim_{|E|\to 0} \frac{1}{|E|}\int_E \Lambda(\tilde{s},t)\mathrm{d}\tilde{s},$$

$$\mu(s)\Lambda(s,t) = \lim_{|E|\to 0} \frac{1}{|E|}\int_E \mu(\tilde{s})\Lambda(\tilde{s},t)\mathrm{d}\tilde{s}$$

and

$$\sigma(s)\Lambda(s,t)\xi(s,t) = \lim_{|E|\to 0} \frac{1}{|E|}\int_E \sigma(\tilde{s})\Lambda(\tilde{s},t)\xi(\tilde{s},t)\mathrm{d}\tilde{s}.$$

Notice that for lattice data, $\Lambda_i(t)$ gives an absolute quantity, not a density: the number of crimes in region D_i, on $(t-1,t]$. Indeed, the spatial integration of an intensity gives the "mass" (usually, in the literature, researchers simply talk about the number of crimes in a region at t for easiness, to really refer to $(t-1,t]$). For geostatistical data, by contrast, $\Lambda_i(t)$ is still an intensity: the surface density of crimes at s_i, on $(t-1,t]$.

Equations in (11) are formulated in $\mathbb{R}$, instead of $\mathbb{R}^2$. Each $\xi_i(t)$ can be defined as a one-dimensional Gaussian white-noise process, given by the derivative of a one-dimensional Brownian motion, $B_i'(t)$. Therefore, (11) is rewritten in differential form as

$$\mathrm{d}\Lambda_i(t) = \mu_i \Lambda_i(t)\mathrm{d}t + \sigma_i \Lambda_i(t)\mathrm{d}B_i(t), \tag{12}$$

for $i = 1, \ldots, n$. The closed-form solutions are

$$\Lambda_i(t) = \Lambda_i(0)\mathrm{e}^{(\mu_i - \frac{1}{2}\sigma_i^2)t + \sigma_i B_i(t)}. \tag{13}$$

The specification of a log-normal process is equivalent to the specification of a Gaussian process. We note that if $\Lambda_i(t) = \mathrm{e}^{z_i(t)}$, then (12) is equivalent to $\mathrm{d}z_i(t) = (\mu_i - \frac{1}{2}\sigma_i^2)\mathrm{d}t + \sigma_i \mathrm{d}B_i(t)$. This identity is related to the stochastic integro-difference equation proposed by Zammit-Mangion et al. (2012), not for the intensity itself with (8), but for the logarithm of the intensity. In Euler discrete form, $z_i(t_{k+1}) \approx z_i(t_k) + \varepsilon_i(t_k)$, where the mean of the Gaussian variable $\varepsilon_i(t_k)$ is $(\mu_i - \frac{1}{2}\sigma_i^2)\Delta t_k$ and the variance is $\sigma_i^2 \Delta t_k$. The covariance structure between two indices i, j, under spatial heterogeneity, will be specified later. Note that this model is the stochastic integro-difference equation with Dirac-delta kernel K and identity map F.

The original two-dimensional noise $\xi(s,t)$ is given by

$$\xi(s,t) = \sum_{i=1}^{n} \mathcal{X}_{\{s_i\}}(s) B_i'(t), \quad s \in \{s_1, \ldots, s_n\},$$

for geostatistical data, and by

$$\xi(s,t) = \sum_{i=1}^{n} \mathcal{X}_{D_i}(s) B_i'(t), \quad s \in \bigcup_{i=1}^{n} D_i = D,$$

for lattice data, where $\mathcal{X}_{\cdot}(\cdot)$ denotes the indicator function over a set. Under a higher level of abstraction, we note that bivariate noise is a particular form of the appealing expression

$$\xi(s,t) = \sum_i w_i(s)\eta_i(t),$$

where η_i are temporal noises with a certain interdependence over i, and w_i are spatial functions (weights) that distribute those temporal noises in space. Separation of variables decomposes the domain $D \times [0,T]$ into D and $[0,T]$ and simplifies the problem [3].

[3] The decomposition of Cartesian-product domains for random-field representations is a common tool in stochastic modeling. For instance, similar approaches are adopted for stochastic systems driven by parametric uncertainty, with Karhunen-Loève expansions or polynomial chaos expansions (Lord, Powell and Shardlow, 2014; Xiu, 2010).

When w_i are indicator functions, the space is then discretized and the stochastic partial differential equation becomes a set of stochastic differential equations.

In the notation of lattice data, but the same applies for geostatistical data, time-value properties of our ξ are: (a) $\xi(s,*)$ is a Gaussian process; (b) the expectation becomes $\mathbb{E}[\xi(s,t)] = \sum_{i=1}^n \mathcal{X}_{D_i}(s)\mathbb{E}[B_i'(t)] = 0$; (c) and the covariance can be written as

$$\begin{aligned}\mathbb{C}\text{ov}[\xi(s,t_1),\xi(s,t_2)] &= \sum_{i,j=1}^n \mathcal{X}_{D_i}(s)\mathcal{X}_{D_j}(s)\mathbb{C}\text{ov}[B_i'(t_1),B_j'(t_2)] \\ &= \sum_{i=1}^n \mathcal{X}_{D_i}(s)\mathbb{C}\text{ov}[B_i'(t_1),B_i'(t_2)] \\ &= \delta(t_1-t_2)\sum_{i=1}^n \mathcal{X}_{D_i}(s) = \delta(t_1-t_2),\end{aligned}$$

where δ is the Dirac delta function. That is, $\xi(s,*)$ is a Gaussian white-noise stochastic process.

To set the properties of $\xi(*,t)$, since $\xi(s,t)$ exhibited spatial association, the n Brownian motions are correlated. This feature couples the n equations (12). No coupling is due to the drift or the volatility, which does not seem to contribute to lower fitting or prediction accuracy in practice.

To our knowledge, such a decomposition of the spatio-temporal noise in stochastic partial differential equations has not been used previously.

When $n = 2$ or, in other terms, we work pairwise, one can define

$$\mathrm{d}B_2(t) = \rho_{1,2}(t)\mathrm{d}B_1(t) + \sqrt{1-\rho_{1,2}(t)^2}\mathrm{d}\tilde{B}_2(t),$$

where $\rho_{1,2}(t) \in [-1,1]$ is a function of t and $\tilde{B}_2(t)$ is an auxiliary Brownian motion that is independent of $B_1(t)$. This differential equation is rigorously interpreted by Itô integration,

$$B_2(t) = \int_0^t \rho_{1,2}(\tilde{t})\mathrm{d}B_1(\tilde{t}) + \int_0^t \sqrt{1-\rho_{1,2}(\tilde{t})^2}\mathrm{d}\tilde{B}_2(\tilde{t}).$$

By the bilinearity of the covariance operator, it is easy to see that the correlation becomes

$$\mathrm{corr}[\mathrm{d}B_1(t),\mathrm{d}B_2(t)] = \rho_{1,2}(t).$$

For $\Lambda_i(t)$, this property translates into

$$\mathrm{corr}[\mathrm{d}\Lambda_1(t),\mathrm{d}\Lambda_2(t)|\Lambda_1(t),\Lambda_2(t)] = \rho_{1,2}(t),$$

where the vertical bar denotes a conditional quantity. That is, $\rho_{1,2}$ measures how similar the (infinitesimal) variations of $\Lambda_1(t)$ and $\Lambda_2(t)$ are. In practice, this is a key quantity: if $\rho_{1,2} > 0$ is somewhat near 1, then a significant variation in crime incidence within region D_1 should make the public authorities and the police put their attention on D_2 also.

In the general case, the theoretical construction of n correlated Brownian motions is as follows. Given n independent auxiliary Brownian processes $\tilde{B}_1(t) = B_1(t)$, $\tilde{B}_2(t), \ldots,$ $\tilde{B}_n(t)$, define

$$\begin{pmatrix} \mathrm{d}B_1(t) \\ \vdots \\ \mathrm{d}B_n(t) \end{pmatrix} = \mathrm{L}(t) \begin{pmatrix} \mathrm{d}\tilde{B}_1(t) \\ \vdots \\ \mathrm{d}\tilde{B}_n(t) \end{pmatrix},$$

where $\mathrm{L}(t)$ is a lower-triangular matrix with transpose $\mathrm{L}^\top(t)$ and $\mathrm{A}(t) = (\rho_{i,j}(t))_{i,j} = \mathrm{L}(t)\mathrm{L}^\top(t)$ is the Cholesky decomposition for the symmetric and positive definite matrix $\mathrm{A}(t)$ of correlation functions. Indeed, the relations

$$\mathrm{corr}[\mathrm{d}B_i(t), \mathrm{d}B_j(t)] = \rho_{i,j}(t) \tag{14}$$

and

$$\mathrm{corr}[\mathrm{d}\Lambda_i(t), \mathrm{d}\Lambda_j(t) | \Lambda_i(t), \Lambda_j(t)] = \rho_{i,j}(t)$$

hold. Also, for (14), notice that

$$\Sigma_{\mathrm{d}B} = \mathrm{L}\Sigma_{\mathrm{d}\tilde{B}}\mathrm{L}^\top = \mathrm{d}t\mathrm{L}\mathrm{L}^\top = \mathrm{d}t\mathrm{A},$$

where Σ denotes the covariance matrix, $B = (B_1, \ldots, B_n)^\top$ and $\tilde{B} = (\tilde{B}_1, \ldots, \tilde{B}_n)^\top$.

As a consequence, if $s \in D_k$ (or $s = s_k$) and $r \in D_l$ (or $r = s_l$), then

$$\begin{aligned} \mathbb{C}\mathrm{ov}[\xi(s,t)\mathrm{d}t, \xi(r,t)\mathrm{d}t] &= \sum_{i,j=1}^{n} \mathcal{X}_{D_i}(s)\mathcal{X}_{D_j}(r)\mathbb{C}\mathrm{ov}[B_i'(t)\mathrm{d}t, B_j'(t)\mathrm{d}t] \\ &= \sum_{i,j=1}^{n} \mathcal{X}_{D_i}(s)\mathcal{X}_{D_j}(r)\mathbb{C}\mathrm{ov}[\mathrm{d}B_i(t), \mathrm{d}B_j(t)] \\ &= \sum_{i,j=1}^{n} \mathcal{X}_{D_i}(s)\mathcal{X}_{D_j}(r)\rho_{i,j}(t)\mathrm{d}t = \rho_{k,l}(t)\mathrm{d}t \end{aligned}$$

and

$$\mathrm{corr}[\xi(s,t)\mathrm{d}t, \xi(r,t)\mathrm{d}t] = \rho_{k,l}(t).$$

This is the spatial structure of the noise term in the stochastic partial differential equation. Then, the spatial structure of the intensity becomes

$$\begin{aligned} \mathbb{C}\mathrm{ov}[\mathrm{d}\Lambda(s,t), \mathrm{d}\Lambda(r,t) | \Lambda(s,t), \Lambda(r,t)] &= \sigma(s)\sigma(r)\Lambda(s,t)\Lambda(r,t)\mathbb{C}\mathrm{ov}[\xi(s,t)\mathrm{d}t, \xi(r,t)\mathrm{d}t] \\ &= \sigma(s)\sigma(r)\Lambda(s,t)\Lambda(r,t)\rho_{k,l}(t)\mathrm{d}t \end{aligned}$$

implying that

$$\mathrm{corr}[\mathrm{d}\Lambda(s,t), \mathrm{d}\Lambda(r,t) | \Lambda(s,t), \Lambda(r,t)] = \rho_{k,l}(t),$$

where the differential is taken with respect to t. At different spatial locations, the infinitesimal increments of the intensity over time are dependent via such a ρ-function.

2.2. Stochastic correlations

The most basic spatial correlation model assumes that $\rho_{i,j}(t) \equiv \rho_{i,j}$ is time independent (Calatayud et al., 2023b). This is a good option from an "averaged" point of view, when one uses a single quantity to summarize the relationship between the two historical time series. Moreover, a confidence interval for the correlation gives a region where the (constant) quantity lies with high confidence (Davison and Hinkley, 1997). However, in general, the correlation changes with time. This fact has been observed for data in finance (van Emmerich, 2006; Teng et al., 2016), and for crime dynamics the same seems to occur. It is interesting to notice that, in terms of time-series modeling, spatial dependencies between regions are somewhat similar to assets' dependencies between companies in the same financial market; hence the mathematical connection between spatial statistics and finance. After selecting a time lag $\tau > 0$, sample correlations of the log returns can be computed at blocks $[t-\tau, t]$ (moving-window technique), and it is observed that a noisy time series describes the dynamics of the empirical correlation; when τ is lower, the fluctuations are higher.

Thus, for completeness we better consider the correlation $\rho_{i,j}(t)$ as a stochastic process. Since it must belong to the interval $[-1,1]$, an appropriate transformation is used. In particular, making use of the hyperbolic tangent and the logarithm functions, we define

$$\rho_{i,j}(t) = \tanh(\log(y_{i,j}(t))) \in [-1,1], \tag{15}$$

where $y_{i,j}(t)$ is described by a log-normal model, as in (3), and it is given by

$$\mathrm{d}y_{i,j}(t) = \mu_{i,j} y_{i,j}(t)\mathrm{d}t + \sigma_{i,j} y_{i,j}(t)\mathrm{d}W_{i,j}(t), \tag{16}$$

where $W_{i,j}(t)$ are independent Brownian motions (hence the correlation processes operate independently), $\mu_{i,j} \in \mathbb{R}$, and $\sigma_{i,j} > 0$, for $i, j = 1, \ldots, n$. Notice that the hyperbolic tangent function has the image in $(-1,1)$, so $\rho_{i,j}$ is well-defined.

After combining (12), (14), (15) and (16), the final model we propose in this paper becomes hierarchical, and takes the form

$$\begin{cases} \mathrm{d}\Lambda_i(t) = \mu_i \Lambda_i(t)\mathrm{d}t + \sigma_i \Lambda_i(t)\mathrm{d}B_i(t), \quad i = 1, \ldots, n, \\ \mathrm{corr}[\mathrm{d}B_i(t), \mathrm{d}B_j(t) | W_{i,j}] = \mathrm{corr}[\mathrm{d}\Lambda_i(t), \mathrm{d}\Lambda_j(t) | \Lambda_i(t), \Lambda_j(t), W_{i,j}] = \rho_{i,j}(t), \\ \rho_{i,j}(t) = \tanh\left(\log\left(y_{i,j}(t)\right)\right), \\ \mathrm{d}y_{i,j}(t) = \mu_{i,j} y_{i,j}(t)\mathrm{d}t + \sigma_{i,j} y_{i,j}(t)\mathrm{d}W_{i,j}(t), \quad i, j = 1, \ldots, n. \end{cases} \tag{17}$$

Compared to (14), the second line in the hierarchy now conditions on $W_{i,j}$, which is the source of randomness for $\rho_{i,j}$; each path of $W_{i,j}$ defines a time function $\rho_{i,j}$. A loose schematic view on (17) is the following

$$\mathrm{data}_i \sim [\Lambda_i | B_1, \ldots, B_n; \mu_i, \sigma_i] \times [B_i, B_j | \rho_{i,j}] \times [\rho_{i,j} | W_{i,j}; \mu_{i,j}, \sigma_{i,j}] \times [W_{i,j}],$$

where the vertical bar conditions on random quantities, the semicolon conditions on (non-random) parameters, and the products multiply probability laws. Recall that the subscripts i,j refer to spatial conditions: for geostatistical data, i refers to the fixed spatial location $s_i \in D$, whereas for lattice data, i refers to the region $D_i \subseteq D$. On the other hand, by Itô's lemma, one may derive a stochastic differential equation for $\rho_{i,j}(t)$ after some computations, without $y_{i,j}(t)$, but it will not be used for calibration purposes.

Mechanistically, model (17) is interpreted in the context of crime dynamics as follows. Omitting the subscript i for simplicity, the deterministic part of the first equation is $\Lambda'(t) = \mu\Lambda(t)$, for the region i. By relating number of crimes with number of criminals as $\Lambda(t) = \alpha x(t)$, where $\alpha > 0$ is a proportionality constant that reflects the average quantity of committed crimes per criminal in region i (Jane White et al., 2021), a model for the number of criminals is $x'(t) = \mu x(t)$. This equation takes into account: the social influence of criminality (Burgess and Akers, 1966; Esiri, 2016; Harkins et al., 2017), with an inflow $\mu_{\text{in}} x(t)$ that generates new offenders, and the cessation of criminal activity, with an outflow $\mu_{\text{out}} x(t)$, that gives rise to the continuous evolution

$$x(t+\mathrm{d}t) = x(t) + (\mu_{\text{in}} - \mu_{\text{out}})x(t)\mathrm{d}t.$$

Parameter μ, rooted in birth-death processes of populations, is thus interpretable as the balance between social influence and cessation of criminality, and it determines x when viewed as a deterministic function. However, there are uncertainties associated to crime evolution, hence the deterministic formulation for the relative change

$$\frac{x(t+\mathrm{d}t) - x(t)}{x(t)} = \mu\mathrm{d}t$$

is normally perturbed as

$$\frac{x(t+\mathrm{d}t) - x(t)}{x(t)} \sim \mathrm{N}(\mu\mathrm{d}t, \sigma^2\mathrm{d}t).$$

Parameter σ represents the volatility of the dynamics, for each region i. The mathematical formalization of this last equation gives rise to the stochastic differential equation of the Itô type. Now, the distinct regions are not independent and are related via correlated noises. For a higher fitting potential, correlations are also taken to be stochastic, but we employ a phenomenological formulation for them because we have no prior sociological information about the evolution. The use of the auxiliary functions tanh and log are mathematical artifacts to ensure that the correlation processes remain within $(-1,1)$.

3. Crime data dynamics from calls to 112-emergency phone

This section presents the application of our methodological approach to the analysis of crime dynamics in the city of Valencia, Spain. In particular, we develop the framework and strategy for Bayesian inference to fit our stochastic differential equations to our crime data. In Section 3.1, we present the data and provide some notations. Then Section 3.2 outlines the Bayesian strategy for statistical inference, and presents the results: posterior distributions, fitting to the historical data, and predictions.

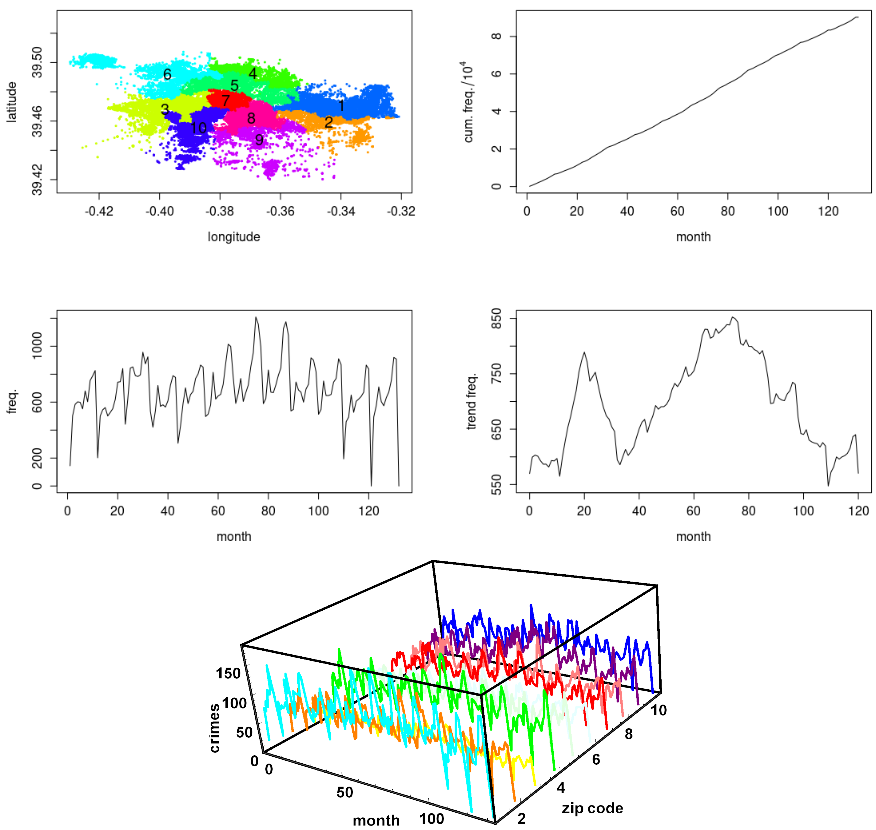

Figure 1. First panel*: Locations of crime events in Valencia partitioned in colors per each one of the* 10 *postal codes, for the period from January* 2010 *to December* 2020*. Then, each zip code has a time series of monthly crime counts.* Second panel*: Cumulative number of crime incidents in Valencia.* Third panel*: Frequency of crime events in Valencia.* Fourth panel*: Trend frequency (annual moving average) of crime events in Valencia.* Fifth panel*: 3D box showing the ten temporal series for each postal code.*

3.1. Crime-related setting: methodological setup

We work in the context of the recent papers by Calatayud et al. (2023a,b). The data consist of 90247 street crime incidents communicated to the 112-emergency phone in the Mediterranean city of Valencia, Spain, from January 2010 to December 2020. Essentially, these correspond to violent, smooth robberies and thefts in the streets. Valencia has around $800,000$ inhabitants, ranked third in Spain, and it is the capital of the Valencian region; as a large and populated city, crime patterns stand as an important societal issue. Monthly counts of crime (132 measurements along 11 years) are available, positioned here in 10 regions of the city of Valencia based on their postal or zip codes.

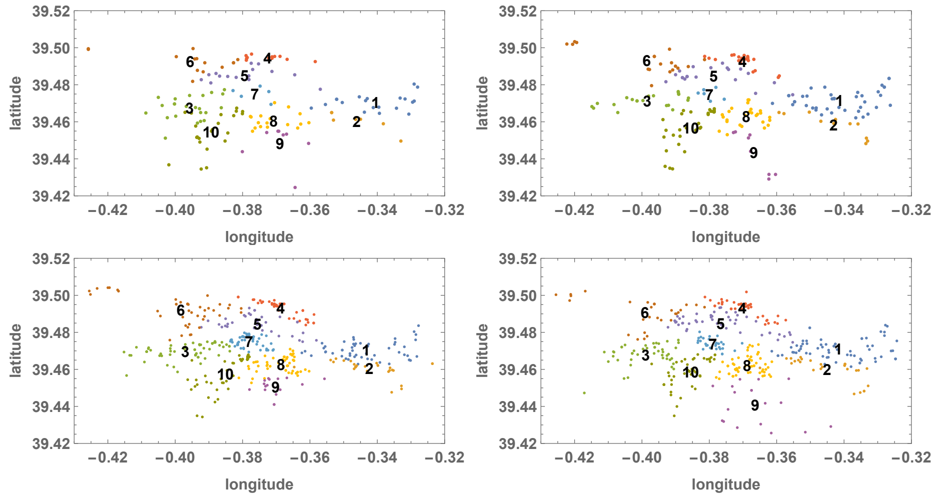

Figure 2. *Monthly spatial distribution of crime for the first four months of the data for all years: January (top-left), February (top-right), March (bottom-left), and April (bottom-right).*

The city is represented in the plane by the set D. The regional partition into 10 zip codes provides the context for lattice-data analysis. We have districts $D_i \subseteq D$, for $i = 1,\ldots,n$, where $n = 10$ and $D = \cup_{i=1}^{n} D_i$. The random-field intensity $\Lambda(s,t)$, defined as the expected number of crimes per unit area at s during month t (temporal interval $(t-1,t]$), is used in terms of aggregated intensity ("mass"): $\Lambda_i(t) = \Lambda(D_i,t) = \int_{D_i} \Lambda(s,t)\mathrm{d}s$. The integration of the density gives $\Lambda_i(t)$, representing the number of crimes in district D_i, at month t (temporal interval $(t-1,t]$). We assume a proportional relationship in terms of average number of crimes committed per criminal (Jane White et al., 2021), per zip code: incidents $= \alpha \times$ offenders, $\alpha > 0$; we then focus on the number of crimes. The time horizon is $T = 131$, and the time interval $[0,131]$ is partitioned into $t_1 < \ldots < t_m$, $m = 132$, with $t_{k+1} - t_k = \Delta t = 1$. The empirical data on street crimes are represented by λ_{i,t_k}. There are $n = 10$ time series of interest $\{\lambda_{i,t_k}\}_{k=1}^{m}$, for $i = 1,\ldots,n$. The raw time series are very noisy and Itô processes do not fit well, so an annual moving average is applied from the beginning to remove seasonality and outliers, and to accommodate geometric Brownian motions to trends, giving $\lambda_{i,t_k} \leftarrow \frac{1}{12}\sum_{\ell=0}^{11} \lambda_{i,t_{k-\ell}}$ and $m = 121$. This type of data processing was suggested by Cao et al. (2013). The new and smoother trend time series $\{\lambda_{i,t_k}\}_{k=12}^{m}$ are those used to fit our model. Indeed, while the raw time series are similar to a white noise due to the abrupt variability observed, the corresponding trend time series are better described as an Itô process (Cao et al., 2013; Calatayud et al., 2023b). The locations of crimes in Valencia are depicted in Figure 1, indicating each of the 10 zip codes that we are considering here. The bottom panel of Figure 1 shows basic information on the temporal evolution of the incidents, based on the cumulative frequency in Valencia. The third picture depicts the overall crime frequency in the city, in form of a time series. In the fourth panel, the time series is smoothed with the annual moving average. The fifth panel displays a $3D$ box containing the ten temporal series for each postal code, enabling a simultaneous view of crime across both spatial and temporal dimensions.

To further investigate the temporal and spatial dynamics, Figure 2 presents the spatial distribution of crimes for the first four months for all years. Finally, Figure 3 shows the spatial distribution of crimes for the first four years. Each map represents a single year, allowing to identify and compare long-term spatial patterns.

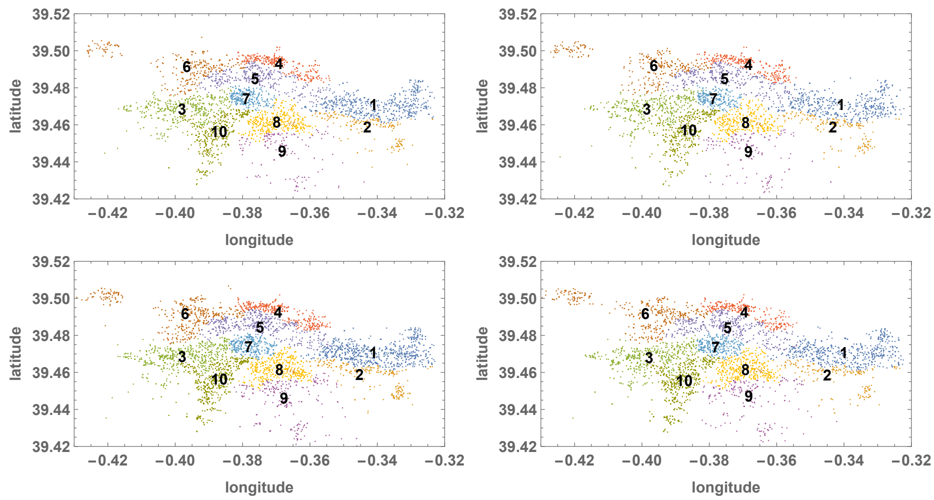

Figure 3. *Annual spatial distribution of crime for the first four years: Year 1 (top-left), Year 2 (top-right), Year 3 (bottom-left), and Year 4 (bottom-right).*

These figures collectively contribute to a more comprehensive understanding of the spatial distribution of crime in Valencia and its temporal evolution that will be interpreted through our stochastic model.

3.2. Bayesian inference for hierarchical model estimation

We assume discrete data labeled as $\{\lambda_{i,t_k}\}$, with $k = 1,\dots,m$ and $i = 1,\dots,n$, where subscript i refers to space, t_k represents an instant of time, and the time lags $\Delta t = t_{k+1} - t_k = 1$ are constant. These records are regarded as realizations of our hierarchical model of the Itô type.

We used the software Mathematica (Wolfram Research, 2020), for some preliminary computations with the data, and R (R Core Team, 2023) for the complete self-implementation of the Bayesian inference framework.

Given the data at the i-th location, $\lambda_{i,t_1},\dots,\lambda_{i,t_m}$, we use a Bayesian approach following a Markov Chain Monte Carlo (MCMC) method for parameter estimation. Specifically, we use Gibbs sampling for the one-dimensional parameters and slice sampling for the multivariate settings (i.e., $Z_{i,t}$ and $W_{i,j}$, as coming next) (Brooks et al., 2011; Neal, 2003). Furthermore, we note that Λ_i, as defined in (17), may show negative intensity values when the stochastic differential equation is discretized with an Euler-Maruyama scheme of the form

$$\begin{aligned}\Lambda_i(t_{k+1}) &= \Lambda_i(t_k) + \mu_i \Lambda_i(t_k) + \sigma_i \Lambda_i(t_k)(B_i(t_{k+1}) - B_i(t_k)) \\ &\sim \mathrm{N}\left(\Lambda_i(t_k) + \mu_i \Lambda_i(t_k), \sigma_i^2 \Lambda_i(t_k)^2\right),\end{aligned}$$

$$\text{support}(\Lambda_i(t_{k+1})) \in (-\infty, \infty).$$

which is not realistic. Therefore, we use a logarithmic transformation by taking advantage of Itô's lemma, see (13), and we propose the following hierarchical model, originated from (17):

$$\lambda_i(t)|\Lambda_i(t) \sim \mathrm{Po}\left(\Lambda_i(t)\right) \tag{18}$$

with Po() standing for a counting Poisson distribution, and

$$\log\left(\Lambda_i(t+1)\right) = \log\left(\Lambda_i(t)\right) + \left(\mu_i - \frac{1}{2}\sigma_i^2\right) + \sigma_i Z_{i,t}, \tag{19}$$

with

$$(Z_{1,A_l}, \ldots, Z_{n,A_l}) \sim \mathrm{N}_n(0, \Sigma_{A_l}), \quad \Sigma_{A_l} = (\rho_{i,j,A_l})_{1 \le i,j \le n}, \tag{20}$$

where A_l are chosen subsets of the complete temporal interval. In our case, we divided the total temporal region into $L = 10$ subintervals A_l, each corresponding to one year, so $l = 1, \ldots, L = 10$. In addition,

$$\begin{aligned}\rho_{i,i,A_l} &= 1, \qquad \forall 1 \le i \le n, \\ \rho_{i,j,A_l} &= \tanh(\log(y_{i,j}(A_l))),\end{aligned} \tag{21}$$

where

$$\log\left(y_{i,j}(A_{l+1})\right) = \log\left(y_{i,j}(A_l)\right) + \left(\mu_{i,j} - \frac{1}{2}\sigma_{i,j}^2\right) + \sigma_{i,j} W_{i,j}. \tag{22}$$

Equation (18), added to (17) in practice, describes events that appear randomly and independently on a continuous space, characterized by the occurrence rate (the "intensity"), which is the expected number of cases per unit area.

We stress that in this model with n subregions (postal codes) and L partitions A_l of the overall temporal region, the number of parameters amounts to $3\binom{L}{2} + 3n$ plus those of the slice sampling over Z and W. In our case, both L and n are 10, resulting in a total of 165 parameters. This large number over-parametrizes our model, posing difficulties in the MCMC convergence and enlarging computational times unnecessarily. In consequence, we consider two stages. We first assume that the $Z_{i,t}$ are independent with respect to space, meaning that we estimate each time series separately and take the posterior median of each parameter as the correct value. Once this estimation is done, we then consider the spatial dependence of the $Z_{i,t}$ as defined in equation (20).

We also emphasize the fact that, although the conception of model (17) is of continuous type, the Bayesian implementation requires a discretization. Essentially, stochastic processes become finite-dimensional random vectors and stochastic differential equations become random difference equations. Euler-type discretizations are broadly employed in stochastic computing; they are the simplest convergent methods, with strong convergence of order $1/2$. Milstein scheme, of higher convergence order in general, is not needed for parameter estimation, at least in our context.

On the other hand, we notice that Calatayud et al. (2023b) assumed constant correlations in time, and inference proceeded with the method of moments due to the simplicity of the solution in (13). In Calatayud et al. (2023b), correlations were estimated pairwise, by the underlying multivariate-Gaussian structure. Here, by contrast, the distributions of the parameters are estimated jointly. The downside is that, because of the higher complexity, blocks of time A_l and of space need to be considered.

The MCMC process was conducted over 20,000 iterations, with the initial 5,000 used as a "burn-in" period to ensure convergence. Samples of the parameters were taken at intervals of 50 iterations (thinning to reduce autocorrelation), resulting in a total of 300 values that provide an approximation of the posterior marginal distribution of all parameters involved in the process. We considered non-informative prior distributions for the spatial trends $\mu_i, \mu_{i,j}$ (a Gaussian distribution with a large variance), and informative priors for $\sigma_{i,j}, \Lambda_i(0)$, as follows:

$$\mu_i \sim \mathrm{N}(0,10),$$

$$\sigma_i \sim \mathrm{U}(0.01,0.09),$$
$$\Lambda_i(0) \sim \mathrm{Ga}(10,1),$$
$$\mu_{i,j} \sim \mathrm{N}(0,10),$$
$$\sigma_{i,j} \sim \mathrm{Ga}(1,1),$$
$$y_{i,j}(A_0) \sim \mathrm{Ga}(1,1),$$
$$W_{i,j} \sim \mathrm{N}(0,1),$$

for all postal codes, where N, Ga and U refer to normal, gamma and uniform random variables, respectively. The corresponding informative priors were selected empirically, to reduce the computational time and better adjust the domains of the outputs. We indeed stress that we have consistently employed non-informative priors throughout the modeling process. However, as the number of iterations in the Monte Carlo process increased, we observed that keeping excessively broad parameter domains led to a significant increase in computational cost, and posed challenges to model convergence. In order to improve both the stability and efficiency of the model, we empirically refined the parameter domains as the iterations progressed. This refinement was solely based on the behavior of the model during the initial stages of convergence and was not derived from any subset of the data used later for model fitting or predictive performance evaluation. This approach helped ensure computational feasibility while avoiding any risk of information leakage from the data used in subsequent phases of the analysis.

Figure 4 shows a comparison between the expected number of cases obtained from the posterior median of each parameter and the number of current cases, for some selected cases. Recall that we are working with trend time series. The estimates are quite accurate and capture well the temporal behavior of the observed number of cases over time.

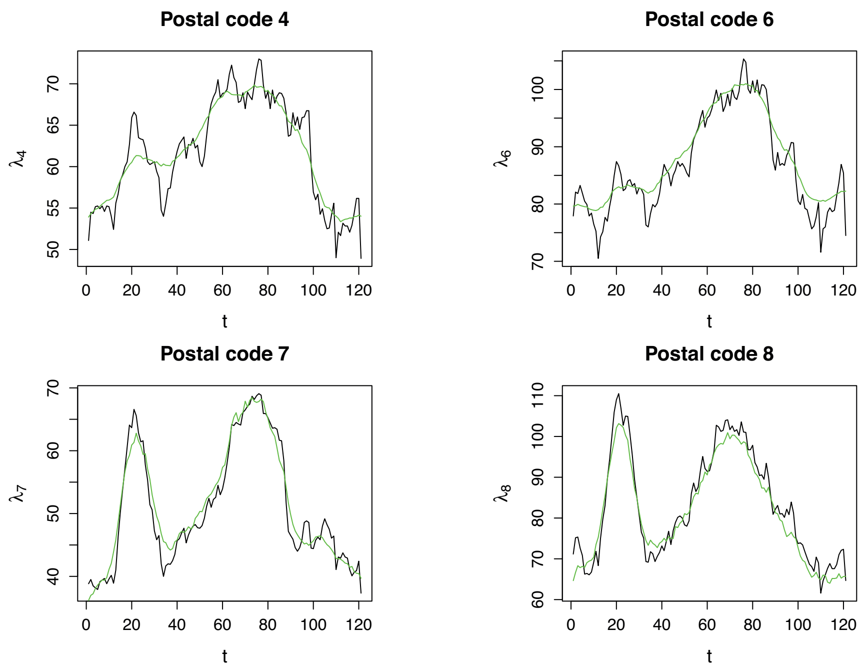

Figure 4. *Comparison between the estimate of cases (green line) and the number of real cases (black line), for postal codes 3, 6, 7 and 8.*

Our model also produces estimates of the correlations between any two postal codes, i.e., up to 45 pairs. To save space, here we focus on the pair $(1,5)$ and show the posterior densities of the parameters associated with the correlation between these two postal codes. Figure 5 depicts such posterior densities, specifically $\mu_{1,5}$, $\sigma_{1,5}$, $y_{1,5}(A_0)$ and $\rho_{1,5,A_0}$, noting that when using non-informative priors, the likelihood is very informative and drives the posterior distribution to the right place, with a shape similar to a bell. We show in the bottom right corner the posterior density of the correlation between postal codes 1 and 5. We obtained this sample by calculating the correlation value with equation (21) employing the corresponding parameter values in each iteration. From this sample, we obtained the estimate for each of the correlations by choosing the posterior median and the credibility intervals which we show in Figure 6, where we observe the behavior of the correlation across the subintervals A_l. We note the temporal variation of such correlations that our model has captured considering them stochastic processes themselves. The uncertainty of the estimates is given by the credible intervals. In general, the zero-correlation value does not lie within the credible region, suggesting a spatial effect in criminal behavior.

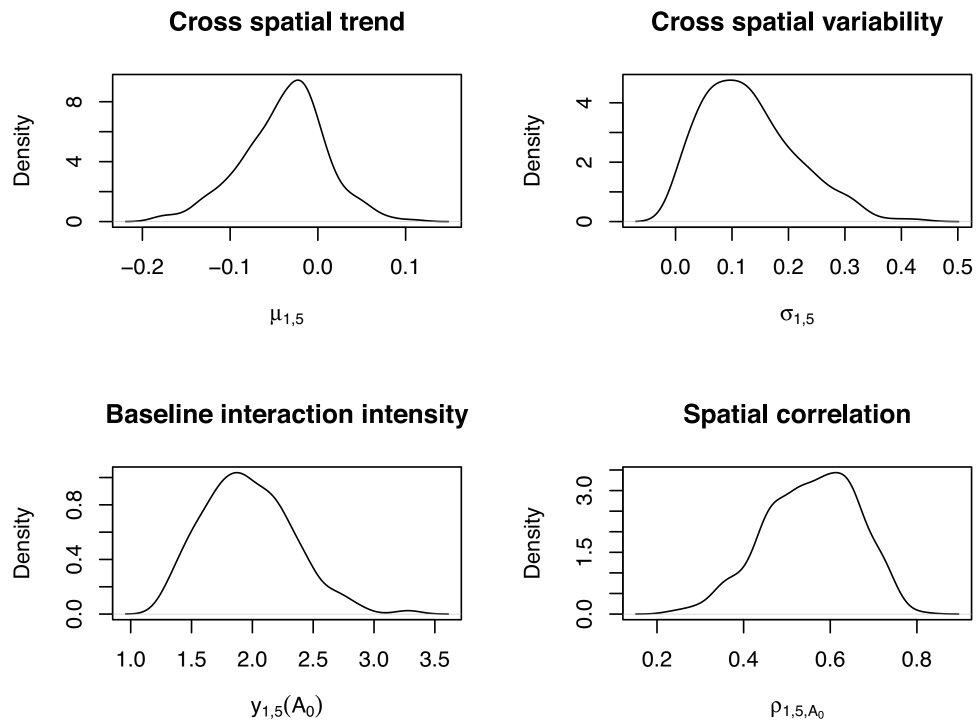

Figure 5. *Posterior densities of the parameters associated with the correlation between postal codes* 1 *and* 5*, as well as the posterior distribution of the correlation in the first subinterval* A_1*. The parameters are defined in equations* (21) *and* (22)*.*

Finally, we assess the accuracy of future predictions on the number of crime cases. We proceed in two ways: complete extrapolation and month-by-month extrapolation. For the former way, we use the estimated model for the first 9 years and predict the expected number for the last year (12 temporal instances), comparing them with the current data. Figure 7 depicts such predictions for the last year in four different postal codes, observing that the black line associated with the number of observed cases is almost always within the 0.95 interval. This is an indication that our proposed model has been accurately estimated and fits the crime data in Valencia well. The credible bands tend to widen because we are extrapolating for a long time window, along a complete year, hence there is a large uncertainty. Mathematically, the behavior is a consequence of the linearly increasing variance of the Brownian motion.

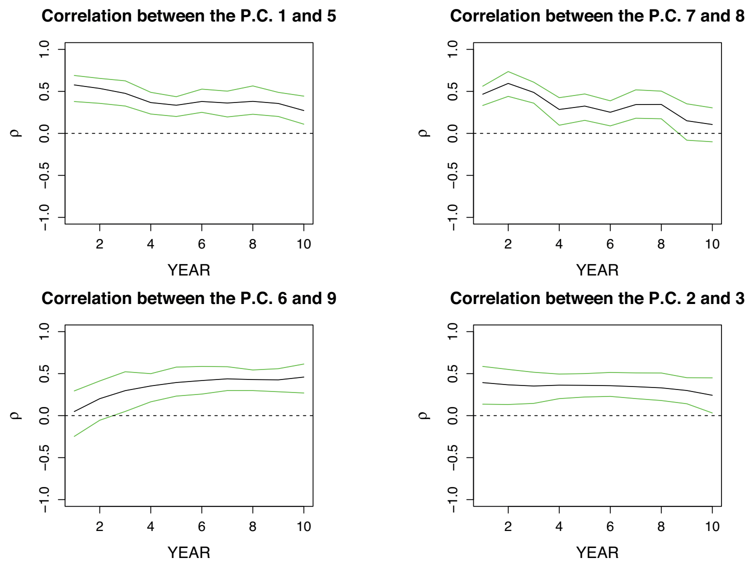

Figure 6. *Estimated pairwise correlations between four selected pairs of different postal codes and their corresponding* 0.95 *credibility intervals (green lines).*

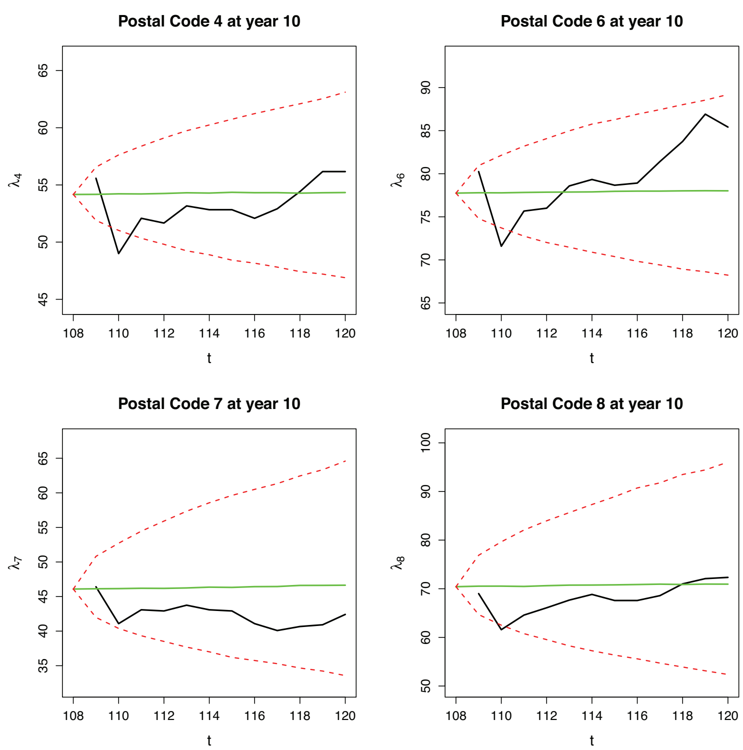

Figure 7. *Complete-one-year predictions (green line) and their credibility intervals (red line) for the last 12 months in 4 different areas, along with the observed cases (black line).*

However, if one forecasts month by month, the magnitude of the credible intervals is that of one time step. This is the case of the second approach: we predict the instant "9 years + ith month" from the data at instants $\leq$ "9 years + $(i-1)$th month", for $i \geq 1$. Figure 8 displays the new forecasts for the last year, this time for all zip codes for generality. The mean values and the credible intervals now move with the data, exhibiting narrower uncertainty. We remark that the last year corresponds to a very complicated scenario, coinciding with the COVID-19 pandemic and the lockdown in Spain (BOE, 2020; Wu et al., 2020). The second black data value of the pictures, that seems to be an outlier, is influenced by March 2020, where the restrictions on movement were imposed and reported criminality suddenly decreased. The fact that our model is successful for such a year makes us think that the proposed methodology is certainly useful for delineating crime dynamics in the short term. We note that the green and red lines advance in parallel, but it seems that there is a one-month lag with the black curve. Such a behavior is linked with the essence of prediction. For example, if there is a decrease of data from $t-1$ to t, this affects the parameter estimation for the times up to t, hence there will probably be a decrease in the prediction on $(t,t+1]$. If data increase from t to $t+1$, then the predicted values will likely rise on $(t+1,t+2]$.

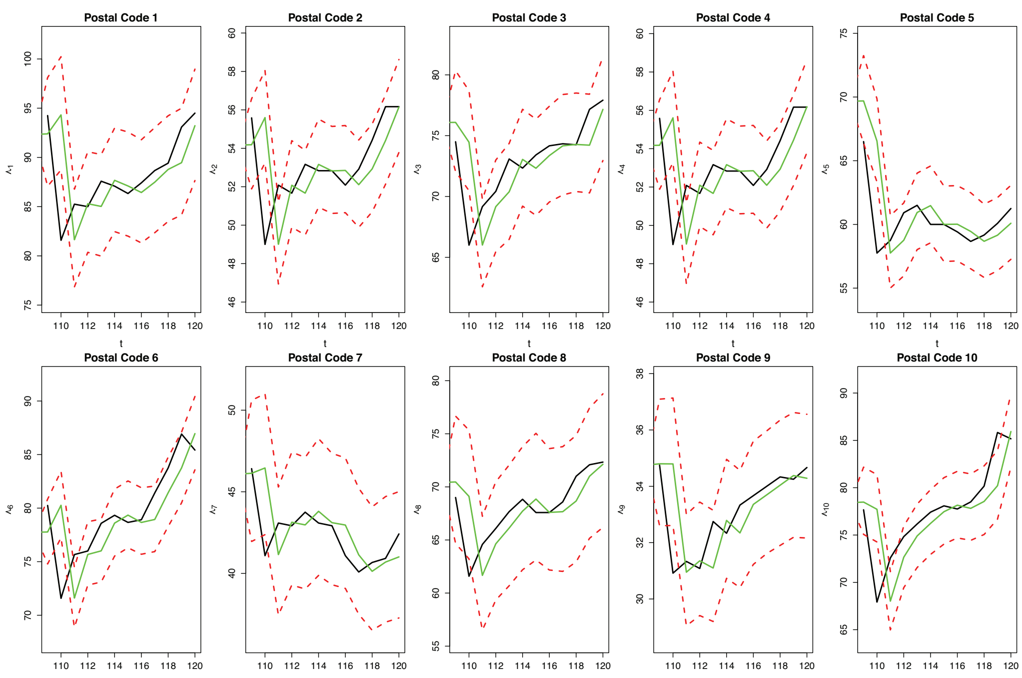

Figure 8. *Month-by-month predictions (green line) and their credibility intervals (red line) for the last 12 months (for year 10), along with the observed cases (black line).*

3.3. Comparison with alternative simpler models

To strengthen our approach, we offer a brief comparison among our model (HDEM), an autoregressive AR(1) structure, and a random forest (RF), as a widely used machine learning method.

Figure 9 presents comparatively credibility intervals (CI) for AR(1), random forest (RF), and HDEM for the case of a one-year prediction for some selected postal codes (3, 4, 6, 10) and year 10. We only show these four cases as an example to reinforce our arguments in favor of our model, since all the cases show similar behavior. We note that in terms of CIs, our HDEM and AR are quite close, suggesting similar levels of precision and reliability in the predicted outcomes. However, the CI of the RF seems to go away from the true values, indicating wrong predictions in all cases. We also considered the case of AR(p) for various orders, and the results were even worse compared to AR(1).

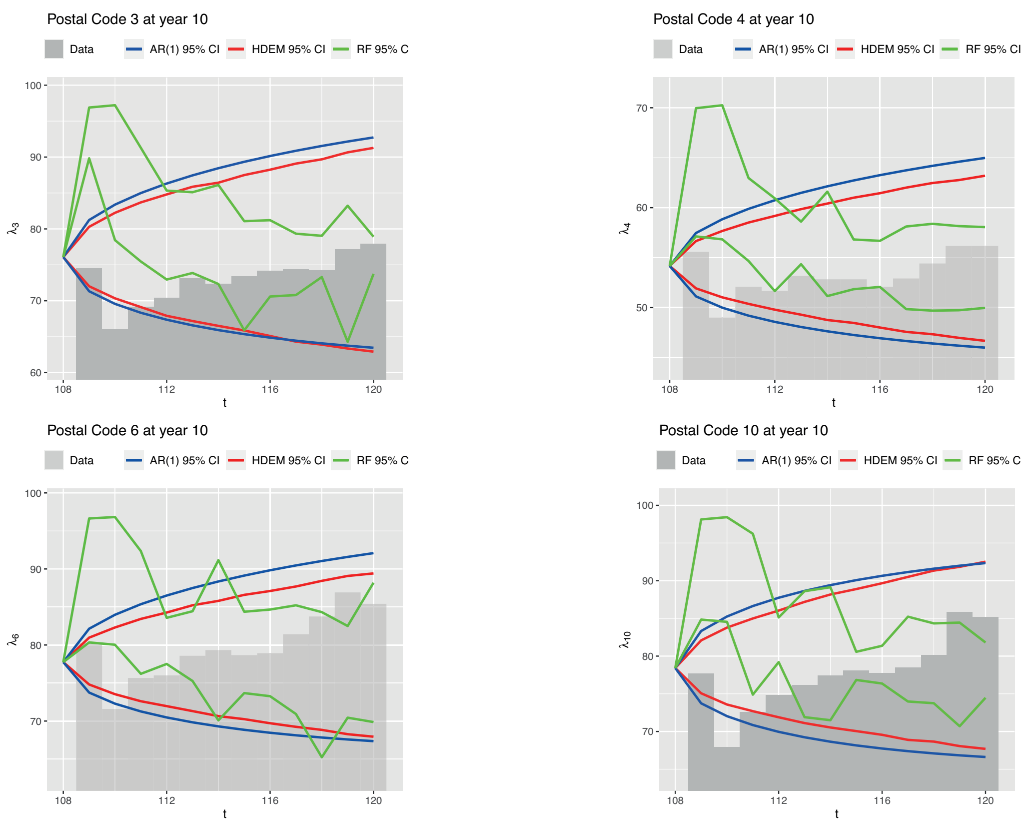

Figure 9. *Comparison between the credibility intervals of the proposed model (HDEM), the AR(1), and the Random Forest (RF) for one-year predictions for year 10 and some selected postal codes (3, 4, 6, 10).*

It is important to stress that one of the key strengths of our model is its ability to account for the spatial correlation between different areas. This feature is crucial for capturing the interactions between neighboring regions, which are often an essential component in the spatial dynamics of crime. Although an AR structure may show strong predictive capabilities in certain contexts, it lacks the spatial interpretability and granularity that our model offers (also true in the case of an RF). This makes our approach particularly valuable for criminologists and law enforcement agencies who want not only predict but also gain a deeper understanding of the spatial patterns and underlying dynamics of crime.

We additionally calculated the root mean squared error (RMSE) of the predictions for each model by comparing the estimates with real values for the last year. By doing this across the 10 zones, we obtain an error for each of them. In Table 1, we show the average error over these 10 zones, as well as the 90% confidence interval of these errors. The results highlight that the RMSE of our model is lower than that of the other two classical models, supporting the better precision and reliability of the HDEM approach over traditional models.

Model	Average error over 10 zip codes	90%
HDEM	15.16359	[7.63, 26.61]
AR(1)	15.38856	[7.97, 26.80]
RF	27.06137	[17.71, 39.59]

Table 1. *Root mean square error (RMSE) for each model based on predictions during the final* 10*th year. The training period spans the previous years (*< 10*).*

4. Discussion and conclusions

We have detailed a methodology to deal with crime data, based on spatio-temporal stochastic differential equations. The data were point-referenced or distributed on patches. From a stochastic partial differential equation for the intensity with temporal log-Gaussian evolution, we obtained stochastic differential equations with correlated Wiener noises. The correlations were stochastic processes too, with a transformed log-Gaussian distribution. We have proposed a hierarchical Bayesian structure for joint inference, and model fitting was conducted for the lattice data in Valencia, which contains a sufficiently large number of events. Time series on the irregular grid were fitted for the past (interpolation) and for the future (prediction, either for a long time window or on an instant-by-instant basis). Our work combined ideas from stochastic differential equations, spatial statistics, time series, and quantitative finance. It is the first attempt to incorporate the use of stochastic partial differential equations in mathematical/statistical criminology with real data, as a direct modeling tool. Other areas of application could be epidemiology or ecology.

Our proposal is very much focused on modeling the dynamics of the stochastic intensity function through differential equations rather than through an underlying Gaussian process. In this way, we can better control the stochastic term through the drift parameter $\mu()$ that captures the growth rate, and the volatility parameter $\sigma()$ that captures the magnitude of fluctuations and the uncertainty of future values. These two parameters are inspired by the evolution of crime incidence and social crime, where there is a rate for the growth of criminality and a volatility for random fluctuations. Note that our approach goes more in the line of Zammit-Mangion et al. (2012), which can be further embedded into an LGCP framework in the context of stochastic integro-difference equations. This is not the usual framework in LGCPs. Classical and widely used LGCPs are based on an underlying Gaussian process (GP) and the second-order properties are inherited from the covariance structure of this GP. Using covariances in space and time is much less flexible, carries a computational burden, and does not model the dynamics of the process as a differential equation does. A final aspect is that with our approach we can model stochastic correlations varying in time, something that becomes very troublesome in classical LGCPs.

The stochastic log-normal model has been of use in finance for stock-price evolution with correlations. The time series of crime studied in this paper have similar fluctuations to those encountered in financial modeling, hence our model building. An important feature of our proposal is that parameter estimation is conducted within the Bayesian paradigm, enabling the delineation of the uncertainty in each involved parameter. We are not aware of similar differential-equation models in criminology, with such a statistical treatment.

Further investigation of this type of models would be of interest when the number of sampling sites or regions of study augments, producing a high increase in the number of parameters. Convergence and computational times are issues to be solved to make these models useful in practice. The case of point patterns, for which locations and times arise randomly, is left for future research. The analysis of point patterns would help to identify crime hotspots within the zip codes. Otherwise, we are constrained to uniformly distributed events within the spatial districts, with the quantity given by the Poisson realization of the aggregate intensity. Finally, there is still work to be done for the inclusion of covariates, probably through link functions on the drift coefficients, to seek higher predictability (with no overfitting): more accurate mean values from a pointwise sense and narrower (less uncertain) probabilistic intervals. Overfitting occurs when models are excessively complex and match the data too perfectly (Duan et al., 2009, Footnote 3), with some parameters that may be unidentifiable (Smith, 2013). Note that in Figure 8 the model does an effective and practical job when predicting month by month for real data, and the incorporation of covariates shall be of use to expand the prediction window.

Funding

This paper has been supported by grants UJI-B2021-37 from Universitat Jaume I and PID2022-141555OB-I00, PID2023-146836NB-I00 from Spanish Ministry of Science.

Data availability statement

The data analyzed in this study are available from the authors upon reasonable request.

Disclosure statement

The authors declare that there is no conflict of interests regarding the publication of this article.

References

Abbas, S., Tripathi, J. P. and Neha, A. A. (2017). Dynamical analysis of a model of social behavior: criminal vs non-criminal population. Chaos Solitons & Fractals, 98, 121–129.

Acedo, L., Díez-Domingo, J., Morano, J. A. and Villanueva R. J. (2010). Mathematical modelling of respiratory syncytial virus (RSV): vaccination strategies and budget applications. Epidemiology & Infection, 138(6), 853–860.

Allen, E. (2007). Modeling With Itô Stochastic Differential Equations. Springer Science & Business Media, Dordrecht, Netherlands.

Banerjee, S., Carlin, B. P. and Gelfand A. E. (2014). Hierarchical Modeling and Analysis for Spatial Data, 2nd Ed. Chapman & Hall/CRC, Boca Raton.

Brauer, F. (2008). Compartmental models in epidemiology. In: Brauer, F., van den Driessche, P. and Wu, J. (eds), Mathematical Epidemiology. Springer, Berlin, Heidelberg.

Braumann, C. A. (2007). Itô versus Stratonovich calculus in random population growth. Mathematical Biosciences, 206(1), 81–107.

Brooks, S., Gelman, A., Jones G. L. and Meng X. L. (2011). Handbook of Markov Chain Monte Carlo, 1st Ed. Chapman & Hall/CRC, Boca Raton.

Burgess, R. L. and Akers, R. L. (1966). A differential association-reinforcement theory of criminal behavior. Social Problems, 14(2), 128–147.

Calatayud, J., Jornet, M. and Mateu, J. (2025a). Spatial modeling of crime dynamics: Patch and reaction-diffusion compartmental systems. Mathematical Methods in the Applied Sciences, 48(7), 7440-7459.

Calatayud, J., Jornet, M. and Mateu, J. (2023a). Modeling noisy time-series data of crime with stochastic differential equations. Stochastic Environmental Research and Risk Assessment, 37, 1053–1066.

Calatayud, J., Jornet, M. and Mateu, J. (2023b). Spatio-temporal stochastic differential equations for crime incidence modeling. Stochastic Environmental Research and Risk Assessment, 37, 1839–1854.

Calatayud, J., Jornet, M. and Mateu, J. (2025b). A dynamical mathematical model for crime evolution based on a compartmental system with interactions. International Journal of Computer Mathematics, 102(1), 44-59.

Cao, Y., Dong, K., Siercke, B. and Wilber, M. (2013). Final Report: Crime Modeling. UCLA, LA, USA. https://www.math.ucla.edu/~bertozzi/WORKFORCE/REU%202013/Crime%20Fighters%20%5BGroup%20Awesome%5D/FinalReport.pdf. Last accessed 14 July 2023.

Cervelló, R., Cortés, J. C., Santonja, F. J. and Villanueva R. J. (2014). The dynamics over the next few years of the Spanish mobile telecommunications market share: a mathematical modelling approach. Mathematical and Computer Modelling of Dynamical Systems, 20(6), 557–565.

Davison, A.C. and Hinkley, D.V. (1997). Bootstrap Methods and Their Application. Cambridge University Press, London.

Dorini, F. A. and Sampaio, R. (2012). Some results on the random wear coefficient of the Archard model. Journal of Applied Mechanics, 79(5), 051008–051014.

Duan, J. A., Gelfand, A. E. and Sirmans, C.F. (2009). Modeling space-time data using stochastic differential equations. Bayesian Analysis, 4(4), 733–758.

Esiri, M. O. (2016). The influence of peer pressure on criminal behaviour. Journal of Humanities and Social Science, 21(1), 08–14.

Evans, L. C. (2012). An Introduction to Stochastic Differential Equations. American Mathematical Society, New York.

Gelfand, A. E., Diggle, P. J., Fuentes, M. and Guttorp, P. (2010). The Handbook of Spatial Statistics. Chapman Hall, Boca Raton.

González-Parra, G., Chen-Charpentier, B. and Kojouharov, H. V. (2018). Mathematical modeling of crime as a social epidemic. Journal of Interdisciplinary Mathematics, 21(3), 623–643.

Harkins, S. G., Williams, K. D. and Burger, J. (2017). The Oxford Handbook of Social Influence. Oxford University Press, UK.

Jane White, K. A., Campillo-Funollet, E., Nyabadza, F., Cusseddu, D., Kasumo, C., Imbusi, N. M., Ogesa Juma, V., Meir, A. J., Marijani, T. (2021). Towards understanding crime dynamics in a heterogeneous environment: A mathematical approach. Journal of Interdisciplinary Mathematics, 24(8), 2139–2159.

Kolokolnikov, T., Lloyd, D. J. B. and Short M. (2019). Mathematical Criminology and Security. Workshop, Banff International Research Station, Banff, Canada, 17th March 2019 – 22nd March 2019, pp. 1–7.

Kolokolnikov, T., Ward, M. J. and Wei, J. (2014). The stability of steady-state hot-spot patterns for a reaction-diffusion model of urban crime. Discrete and Continuous Dynamical Systems - B, 19(5), 1373–1410.

Koss, L. (2019). SIR models: differential equations that support the common good. CODEE Journal, 12(1), article 6.

Lacey, A. A. and Tsardakas, M. N. (2016). A mathematical model of serious and minor criminal activity. European Journal of Applied Mathematics, 27(3), 403–421.

Lamberton, D. and Lapeyre, B. (2011). Introduction to Stochastic Calculus Applied to Finance, 2nd Ed. Chapman & Hall / CRC press, London.

Lesaffre, E. and Lawson, A. B. (2012). Bayesian Biostatistics. Wiley, New York.

Lloyd, D. J. and O'Farrell, H. (2013). On localised hotspots of an urban crime model. Physica D: Nonlinear Phenomena, 253, 23–39.

Lord, G. J., Powell, C. E. and Shardlow, T. (2014). An Introduction to Computational Stochastic PDEs. Cambridge Texts in Applied Mathematics, Cambridge University Press, New York.

Mao, X. (2007). Stochastic Differential Equations and Applications. Elsevier, Cambridge.

McMillon, D., Simon, C. P. and Morenoff, J. (2014). Modeling the underlying dynamics of the spread of crime. PLoS ONE, 9(4), e88923.

Murray, J. D. (2002). Mathematical Biology I. 3rd Ed. Springer, New York.

Neal, R. M. (2003). Slice sampling. Annals of Statistics, 31(3), 705–767.

R Core Team. (2023). R: A language and environment for statistical computing, Version 4.3.2. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rackauckas, C. (2014). An Intuitive Introduction For Understanding and Solving Stochastic Differential Equations. Course notes.

Rodriguez, N. and Winkler, M. (2022). On the global existence and qualitative behaviour of one-dimensional solutions to a model for urban crime. European Journal of Applied Mathematics, 33(5), 919–959.

Santonja, F. J., Sánchez, E., Rubio, M. and Morera, J. L. (2010). Alcohol consumption in Spain and its economic cost: a mathematical modeling approach. Mathematical and Computer Modelling, 52(7–8), 999–1003.

Schiesser, W. E. (2019). Spatiotemporal Modeling of Influenza. Partial Differential Equation Analysis in R. Synthesis Lectures on Biomedical Engineering. Morgan & Claypool Publishers, USA.

Short, M. B., d'Orsogna, M. R., Pasour, V. B., Tita, G. E., Brantingham, P. J., Bertozzi, A. L. and Chayes, L. B. (2008). A statistical model of criminal behavior. Mathematical Models and Methods in Applied Sciences, 18(01), 1249–1267.

Short, M., Bertozzi, A. and Brantingham, P. (2010a). Nonlinear patterns in urban crime: Hotspots, bifurcations, and suppression. SIAM Journal on Applied Dynamical Systems, 9(2), 462–483.

Short, M. B., Brantingham, P. J., Bertozzi, A. L., Tita, G. E. (2010b). Dissipation and displacement of hotspots in reaction-diffusion models of crime. Proceedings of the National Academy of Sciences, 107(9), 3961–3965.

Smith, R.C. (2013) Uncertainty Quantification: Theory, Implementation, and Applications. SIAM, Philadelphia.

Song, B., Castillo-Garsow, M., Ros-Soto, K. R., Mejran, M., Henso, L. and Castillo-Chávez, C. (2006). Raves, clubs and ecstasy: the impact of peer pressure. Mathematical Biosciences and Engineering, 3(1), 249–266.

Sooknanan, J. and Comissiong, D.M. (2017). When behaviour turns contagious: the use of deterministic epidemiological models in modeling social contagion phenomena. International Journal of Dynamics and Control, 5(4), 1046–1050.

Spanish Official State Bulletin (BOE), number 67, decree 463/2020, reference BOE-A-2020-3692, 14th March 2020. https://www.boe.es/eli/es/rd/2020/03/14/463/con

Srivastav, A. K., Athithan, S. and Ghosh, M. (2020). Modeling and analysis of crime prediction and prevention. Social Network Analysis and Mining, 10(1), 1–21.

Teng, L., Ehrhardt, M. and Günther, M. (2016). Modelling stochastic correlation. Journal of Mathematics in Industry, 6, 1–18.

Tse, W.H. and Ward, M.J. (2016). Hotspot formation and dynamics for a continuum model of urban crime. European Journal of Applied Mathematics, 27(3), 583–624.

van den Driessche, P. (2008). Compartmental models in epidemiology. In: Brauer, F., van den Driessche, P. and Wu, J., (eds), Mathematical Epidemiology. Springer, Berlin, Heidelberg.

van Emmerich, C. (2006). Modelling correlation as a stochastic process. Preprint 06/03, University of Wuppertal.

Voit, J. (2010). The Statistical Mechanics of Financial Markets, 3rd Ed. Springer, Berlin, Heidelberg.

Wang, Q., Wang, D. and Feng, Y. (2020). Global well-posedness and uniform boundedness of urban crime models: One-dimensional case. Journal Of Differential Equations, 269(7), 6216–6235.

White, E. and Comiskey, C. (2007). Heroin epidemics, treatment and ODE modeling. Mathematical Biosciences, 208(1), 312–324.

Wolfram Research, Inc. (2020). Mathematica, Version 12.1. Champaign, IL.

Wu, J. (2008). Spatial structure: Partial differential equations models. In: Brauer, F., van den Driessche, P. and Wu, J. (eds), Mathematical Epidemiology. Springer, Berlin, Heidelberg.

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C. and Zhang, Y. L. (2020). A new coronavirus associated with human respiratory disease in China. Nature, 579(265), 265–269.

Xiu, D. (2010). Numerical Methods for Stochastic Computations: A Spectral Method Approach. Cambridge Texts in Applied Mathematics. Princeton University Press, New York.

Zammit-Mangion, A., Dewar, M., Kadirkamanathan, V. and Sanguinetti, G. (2012). Point process modelling of the Afghan War Diary. Proceedings of the National Academy of Sciences, 109(31), 12414–12419.

Zammit-Mangion, A. and Wikle, C.K. (2020). Deep integro-difference equation models for spatio-temporal forecasting. Spatial Statistics, 37, 100408.