Modeling noisy time-series data of crime with stochastic differential equations

Julia Calatayud [a], Marc Jornet [b], Jorge Mateu [c]

[a] Departament de Matemàtiques, Universitat Jaume I, 12071 Castellón, Spain.
email: calatayj@uji.es
ORCID: 0000-0002-9639-1530

[b] Departament de Matemàtiques, Universitat de València, 46100 Burjassot, Spain.
email: marc.jornet@uv.es
ORCID: 0000-0003-0748-3730

[c] Departament de Matemàtiques, Universitat Jaume I, 12071 Castellón, Spain.
email: mateu@uji.es
ORCID: 0000-0002-2868-7604

Abstract. We develop and calibrate stochastic continuous models that capture crime dynamics in the city of Valencia, Spain. From the emergency phone, data corresponding to three crime events, aggressions, stealing and women alarms, are available from the year 2010 until 2020. As the resulting time series, with monthly counts, are highly noisy, we decompose them into trend and seasonality parts. The former is modeled by geometric Brownian motions, both uncorrelated and correlated, and the latter is accommodated by randomly perturbed sine-cosine waves. Albeit simple, the models exhibit high ability to simulate the real data and show promising for crimes-interaction identification and short-term predictive policing.

Keywords: Crime-incidence assessment; Trend and seasonality; Stochastic differential equation; Inverse problem; Simulations; Correlated data

AMS Classification 2010: 60H10; 34F05; 62F10; 62P25

1. Introduction

Criminality is a serious problem for any region, which risks its economy, security and quality of life. In the field of mathematical modeling, the study of crime events from the point of view of differential equations has been developed in several directions. On the one hand, with partial differential equations, space locations are characterized by a potential of criminal activity, taking into account feasibility, attractiveness, opportunities, and knowledge of offenders about target, vulnerability, victims, area, etc.; the main objective is the study of the dynamics of crime hotspots [4, 10, 12, 16, 24–26, 30]. On the other hand, ordinary differential equations coupled through population compartments provide the mechanisms for the flow and the social transmission between criminality states [1, 9, 18, 20, 27, 28]. Albeit these theories are powerful to get a deeper understanding of crime patterns, fitting the models to actual crime data is not straightforward and therefore their applicability is lessened. In fact, to our knowledge, only two differential equation-based works overtake qualitative aspects and attempt to calibrate parameters to match model output and recorded observations, see [11, 13]. In paper [13], the authors consider serious and minor criminal activities in Manchester, which are influenced by the attractiveness of the place at each time instant, and set a system of ordinary differential equations for model fitting. However, the performance is limited, since the parameters seem to be unidentifiable and the inverse problem is challenging and not uniquely solvable. Further, although a stochastic model is proposed, it is not fully calibrated. Article [11], for its part, models criminality data in an area of South Africa, by dividing the region into high- and low-conflicting zones. A system of two ordinary differential equations is proposed, by assuming certain behavioral and spatial fluxes. However, it is not clear how to divide the area of study in general. Also, data are aggregated on an annual basis, so noisy patterns do not arise and nearly linear models make a very good job at replicating the observations with no need of stochastic effects.

In our paper, we intend to model time series of crime in a city of Spain, Valencia, by dealing with highly noisy patterns and calibrating stochastic effects. In this manner, we seek to supplement the interesting cases investigated in [11, 13]. To provide context, Valencia is a city located in the Mediterranean coast, with 800,000 inhabitants. Even though it is a safe place, it is a major city in Spain and several illegal acts may occur per day. When suffered or witnessed, these activities are communicated to the 112-emergency phone. For the design of the paper, we have access to a list of crime events in the streets of Valencia from 2010 until 2020: aggression (theft with violence), stealing (theft with no violence), women alarms (attack to a woman with violence), and others. Our main goal is the proposal and calibration of stochastic differential equation models that can capture the trends of the crimes and quantify their uncertainties [2, 14, 17], by using standard models from the financial literature on stock price evolution. The ability of our simple stochastic equations to simulate the real data suggests a new view of crime-dynamics modeling. Ideally, for real-world applications seeking predictability by the police, short training periods may be employed for calibration and then forecast a few subsequent times ("predictive policing").

The structure of the paper is as follows. In Section 2, data are presented and decomposed to capture a trend and seasonality. Methods are proposed to model trend by uncorrelated or correlated Itô diffusion, and seasonality. Numerical results for each methodology are reported in Section 3, with tabulated calibrations and graphed model outputs. Section 4 is devoted to the discussion of the main aspects of the paper and a detailed comparison with the literature. Finally, in Section 5, conclusions are drawn.

2. Methods

In this section, we describe the methods followed in the analysis of crime data. After presenting the data, we extract its components for simplification. Then we develop stochastic models that can well capture the new time series.

2.1. Data.

Our dataset contains information about reported criminal events in the city of Valencia for ten complete years, from 2010 to 2020. We have a total of 90247 events communicated to the 112-emergency phone, split into *aggression* (55610 cases), *stealing* (25342 cases), *woman alarm* (454 cases) and *others* (8841 cases). These four categories refer to different types of thefts or robberies in the streets: *aggression* means a theft after hitting a person, *stealing* is a smooth theft with no force used, *woman alarm* is a theft to a woman with violence, and *others* means other thefts or robberies that cannot be considered within the previous three groups. This last category is formed by several events with different types of structures, making it highly variable and difficult to model; thus, we focus on the other three categories.

In Figure 1, we present the data on aggressions, stealing and women alarms. We employ monthly observations, along 132 months. Observe that the time series are very noisy, with some sort of white noise pattern. This motivates the separation of the series into trend (with an Itô-diffusion pattern) and seasonality (with some noisy bias).

2.2. Trend and seasonality.

The time series are split into two components: trend and seasonality. The trend captures the general pattern of the data over time; we obtain it by using a moving average of twelve months (months of periodicity) to smooth out the original time series. On the other hand, seasonality captures periodic patterns over time, in this case annual; we

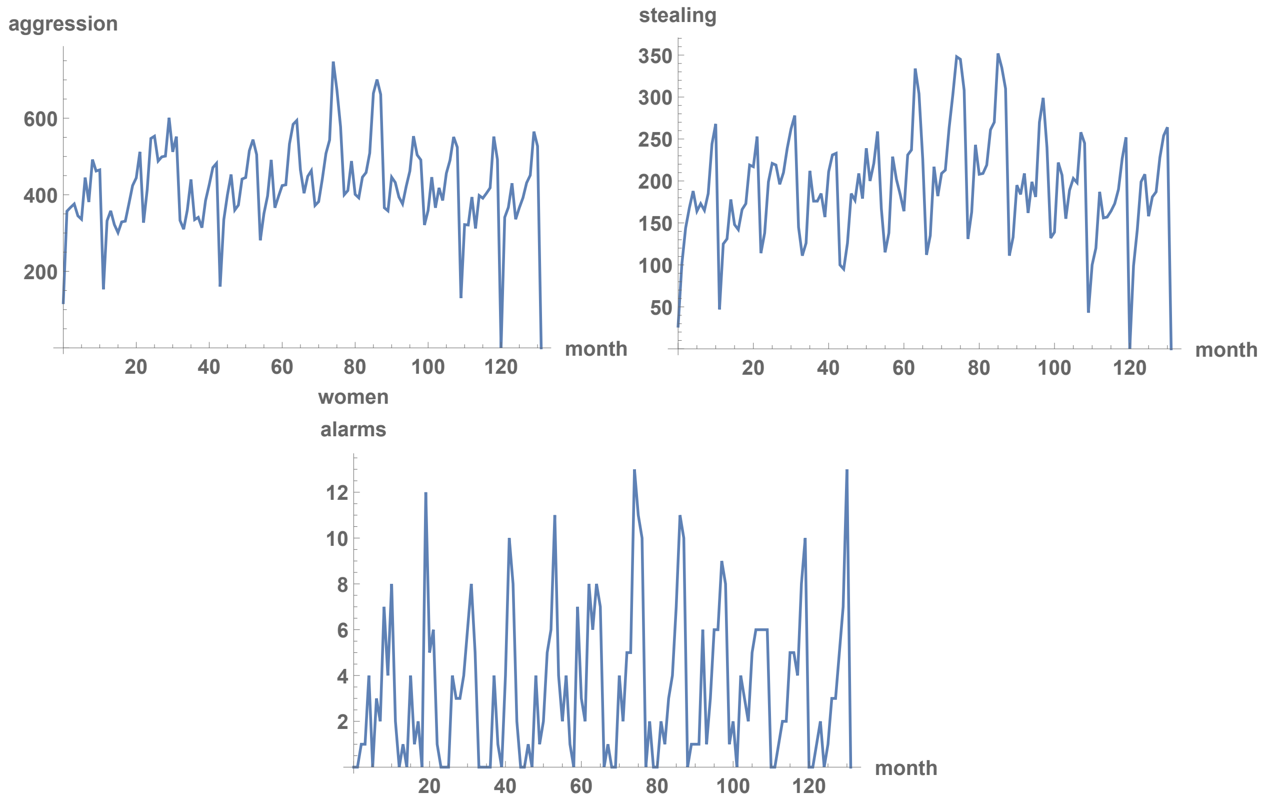

FIGURE 1. Monthly counting of aggressions, stealing and women alarms in the city of Valencia, from January 2010 to December 2020. Source: 112-emergency phone.

obtain it by subtracting the original data and the trend. Both components present noise, due to the inherent uncertainty in the phenomenon and in data collection.

In Figures 2 and 3, we present the data trend and seasonality, respectively. We see that, approximately, the trends increase until summer 2011, decrease until the beginning of 2013, and then augment until a spike at mid 2016, to later show a falling pattern up to December 2020. The three criminal events have a similar evolution, although their incidences are quite different: aggressions double stealing incidents, while women alarms are seldom reported. On the other hand, we observe distinct yearly upward spikes in the seasonality time series.

Due to the smoothing of the original noise and the fluctuations observed in Figure 2, we attempt to describe the trend by an Itô-diffusion process, rather than a white noise process. Specifically, as in the financial literature of stock price evolution, we employ a geometric Brownian motion process to fit the data trend. The seasonality, by contrast, will be given a noise complementing a deterministic Fourier series.

2.3. **Modeling of trend with a geometric Brownian motion.** Given any of the three trends, to be described by x_t = modeled value of the real trend at instant t, we start with the ordinary differential equation model

$$x'_t = \mu x_t, \tag{2.1}$$

where the prime denotes the derivative with respect to time. Parameter $\mu \in \mathbb{R}$ may be interpreted as the instantaneous relative risk of criminality. It is assumed to be constant over time. However, life is inherently uncertain, and there are certainly random factors that may affect the risk along time. Thus, parameter μ is perturbed through a Gaussian white noise process

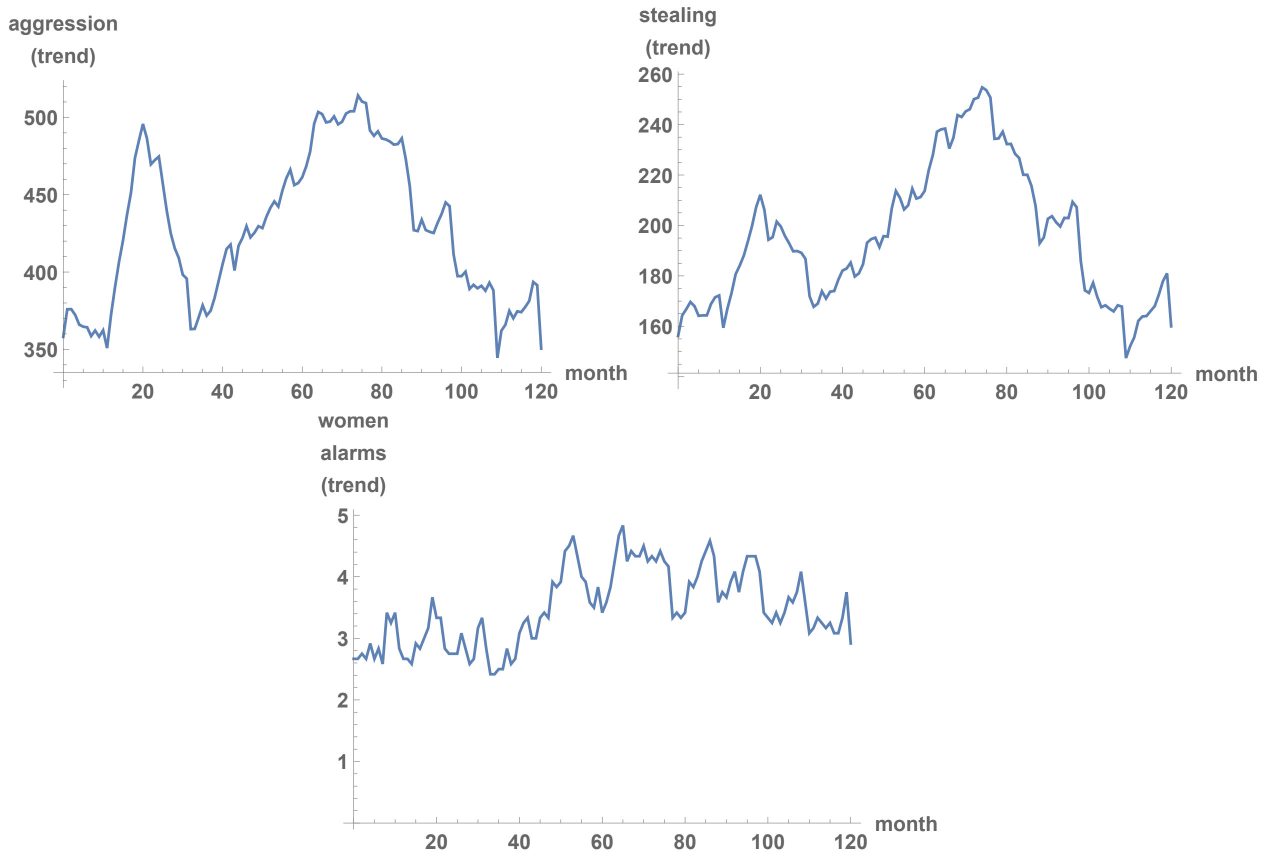

FIGURE 2. Trend component of aggressions, stealing and women alarms in the city of Valencia. The raw data sets were smoothed by using a moving window average.

with intensity (magnitude) $\sigma > 0$:

$$\mu \leftarrow \mu + \sigma B_t'.$$

The noise B_t', uncorrelated with infinite variance and zero mean, is the formal derivative of a standard Brownian motion, or Wiener process, B_t. This Brownian motion has the properties of zero mean and covariance given by the minimum of the two time instants; its trajectories are continuous but nowhere differentiable or monotone. Since B_t is nowhere differentiable, the white noise B_t' is idealized and its properties are derived from merely formal calculations; actually, B_t' is only well-defined as a Schwartz distribution or generalized process. The model (2.1) for the trend becomes a stochastic differential equation

$$x_t' = \mu x_t + \sigma x_t B_t'. \tag{2.2}$$

The white noise is multiplied by the population, so that both are proportional; greater oscillations occur when there are higher rates of crimes. In differential notation, the model (2.2) is

$$\mathrm{d}x_t = \mu x_t \,\mathrm{d}t + \sigma x_t \,\mathrm{d}B_t, \tag{2.3}$$

which is interpreted in integral form under the theory of Itô calculus. Another viewpoint for the Itô stochastic differential equation (2.3) is the continuous limit of the discrete system

$$\Delta x_t = \mu x_t \,\Delta t + \sigma x_t \sqrt{\Delta t}\, Z_t,$$

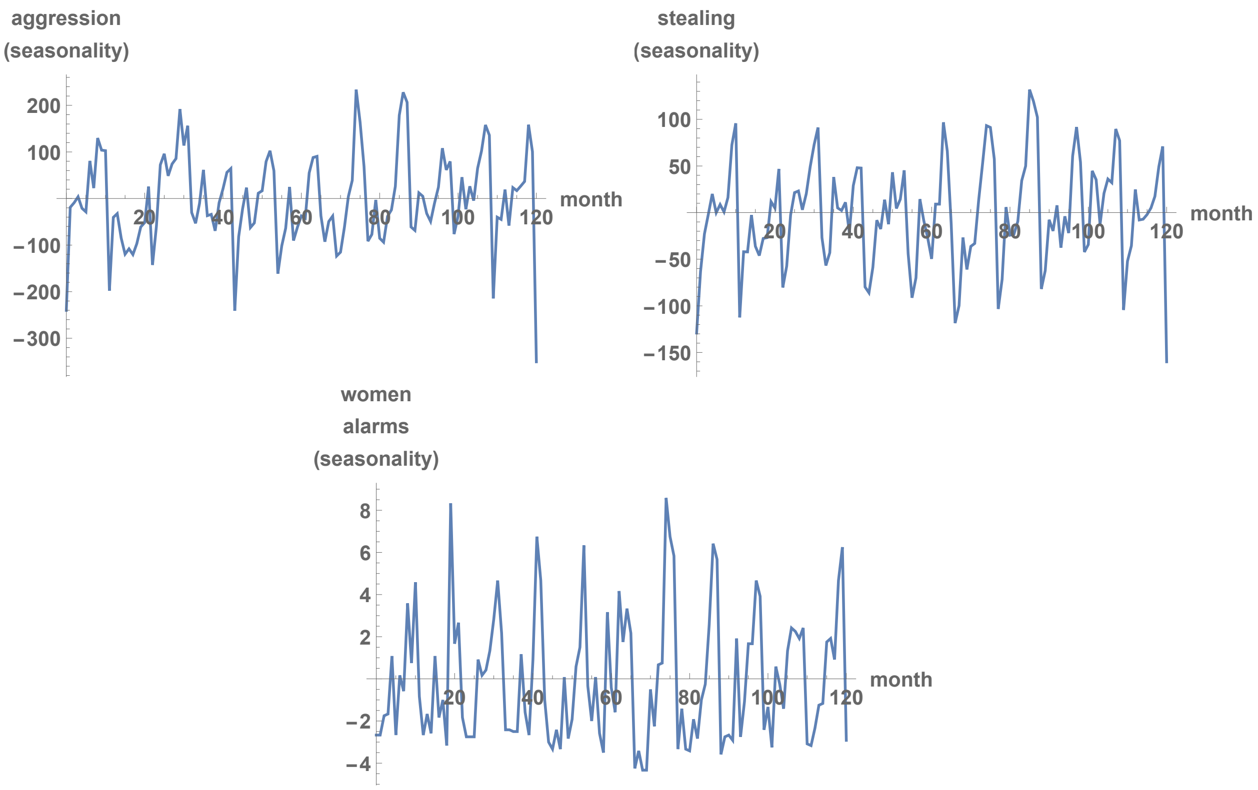

FIGURE 3. Seasonality component of aggressions, stealing and women alarms in the city of Valencia. The trends were extracted from the raw data sets.

given fine partitions, where $Z_t \sim \text{Normal}(0,1)$ is an uncorrelated process. Now x_t is a stochastic process, called geometric Brownian motion. By Itô lemma, which extends the standard chain rule theorem for non-differentiable processes, the solution to (2.3) is given by

$$x_t = x_0 \mathrm{e}^{(\mu - \frac{1}{2}\sigma^2)t + \sigma B_t}, \tag{2.4}$$

where $x_0 > 0$ is the initial, deterministic state. Interestingly, the expected value of x_t coincides with the solution to the deterministic model. The stochastic solution serves to indicate random variability and is qualitatively closer to data. Its trajectories are positive and continuous but nowhere differentiable or monotone.

We fit the real trend time series $\{s_t\}_{t \geq 0}$ at times $0 < t_1 < t_2 < \ldots$, by matching s_t and the model (2.4) x_t and calibrating μ and σ. The simplest method to derive estimates of these two parameters is based on statistical moments. By using (Napierian) log-returns $u_t = \log s_t - \log s_{t-1}$, and by equating the sample mean and variance, $\overline{u}$ and d^2 respectively, to the distributional mean and variance, the estimates obtained are

$$\hat{\mu} = \frac{\overline{u} + d^2/2}{\Delta t}, \quad \hat{\sigma} = \frac{d}{\sqrt{\Delta t}}. \tag{2.5}$$

We will consider times $0 < 1 < 2 < \ldots$ and $\Delta t = 1$. As will be perceived, the realizations of the geometric Brownian motion (2.4) will mimic the trends qualitatively, which justifies the use of stochastic differential equations of Itô type.

2.4. **Modeling of seasonality with Fourier series and noise.** In this part, sine-cosine waves are used to accommodate the seasonal pattern of crimes. Unspecified features of each month are represented by a random effect.

Seasonality is modeled through a truncated Fourier series of period 12 plus a noise,

$$y_t = \frac{a_0}{2} + \sum_{k=1}^{K} \left(a_k \cos\left(\frac{2k\pi t}{12}\right) + b_k \sin\left(\frac{2k\pi t}{12}\right)\right) + \epsilon_t, \tag{2.6}$$

where $\epsilon_t \sim \text{Normal}(0, \sigma)$ is an uncorrelated process with homogeneous variance σ^2 (distinct from the trend case).

The Fourier coefficients $a_0, a_1, \ldots, a_K, b_1, \ldots, b_K$ in (2.6) are estimated by least-squares minimization from the seasonality time series. Since the problem is linear with respect to the coefficients, there is one best-fit solution. The standard deviation σ is then simply estimated from the standard deviation of the residuals sample.

2.5. **Modeling of correlated trends with correlated geometric Brownian motions.** In Figure 2, one notes that time series exhibit cross-correlation. For example, the evolution patterns of aggressions and stealing are similar, as both may be viewed as serious and minor acts of the same criminal activity. Thus, instead of working with independent geometric Brownian motion processes, one may consider certain dependencies. Given two ordinary differential equations

$$x'_{1,t} = \mu_1 x_{1,t}$$

and

$$x'_{2,t} = \mu_2 x_{2,t}$$

for the trend of aggressions and stealing, respectively, the parameters are perturbed as

$$\mu_1 \leftarrow \mu_1 + \sigma_1 B'_{1,t}$$

and

$$\mu_2 \leftarrow \mu_2 + \sigma_2 B'_{2,t},$$

where $B_{1,t}$ and $B_{2,t}$ are correlated Brownian motions and $\sigma_1, \sigma_2 > 0$ are the intensities (magnitudes) of the noises. Indeed, the random factors that may affect the risk of aggression or stealing are not entirely independent. To build the two correlated Brownian motions, one starts with a Brownian process $B_{1,t}$ and then defines

$$B_{2,t} = \rho B_{1,t} + \sqrt{1-\rho^2}\, B_{3,t},$$

where $B_{3,t}$ is an auxiliary Brownian motion that is independent of $B_{1,t}$. Parameter ρ is the resulting correlation between $B_{1,t}$ and $B_{2,t}$, which is homogeneous in time:

$$\begin{aligned} \text{cov}[B_{1,t}, B_{2,t}] &= \text{cov}[B_{1,t}, \rho B_{1,t} + \sqrt{1-\rho^2}\, B_{3,t}] \\ &= \rho\, \text{cov}[B_{1,t}, B_{1,t}] + \sqrt{1-\rho^2}\, \text{cov}[B_{1,t}, B_{3,t}] \\ &= \rho\, t \end{aligned}$$

and

$$\text{corr}[B_{1,t}, B_{2,t}] = \frac{\text{cov}[B_{1,t}, B_{2,t}]}{\sqrt{\text{var}[B_{1,t}]\text{var}[B_{2,t}]}} = \frac{\rho\, t}{\sqrt{t \cdot t}} = \rho.$$

In differential form, the models for both trends are

$$\begin{aligned} \mathrm{d}x_{1,t} &= \mu_1 x_{1,t}\, \mathrm{d}t + \sigma_1 x_{1,t}\, \mathrm{d}B_{1,t}, \\ \mathrm{d}x_{2,t} &= \mu_2 x_{2,t}\, \mathrm{d}t + \sigma_2 x_{2,t}\, \mathrm{d}B_{2,t}. \end{aligned}$$

Itô lemma yields the solutions

$$x_{1,t} = x_{1,0} \mathrm{e}^{(\mu_1 - \frac{1}{2}\sigma_1^2)t + \sigma_1 B_{1,t}}, \tag{2.7}$$

$$x_{2,t} = x_{2,0}\mathrm{e}^{(\mu_2 - \frac{1}{2}\sigma_2^2)t + \sigma_2 B_{2,t}} = x_{2,0}\mathrm{e}^{(\mu_2 - \frac{1}{2}\sigma_2^2)t + \sigma_2\rho B_{1,t} + \sigma_2\sqrt{1-\rho^2}\, B_{3,t}}. \tag{2.8}$$

To estimate the five parameters μ_1, μ_2, σ_1, σ_2 and ρ in (2.7) and (2.8), log-returns are considered. If $\{s_{1,t}\}_{t\geq 0}$ and $\{s_{2,t}\}_{t\geq 0}$ denote the real trend time series at time instants $0 < 1 < 2 < \ldots$, with $\Delta t = 1$, the log-returns $u_{1,t} = \log s_{1,t} - \log s_{1,t-1}$ and $u_{2,t} = \log s_{2,t} - \log s_{2,t-1}$ are considered. The method of moments is used. By equating the sample means and variances, $\overline{u}_1$, $\overline{u}_2$, d_1^2 and d_2^2 respectively, to the distributional means and variances, the estimates obtained are

$$\hat{\mu}_1 = \frac{\overline{u}_1 + d_1^2/2}{\Delta t}, \quad \hat{\sigma}_1 = \frac{d_1}{\sqrt{\Delta t}}, \tag{2.9}$$

$$\hat{\mu}_2 = \frac{\overline{u}_2 + d_2^2/2}{\Delta t}, \quad \hat{\sigma}_2 = \frac{d_2}{\sqrt{\Delta t}}. \tag{2.10}$$

These values coincide with those in the case of no correlation, see (2.5). This is an important feature of our approach for dealing with cross-correlation; since interactions arise from the noises' correlation ρ only, the estimates for the remaining parameters do not change. The estimate for the correlation between the two Brownian motions is

$$\hat{\rho} = \frac{d_{1,2}}{\hat{\sigma}_1\hat{\sigma}_2\Delta t}, \tag{2.11}$$

where $d_{1,2}$ is the sample covariance between $\{u_{1,t}\}_t$ and $\{u_{2,t}\}_t$. When $\hat{\rho} \neq 0$, we are identifying interaction between the two crimes.

3. Results

In this section, we describe the main results obtained in the analysis of the crime data. Specifically, trend time series modeled by uncorrelated and correlated geometric Brownian motions, and seasonality time series modeled by truncated Fourier series with random effects. We use the software Mathematica® [31].

3.1. **Fitting of trend with a geometric Brownian motion.** In Figures 4 and 5, we show how geometric Brownian motion (2.4) accommodates the aggression trend. In both plots, the mean and a 0.95 probabilistic interval are represented. Recall that the mean is the curve of a deterministic exponential model, (2.1). The interval gathers the trajectories and becomes wider as time passes, by the linear increase of the variance of Brownian motion with time; indeed, as we move away from the initial condition, the uncertainty in the output estimation raises. In Figure 4, two realizations of (2.4) are depicted as an example, which mimic the fluctuations of the trend qualitatively. In Figure 5, the optimal path among an ensemble of 10^5 trajectories of (2.4) is drawn, which provides a good fit of the time series quantitatively. The optimal path, say x_t^{opt}, minimizes the sum of the squared differences between the simulated values x_t and the trend data s_t:

$$x^{\mathrm{opt}} = \underset{10^5 \text{ trajectories } x}{\operatorname{argmin}} \sum_{\text{all } t} (x_t - s_t)^2 . \tag{3.1}$$

The capture of fluctuations would be impossible with deterministic formulations. As the number of runs (i.e. simulated trajectories of (2.4)) increases, it is expected that the least-squares optimal path shows less discrepancy and a better overlap with respect to the trend time series because the ensemble is larger.

For the events of stealing and women alarms, analogous figures are presented. In Figures 6 and 7, we show the fit of the stealing trend. In Figures 8 and 9, the trend of women alarms is

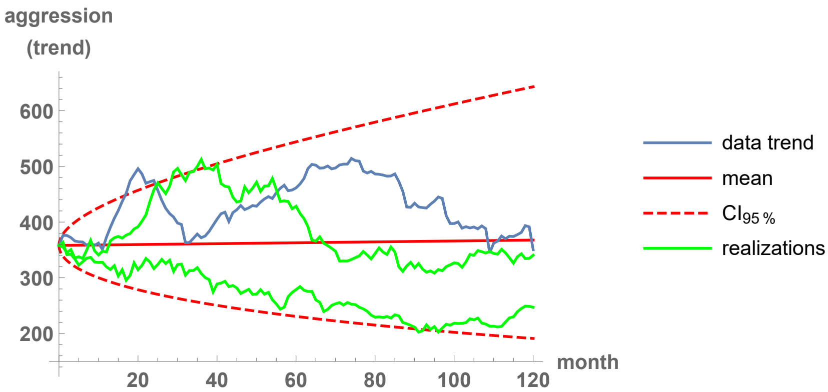

FIGURE 4. Trend-component fitting of aggressions in the city of Valencia. Mean, 0.95 probabilistic interval, and two realizations as an example.

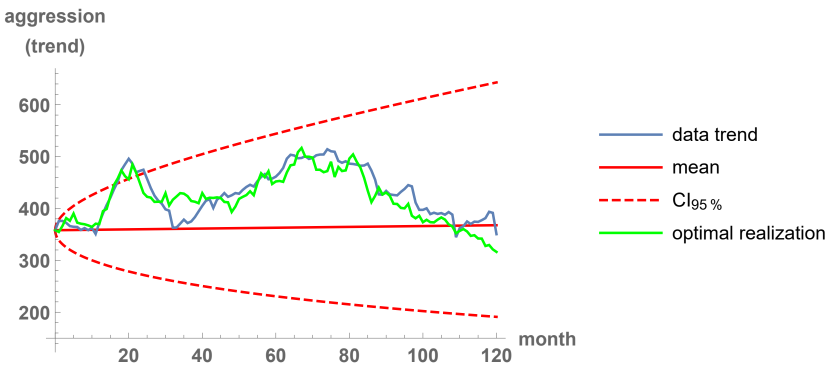

FIGURE 5. Trend-component fitting of aggressions in the city of Valencia. Mean, 0.95 probabilistic interval, and least-squares optimal realization among 10^5 runs.

227 modeled. In this part of *Results*, the three crime events are considered to be independent; they
228 are fitted separately, as detailed in Subsection 2.3.

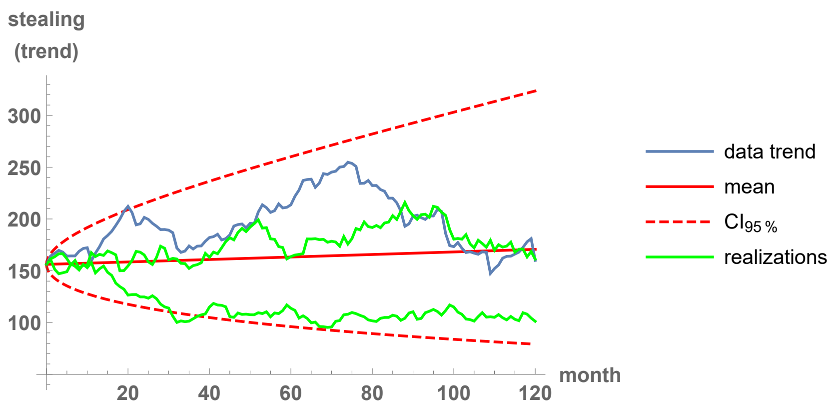

FIGURE 6. Trend-component fitting of stealing in the city of Valencia. Mean, 0.95 probabilistic interval, and two realizations as an example.

229 The estimates of the parameters μ and σ obtained by the method of moments, see (2.5), are
230 given in Table 1. For the three types of events, the estimated global growth rate $\hat{\mu}$ is positive,

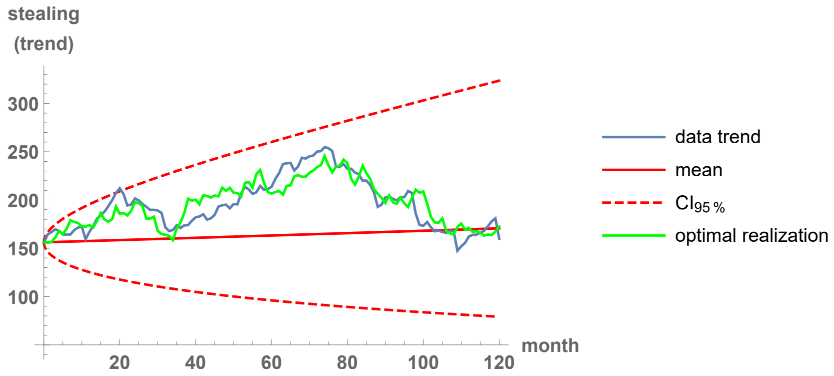

FIGURE 7. Trend-component fitting of stealing in the city of Valencia. Mean, 0.95 probabilistic interval, and least-squares optimal realization among 10^5 runs.

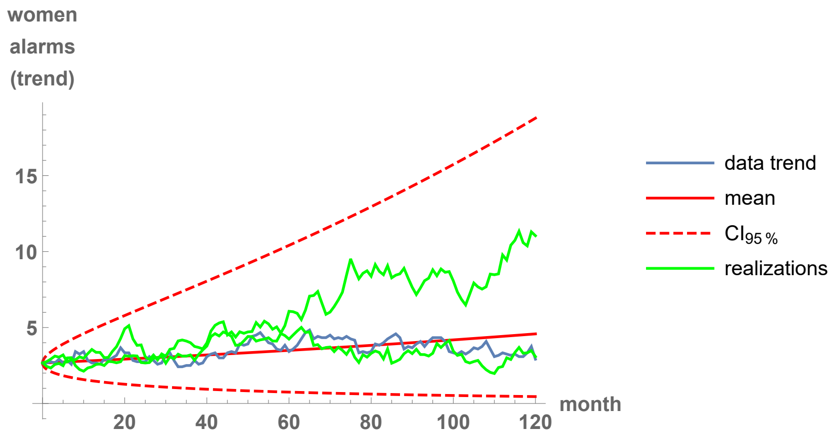

FIGURE 8. Trend-component fitting of women alarms in the city of Valencia. Mean, 0.95 probabilistic interval, and two realizations as an example.

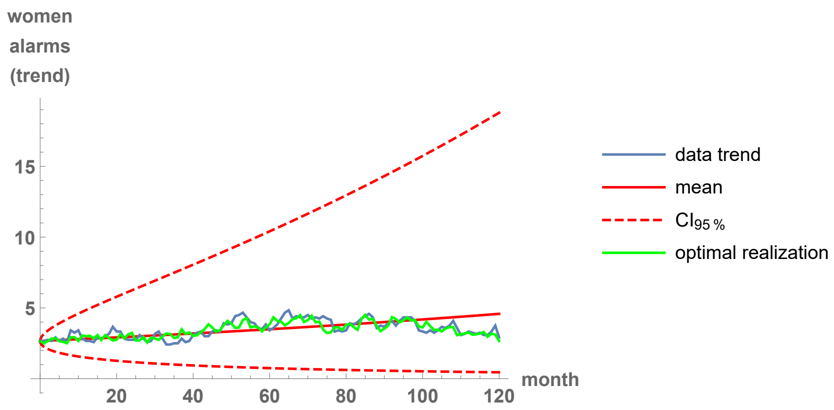

FIGURE 9. Trend-component fitting of women alarms in the city of Valencia. Mean, 0.95 probabilistic interval, and least-squares optimal realization among 10^5 runs.

although nearly zero. This indicates that criminality is similar at the beginning and at the end of the whole time period. The value of $\hat{\sigma}$ gives the magnitude of the infinitesimal standard deviation.

	aggressions	stealing	women alarms
$\hat{\mu}$	0.000220781	0.000740088	0.00450867
$\hat{\sigma}$	0.0282544	0.0328026	0.0867399

TABLE 1. Estimates of the parameters for the three trend components, by the method of moments.

The predictive capability of the stochastic model (2.4) is assessed in Figures 10–13. To avoid repetitions, only the case of aggressions is shown. For each figure, several months are fixed for calibrating the parameters μ and σ by (2.5), and then it is checked whether the criminal events of the remaining months are correctly captured. It should be stressed that we are not seeking quantitative, pointwise forecasts, since this is impossible when working with randomly fluctuating phenomena; rather, we are committed to averaged predictions of crimes, with probabilistic bands. In Figures 10–12, we take three, six and eight years of training. It is perceived that, as the training data increase, the prediction may become worse, since changes in the last months may not be correctly captured. Moreover, forecasts may change with training data, especially for large training periods. For instance, the lower limit of the confidence intervals shows a possibility of decreasing criminality when three and eight years of training are used, but for six years the possibility of crime decreasing is very low. Also, for six years the upper limit grows faster. These facts stem from the level of variability within the training span. As shown in Figure 13, the data between the sixth and the eighth years are a better predictor for the last year than the whole time series; in this manner, the decreasing pattern of the last period is properly reflected. For real-life applications seeking predictability of crime trends, short training scales with recent case counts may be employed to cautiously forecast a few subsequent times. The determination of the training span is not easy and would deserve further research, but it seems that it should be some months long (two years according to the last figure).

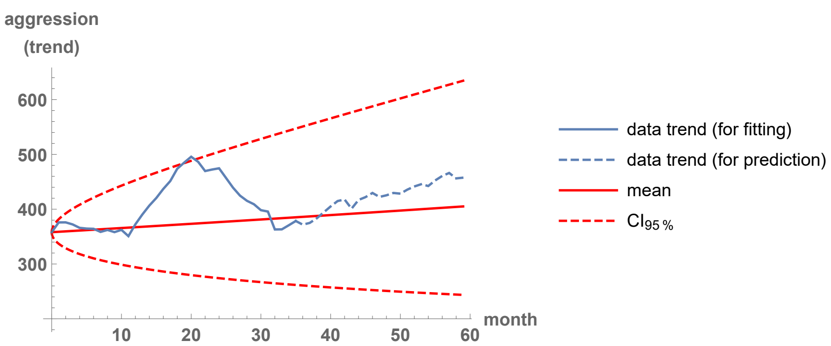

FIGURE 10. Trend-component prediction of aggressions in the city of Valencia, by using three years of training.

3.2. **Fitting of seasonality with Fourier series and noise.** Although it is less interesting for applications, Figures 14–16 show how a noisy, truncated Fourier series (2.6) accommodates the seasonality component. We represent the periodic mean, the 0.95 probabilistic interval,

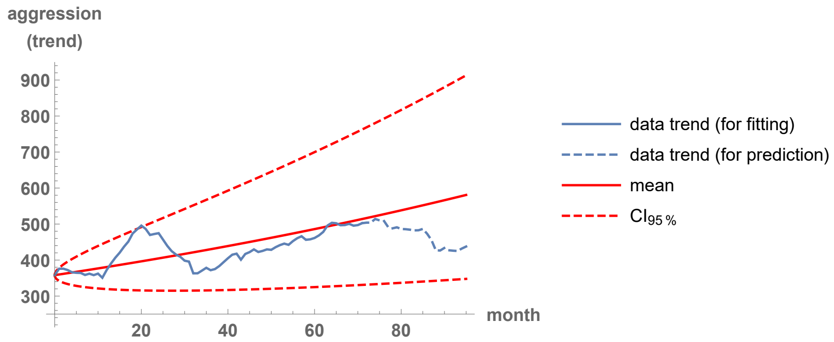

FIGURE 11. Trend-component prediction of aggressions in the city of Valencia, by using six years of training.

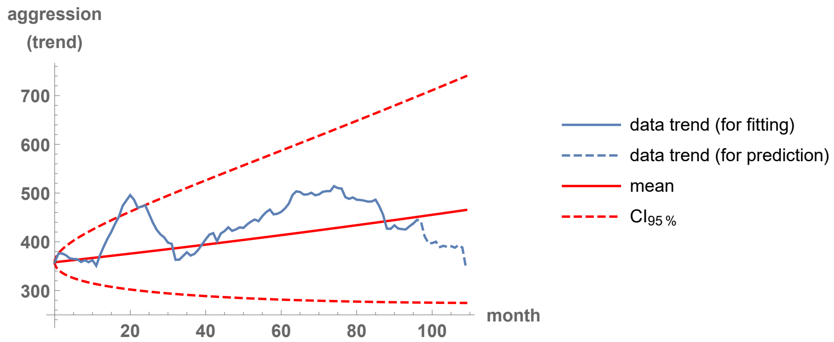

FIGURE 12. Trend-component prediction of aggressions in the city of Valencia, by using eight years of training.

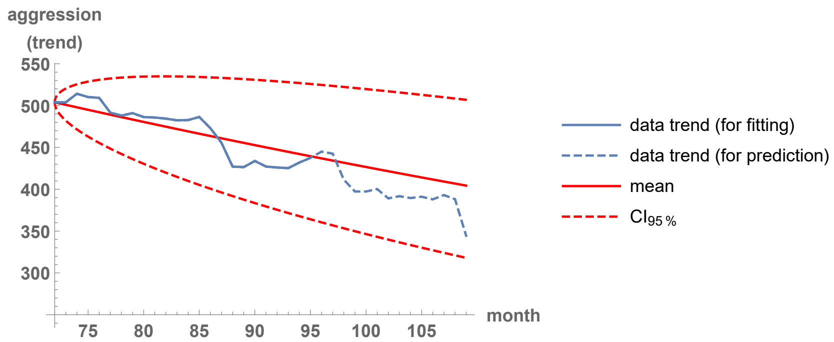

FIGURE 13. Trend-component prediction of aggressions in the city of Valencia, by using training between the sixth and the eighth years.

and the least-squares optimal realization among 10^5 runs (optimality means (3.1)). Of course, the fitting of this type of noise is more difficult than in the Itô-diffusion case of the trend.

We have used the truncation order $K = 4$, and the Fourier coefficients have been calibrated by least-squares optimization. For $K > 4$ harmonic waves, a similar least-squares error is obtained, at the expense of more parameters. The error variance is then fixed as the variance of the residuals sample. In Table 2, the estimates are tabulated for the three criminal events. Observe that the estimated standard deviations $\hat{\sigma}$ are much higher than those for the trends, due to the strongly noisy behavior of seasonality.

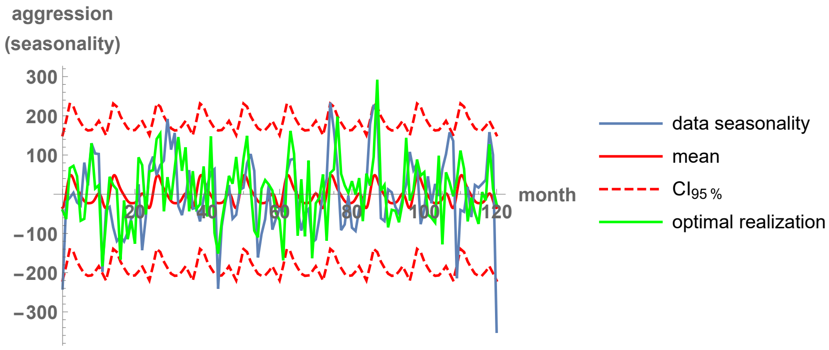

FIGURE 14. Seasonality-component fitting of aggressions in the city of Valencia. Mean, 0.95 probabilistic interval, and least-squares optimal realization among 10^5 runs.

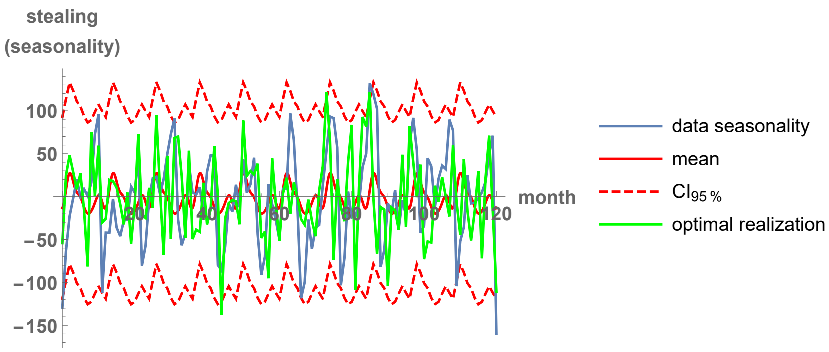

FIGURE 15. Seasonality-component fitting of stealing in the city of Valencia. Mean, 0.95 probabilistic interval, and least-squares optimal realization among 10^5 runs.

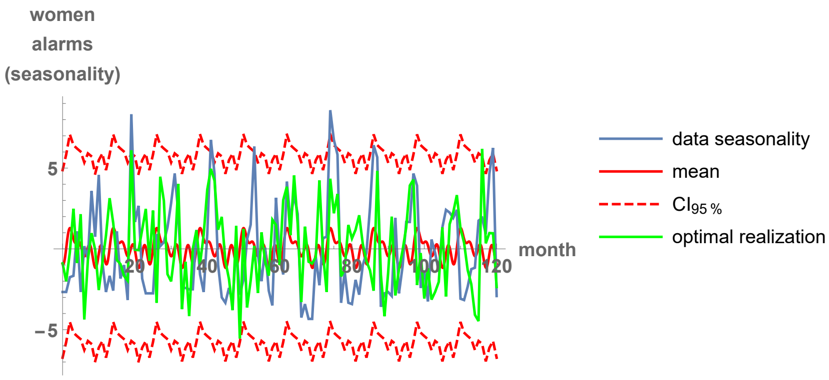

FIGURE 16. Seasonality-component fitting of women alarms in the city of Valencia. Mean, 0.95 probabilistic interval, and least-squares optimal realization among 10^5 runs.

3.3. **Fitting of correlated trends with correlated geometric Brownian motions.** In Figures 17–20, we show the results of modeling the trends of aggression and stealing with two correlated geometric Brownian motion processes, see (2.7) and (2.8). Indeed, as already commented, the evolution patterns of these two events are similar. For each event, we plot the

	aggressions	stealing	women alarms
$\hat{a}_0$	-4.11366	-2.84963	-0.0261963
$\hat{a}_1$	3.49520	6.37358	-0.0518142
$\hat{a}_2$	-19.5921	-8.91074	-0.20953
$\hat{a}_3$	-12.1525	-6.87741	-0.185918
$\hat{a}_4$	-4.29074	-2.34963	-0.467863
$\hat{b}_1$	22.9402	13.7019	0.600703
$\hat{b}_2$	1.13425	-0.0914142	0.137121
$\hat{b}_3$	-1.77500	1.35556	-0.2875
$\hat{b}_4$	-4.29074	-1.2413	-0.185233
$\hat{\sigma}$	94.5516	53.9931	2.96081

TABLE 2. Estimates of the parameters for the three seasonality components, by the method of moments.

mean, a 0.95 probabilistic interval, two examples of realizations, and the least-squares optimal path (with the minimization for the two trend series at the same time) among 10^5 simulations.

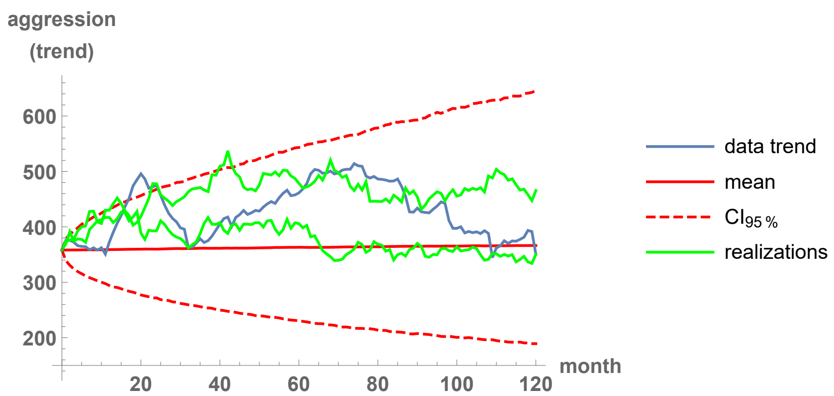

FIGURE 17. Trend-component fitting of aggressions in the city of Valencia, by taking into account correlation between aggression and stealing. Mean, 0.95 probabilistic interval, and two realizations as an example.

The estimates of the parameters are given in Table 3, by using (2.9)–(2.11). The growth rates and the infinitesimal standard deviations are the same as in Table 1. But now, we are identifying the significant correlation between the two Brownian motions, which demonstrates that the use of this model is advisable. For an illustration of the existing interaction, one may jointly sample from $x_{1,t}$ and $x_{2,t}$ at fixed time t (i.e. from (2.7) and (2.8) jointly), and then obtain a scatter plot and the correlation estimate. In Figure 21, scatter plots for $t = 2$ and $t = 100$ are displayed. As t increases, the dispersion of the conditional distribution $x_{2,t}|x_{1,t} = u$ gets larger with u. An approximate functional relationship between $x_{1,t}$ and $x_{2,t}$ may be obtained via a regression line.

3.4. **Summary of the results.** With geometric Brownian motion processes (2.4), the historic time series on trends are fitted for each of the three events separately: aggressions in Figures 4 and 5, stealing in Figures 6 and 7, and women alarms in Figures 8 and 9. The fit consists of the mean value, a 95% probabilistic interval, and realizations. The first figure of each pair simulates two paths, to focus on the qualitative aspects of the fluctuations of the trends. The

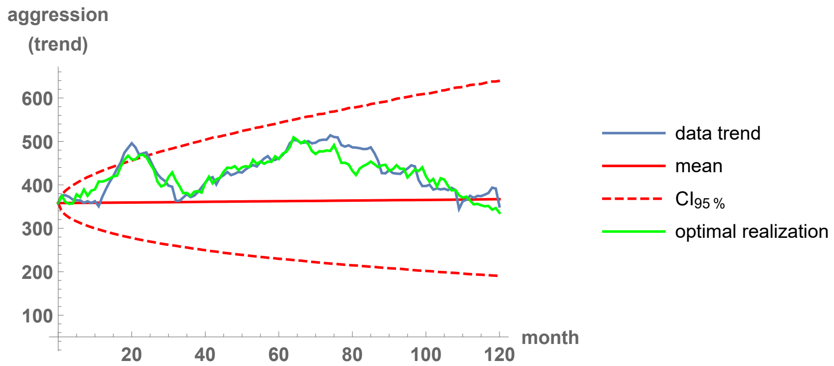

FIGURE 18. Trend-component fitting of aggressions in the city of Valencia, by taking into account correlation between aggression and stealing. Mean, 0.95 probabilistic interval, and least-squares optimal realization among 10^5 runs.

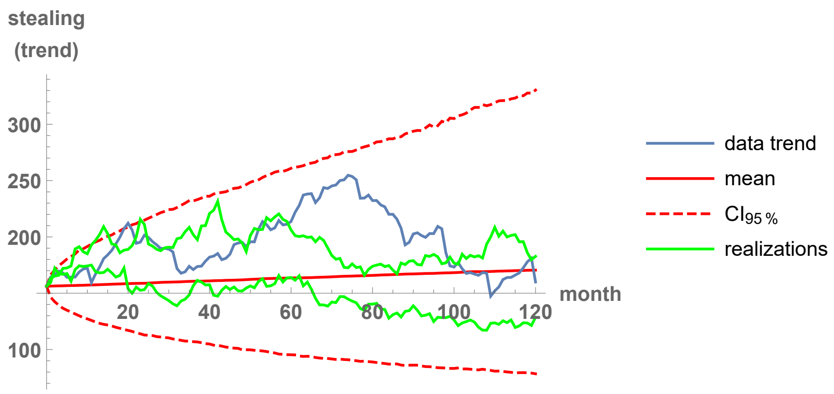

FIGURE 19. Trend-component fitting of stealing in the city of Valencia, by taking into account correlation between aggression and stealing. Mean, 0.95 probabilistic interval, and two realizations as an example.

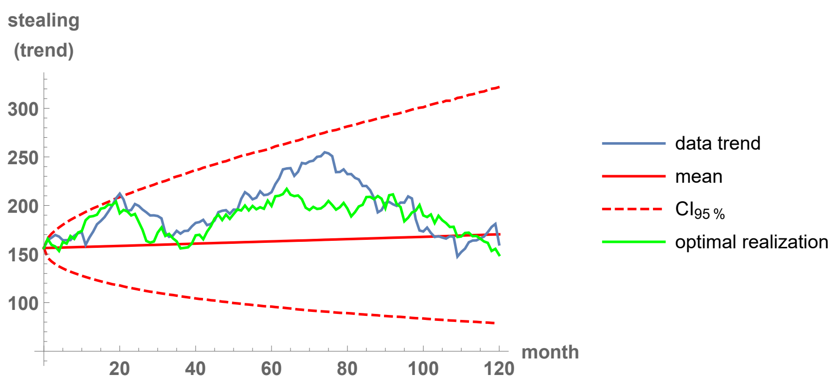

FIGURE 20. Trend-component fitting of stealing in the city of Valencia, by taking into account correlation between aggression and stealing. Mean, 0.95 probabilistic interval, and least-squares optimal realization among 10^5 runs.

	aggression & stealing
$\hat{\mu}_1$	0.000220781
$\hat{\mu}_2$	000740088
$\hat{\sigma}_1$	0.0282544
$\hat{\sigma}_2$	0.0328026
$\hat{\rho}$	0.854833

TABLE 3. Estimates of the parameters when modeling the trends of aggression and stealing with correlations, by using the method of moments.

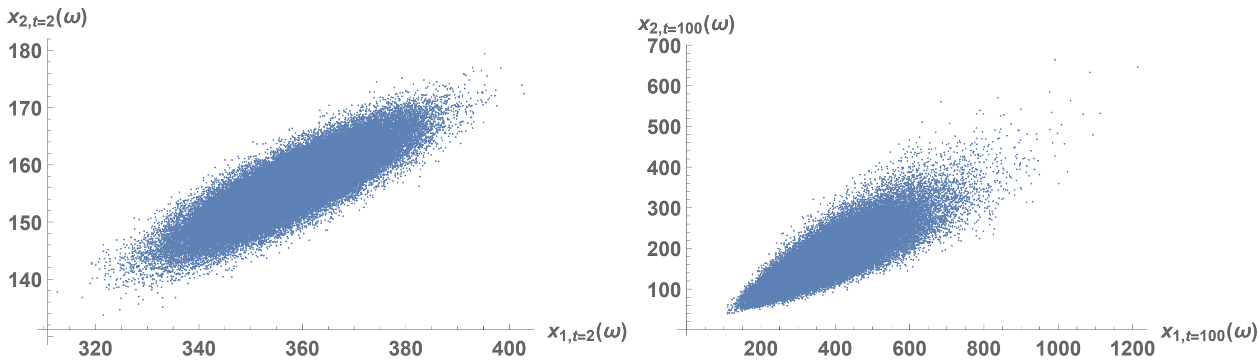

FIGURE 21. Scatter plots for $x_{1,t}$ (aggression) and $x_{2,t}$ (stealing) at $t = 2$ and $t = 100$, by sampling, when modeling the trends of aggression and stealing with correlations.

second figure of each pair plots a least-squares optimal trajectory (3.1) against the trend time series, to focus on quantitative, pointwise fits. Despite its simplicity, the performance of the model is good, since the trend data are nearly reproduced. The estimated parameter values of the model, by the method of moments (2.5), are tabulated in Table 1.

The capability of a model to "view" the future is important. Given a training dataset, which serves for parameter calibration, the incidence of crime in subsequent times is forecast. Figures 10–13 illustrate that matter for aggressions and model (2.4). Future incidences are delimited by probabilistic bands, with average values. Pointwise predictions are not possible. Uncertainty quantification for the model response is devoted to probabilistic measures for outcomes: statistics, regions, thresholds, etc. As the figures show, the selection of the training period is important, because too large periods may not forecast the future well.

Seasonality is studied for the three crime events in Figures 14–16. The seasonality time series are highly noisy. A truncated Fourier series with uncorrelated noise, (2.6), is employed for fitting. The estimated coefficients are given in Table 2.

Finally, aggression and stealing incidents are coupled. This serves as an instance to show the stochastic modeling of any two interacting phenomena. Two non-independent geometric Brownian motions are used to fit the historic trend time series of aggression and stealing, see (2.7) and (2.8). The method of moments renders closed-form estimates for the parameters, by (2.9)–(2.11). Figures 17–20 represent the usual metrics of interest: the mean value, a 95% probabilistic interval, and realizations. Parameter calibrations are detailed in Table 3. Scatter plots for the two events are given in Figure 21. The significant dependence demonstrates the need of introducing a correlation parameter. This new coupled model (2.7)–(2.8) may be used for forecasting too, with mean values and probabilistic intervals as in Figures 10–13.

4. Discussion

As shown in this paper, standard stochastic differential equation models from finance are useful to model crime dynamics. Quantitatively, model trajectories fit historic data along a whole decade. Thus, for short-term predictions, the model may be a useful tool for delineating the incidence of crime, based on mean values and probabilistic regions. Of course, pointwise quantitative forecasts cannot be expected with randomly fluctuating dynamics. We believe that the ability of the model to fit and predict, applied to certain days/weeks/months and neighborhoods/areas/cities, could be an aid for law enforcement.

A critique of our approach might be the lack of mechanistic components, which does not permit understanding social or psychological sources of crime to derive eradication strategies. However, the incorporation of these mechanisms complicates models. As reviewed in the Introduction section, those complex models are restricted to simulating data-independent dynamics [1, 4, 9, 10, 12, 16, 18, 20, 24–28, 30] or entail unidentifiable inverse problems [13], so we are sure that there should be a balance between complexity and applicability. Here is where phenomenological/statistical modeling comes in [15, Section 2.1]. Our adopted approach does not pose any computational difficulty; it allows for fitting and forecasting, and further, it identifies crime interactions (for example, serious and minor events) by simply correlating the noises. Nonetheless, statistical forecasting models are limited by the assumption that future incidence will follow the patterns of incidence observed in the past.

Although phenomenological models of crime based on differential equations have not taken a noticeable place in the literature, these types of models have been widely used in environmental sciences. For example, [6] and [23] employ logistic differential equations to forecast the burden of Zika and Ebola epidemics, respectively; [5] proposes multiple stochastic logistic functions to fit several COVID-19 waves and forecast; and [21] studies the applicability of a stochastic modified Lundqvist-Korf diffusion process to model CO_2 emissions. As our paper shows, differential equation-based statistical models shall be considered a tool to assess the evolution of social behaviors.

Following [11], we tried to spatially divide our city of study into high- and low-criminality zones, but both areas showed similar form of the time series and no gain was clearly perceived. Even so, the inclusion of spatial dependencies, by correlating noises, will be the basis of our future efforts. Here, we are omitting spatial statistics analysis, committed to point patterns from a completely different perspective [7, 8].

The geometric Brownian motion process used for trend evolution mimics the use for stock price evolution. In that financial setting, the variances of the trajectories are unbounded on $[0, \infty)$ and there is no mean reversion, because the prices may rise or diminish indefinitely. An alternative formulation is Vasicek's model, which gives rise to the Ornstein-Uhlenbeck process and possesses the properties of mean reversion and asymptotic finite variance [3]. Used for interest rates in finance [22] since these cannot increase or decrease indefinitely, one may wonder whether the Vasicek's model would be more appropriate for crime dynamics. We tried this model. In terms of pointwise fitting of historic data, we did not find particular differences. Essentially, the difference relied on the probabilistic band, which exhibited bounded amplitude along time. In this sense, the use of one or the other model depends on whether the extent of criminal activities is considered delimited or not.

Some modifications and enhancements of the present paper are here commented. First, the growth-rate parameter μ was considered constant, but it would be more realistic to work with certain dependencies on covariates via link/effect functions [19]. Second, in line with the previous point, covariates could be incorporated as Itô processes into the differential terms instead, by setting a hierarchical stochastic model. While these ideas would help for better forecasts in criminology, the complexity of the model would certainly increase. Third, Poisson jumps could be included in the model, apart from Itô diffusion; as motivated by [29] in the financial setting, at least these jumps may give a better fit of the log-returns. Fourth, independently of the approach followed, it would be of high interest to derive a general methodology for the determination of the training span when forecasting. In our paper, we give some insights on this fourth topic, but it deserves further analysis. And fifth, our stochastic methods could be applicable to spatio-temporal series, by correlating two patches like we did with the two interacting crimes. This last topic is the focus of a future work.

5. Conclusion

The evolution of three time series of criminal activity (aggressions, stealing and women alarms) is analyzed. Our case study corresponds to the calls retrieved by the 112-emergency phone in the city of Valencia, Spain, for the decade 2010–2020. The original noisy time series are decomposed into trend, with an annual moving average, and seasonality. The trend is a smoother version of the raw data and fluctuates as an Itô process. We apply a geometric Brownian motion process with method-of-moments parameter estimation for the three types of events, which also permits analyzing interacting crimes (such as aggression and stealing) by correlating noises and coupling equations. Seasonality is fitted by a randomly perturbed periodic function. Numerical results are essentially based on tabulating parameter estimates and graphing fits of historic data and simulations of forecasts. Our simple approach allows for simulating the real data, rendering short-term predictions, and identifying correlated crimes and risky periods.

Funding

Julia Calatayud has been supported by the postdoctoral contract POSDOC/2021/02 from Universitat Jaume I, Spain (Acció 3.2 del Pla de Promoció de la Investigació de la Universitat Jaume I per a l'any 2021). Jorge Mateu has been supported by the grant PID2019-107392RB-I00 from Spanish Ministry of Science and the grant AICO/2019/198 from Generalitat Valenciana.

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

[1] Abbas S, Tripathi JP, Neha AA (2017) Dynamical analysis of a model of social behavior: Criminal vs non-criminal population. Chaos Soliton. Fract. 98:121–129
[2] Allen E (2007) Modeling With Itô Stochastic Differential Equations. Springer Science & Business Media, Dordrecht, Netherlands
[3] Allen E (2016) Environmental variability and mean-reverting processes. Discrete Cont. Dyn.-B 21(7):2073
[4] Berestycki H, Rodriguez N, Ryzhik L (2013) Traveling wave solutions in a reaction-diffusion model for criminal activity. Multiscale Model. Sim. 11:1097–1126
[5] Calatayud J, Jornet M, Mateu J (2022) A stochastic Bayesian bootstrapping model for COVID-19 data. Stoch. Environ. Res. Risk. Assess. 36:2907–2917. https://doi.org/10.1007/s00477-022-02170-w

[6] Chowell G, Hincapie-Palacio D, Ospina JF, Pell B, Tariq A, Dahal S, Moghadas SM, Smirnova A, Simonsen L, Viboud C (2016) Using phenomenological models to characterize transmissibility and forecast patterns and final burden of Zika epidemics. PLoS Currents 8
[7] Cressie N, Wikle CK (2015) Statistics for Spatio-Temporal Data. John Wiley & Sons, New York
[8] Gelfand AE, Schliep EM (2018) Bayesian inference and computing for spatial point patterns. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 10, Institute of Mathematical Statistics and the American Statistical Association, pp i–125
[9] González-Parra G, Chen-Charpentier B, Kojouharov HV (2018) Mathematical modeling of crime as a social epidemic. J. Interdiscipl. Math. 21(3):623–643
[10] Gu Y, Wang Q, Yi G (2017) Stationary patterns and their selection mechanism of urban crime models with heterogeneous near-repeat victimization effect. Eur. J. Appl. Math. 28(1):141–178
[11] Jane White KA, Campillo-Funollet E, Nyabadza F, Cusseddu D, Kasumo C, Imbusi NM, Ogesa Juma V, Meir AJ, Marijani T (2021) Towards understanding crime dynamics in a heterogeneous environment: A mathematical approach. J. Interdiscipl. Math. 1–21. https://doi.org/10.1080/09720502.2020.1860292
[12] Kolokolnikov T, Ward MJ, Wei J (2014) The stability of steady-state hot-spot patterns for a reaction-diffusion model of urban crime. Discrete Cont. Dyn.-B 19:1373–1410
[13] Lacey AA, Tsardakas MN (2016) A mathematical model of serious and minor criminal activity. Eur. J. Appl. Math. 27(3):403–421
[14] Lamberton D, Lapeyre B (2011) Introduction to Stochastic Calculus Applied to Finance. Second edition, Chapman & Hall / CRC press, London, UK
[15] Lauer SA, Brown AC, Reich NG (2021) Infectious disease forecasting for public health. In: Drake JM, Bonsall MB, Strand MR (eds) Population Biology of Vector-Borne Diseases. Oxford University Press, UK, pp 45–68
[16] Manásevich R, Phan QH, Souplet P (2013) Global existence of solutions for a chemotaxis-type system arising in crime modelling. Eur. J. Appl. Math. 24:273–296
[17] Mao X (2007) Stochastic Differential Equations and Applications. Elsevier
[18] McMillon D, Simon CP, Morenoff J (2014) Modeling the underlying dynamics of the spread of crime. PLoS ONE 9(4):e88923 04
[19] Michelot T, Glennie R, Harris C, Thomas L (2021) Varying-coefficient stochastic differential equations with applications in ecology. J. Agr. Biol. Envir. St. 26(3):446–463
[20] Misra A (2014) Modeling the effect of police deterrence on the prevalence of crime in the society. Appl. Math. Comput. 237:531–545
[21] Nafidi A, El Azri A, Sánchez RG (2022) The stochastic modified Lundqvist-Korf diffusion process: statistical and computational aspects and application to modeling of the CO_2 emission in Morocco. Stoch. Environ. Res. Risk. Assess. 36:1163–1176
[22] Orlando G, Mininni RM, Bufalo M (2020) Forecasting interest rates through Vasicek and CIR models: A partitioning approach. J. Forecasting 39(4):569–579
[23] Pell B, Kuang Y, Viboud C, Chowell G (2018) Using phenomenological models for forecasting the 2015 Ebola challenge. Epidemics 22:62–70
[24] Rodriguez N, Bertozzi A (2010) Local existence and uniqueness of solutions to a PDE model for criminal behavior. Math. Mod. Meth. Appl. S. 20(supp01):1425–1457
[25] Short M, Bertozzi A, Brantingham P (2010) Nonlinear patterns in urban crime: Hotspots, bifurcations, and suppression. SIAM J. Appl. Dyn. Syst. 9(2):462–483
[26] Short MB, Brantingham PJ, Bertozzi AL, Tita GE (2010) Dissipation and displacement of hotspots in reaction-diffusion models of crime. P. Natl. Acad. Sci. USA 107(9):3961–3965
[27] Srivastav AK, Athithan S, Ghosh M (2020) Modeling and analysis of crime prediction and prevention. Social Network Analysis and Mining 10(1):1–21
[28] Srivastav AK, Ghosh M, Chandra P (2019) Modeling dynamics of the spread of crime in a society. Stoch. Anal. Appl. 37(6):991–1011
[29] Synowiec D (2008) Jump-diffusion models with constant parameters for financial log-return processes. Comput. Math. Appl. 56(8):2120–2127
[30] Tse WH, Ward MJ (2015) Hotspot formation and dynamics for a continuum model of urban crime. Eur. J. Appl. Math. 27:583–624
[31] Wolfram Research, Inc. (2020) Mathematica. Version 12.1, Champaign, IL