

Exploring data assimilation and forecasting issues for an urban crime model

DAVID J. B. LLOYD¹, NARATIP SANTITISSADEEKORN¹ and
MARTIN B. SHORT²

¹*Department of Mathematics, University of Surrey, Guildford, GU2 7XH, UK
email: D.J.Lloyd@surrey.ac.uk; n.santitissadeekorn@surrey.ac.uk*

²*School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA
email: mbshort@math.gatech.edu*

(Received 16 January 2015; revised 17 October 2015; accepted 20 October 2015; first published online
2 December 2015)

In this paper, we explore some of the various issues that may occur in attempting to fit a dynamical systems (either agent- or continuum-based) model of urban crime to data on just the attack times and locations. We show how one may carry out a regression analysis for the model described by Short *et al.* (2008, *Math. Mod. Meth. Appl. Sci.*) by using simulated attack data from the agent-based model. It is discussed how one can incorporate the attack data into the partial differential equations for the expected attractiveness to burglary and the criminal density to predict crime rates between attacks. Using this predicted crime rate, we derive a likelihood function that one can maximise in order to fit parameters and/or initial conditions for the model. We focus on carrying out data assimilation for two different parameter regions, namely in the case where stationary and non-stationary crime hotspots form. It is found that the likelihood function is ‘flat’ for large ranges of parameters, and that this has major implications for crime forecasting. Hence, we look at how one might carry out a goodness-of-fit and forecasting analysis for crime rates given the range of parameter fits. We show how one can use the Kolmogorov–Smirnov statistic to assess the goodness-of-fit. The dynamical systems analysis of the partial differential equations proves invaluable to understanding how the crime rate forecasts depend on the parameters and their sensitivity. Finally, we outline several interesting directions for future research in this area where we believe that the combination of dynamical systems modelling, analysis, and data assimilation can prove effective in developing policing strategies for urban crime.

Key words: Crime hotspots; Data assimilation; Maximum likelihood; Point process; Parameter estimation

1 Introduction

One major goal of current research on the mathematics of crime is to develop methods by which crime data can be joined with mathematical models in order to predict future criminal events and aid in developing more effective policing. Towards this end, many sophisticated statistical approaches have been employed that use crime data to estimate spatial and/or temporal risk distributions, which may then be projected forward in time to predict future events. These methods sometimes focus on more stationary crime distributions by correlating crimes with such factors as socio-economic

demographics or spatial proximity to so-called crime generators or attractors [17, 20, 32], which tend to change slowly over time. Other methods focus on the self-exciting nature of crime – the fact that criminal events often increase the risk of further events in the nearby spatio-temporal region – and typically use kernel density estimation techniques to determine the precise way in which future risk is affected by recent criminal events [4, 5, 10, 15, 16].

As a specific example of this statistical approach to crime modelling, consider the ETAS (Epidemic Type Aftershock Sequence) model first employed in [24]. This method attempts to capture both stationary and transitory crime levels using only information on prior crimes; namely, the locations and times of previous criminal events. Within this framework, crimes are interpreted as random events generated by an underlying self-exciting point process, governed by the spatio-temporal intensity (probability density) function

$$\lambda(\mathbf{x}, t) = \mu(\mathbf{x}, t) + \sum_{t_i < t} g(|\mathbf{x} - \mathbf{x}_i|, t - t_i). \quad (1.1)$$

Here, the crime rate density at location \mathbf{x} and time t , $\lambda(\mathbf{x}, t)$, is modelled as a linear superposition of a background rate $\mu(\mathbf{x}, t)$, meant to capture events that occur independent of the crime history (not due to self-excitation), and an excitation kernel g that measures precisely how each prior event i , at location \mathbf{x}_i and time t_i , influences the crime rate at location \mathbf{x} and time t . Under this model, then, every criminal event is either a background event, if it was triggered by background rate μ , or a daughter event of prior crime i , if it was triggered by the excitation kernel g due to event i . Using an Expectation Maximisation method and non-parametric kernel density estimation, one can fit crime data to the model (1.1), revealing the underlying kernels μ and g . These fitted kernels can then be used to predict crimes into the near-future, as has been done recently in Los Angeles and other cities [25]. We note that the model (1.1) falls within a wider class of Hawkes processes where various data assimilation techniques are being investigated; see for example [22].

Though these statistical methods are powerful and can be quite useful, they do suffer from some drawbacks. Chief amongst these is the fact that these methods do not typically elucidate any causal mechanisms for why the kernels are as measured. For example, the ETAS model will return an excitation kernel g , but understanding why the kernel behaves as it does may be far from clear. This is not necessarily an issue if your main goal is simply to predict future events, perhaps so that law enforcement might station patrol units at likely locations in an effort to dissuade crime. However, one might also want to enact other crime control measures directly targeted at altering the structure of g to reduce crime, but, lacking an understanding of what mechanisms are driving the structure of g in the first place, this task becomes rather difficult.

To address concerns such as these, one might, instead of using a statistical approach, turn to more explicit mathematical models of criminal behaviour that have clear causal mechanisms embedded within them. One such model that we will focus on in this paper is that of Short *et al.* [29], which we will refer to as the LA model (it having been constructed by a team of researchers in Los Angeles). In doing this, one gains clear interpretations of parameters and mechanisms, opening more avenues for police strategies to lower crime.

However, we have now traded one difficulty for another – namely, how does one best fit such a model to existing crime data? This is an especially tricky problem in the case of the LA model, since the model's dynamic quantities of interest – criminal density and crime attractiveness – are not directly measurable by themselves. Instead, we have data on the times and locations of criminal events, which are a function of these two quantities.

In this paper, we take a first look at incorporating data into the LA model by focusing specifically on how one may carry out data assimilation for the model. Our aim is not to be exhaustive but to highlight this as an area for more detailed investigation in the future. To explore the issues with data assimilation without getting into the technical difficulties of dealing with real crime data (or worrying about whether the model is a good reflection of reality or not), we use the 1D, agent-based version of the LA model to generate a set of criminal attack times and locations that we will treat as ground truth. The 1D model represents burglary on a street and possesses the same phenomenology as the 2D version i.e., hotspot formation. The task is then to find the parameters and initial conditions for the continuum version of the model only knowing these attack times and locations. Of course, the real goal of any data assimilation is to use the model to forecast possible scenarios. We will explore both the assimilation and forecasting issues mainly in the context where the attack distribution becomes a quasi-stationary process in time and space. However, our assimilation procedure is generalisable to the case when the attack distributions are non-stationary and we will discuss this as well.

There are several major issues that one needs to overcome in order to carry out this analysis. First, as mentioned above, there is the complication that the LA model (and others) only predicts the location of criminals (not known) and the probability for the criminals to attack (a non-physical variable). In either case, the data is simply not known or can never be known. Secondly, agent-based crime models are often stochastic and so fitting a single model run to data makes little sense e.g., how does one deal with an attack when a single model run says there is no criminal present? Hence, one needs to fit probability distributions of the models and this quickly becomes computationally very expensive. For both problems, we show how one can overcome them in the context of the agent-based LA model. We note that data assimilation for agent-based models (ABMs) is an emerging field; see for example [7, 12].

We will show that using just 'optimal' parameter fits could yield poor forecasts and we find from the data assimilation step a feasible region of parameter space that could provide a good model fit. Therefore, one needs to have a good understanding of how the dynamics of the model depend on the parameters in order to provide good forecasts. Here, the dynamical systems investigations of [3, 19, 21, 26, 27] prove invaluable.

The paper is outlined as follows. In Section 2, we briefly review the agent-based LA model [29] and the averaged Partial Differential Equation (PDE) system derived from the ABM. We then explain our setup and methodology in Section 3. The data assimilation is carried out in Section 4, where we generate the attack data and incorporate this into a model to carry out the regression analysis. We then use the model to predict a range of various scenarios that may be observed in Section 4.4. Finally, in Section 5 we discuss our results and outline future directions of research.

2 Review of the LA model

2.1 Agent-based model

The agent-based stochastic model of Short *et al.* [29] simulates two quantities on a lattice; the locations of the criminal agents that will commit the burglaries at the lattice sites and an attractiveness of each lattice site to a burglar, which is to be understood as the rate at which criminals located at that site commit burglaries. The attractiveness field, $A_s(t)$ at the lattice site, s , is modelled as

$$A_s(t) = A^0 + B_s(t), \quad (2.1)$$

where A^0 is the intrinsic attractiveness of site s and $B_s(t)$ is the dynamic attractiveness. Thus, the model attempts to capture the possibility of both static hotspots through A^0 and dynamic hotspots through B_s ; these quantities are similar in spirit to μ and g from the ETAS model (1.1). The dynamic attractiveness will be used to model the self-excitation effects, which are two-fold [28]. First, it is noted that when a specific home s is burgled, the rate of burglaries at s increases in the near-future; this is sometimes referred to as the ‘exact-repeat’ effect. Second, there is a neighbourhood effect, such that a crime at s increases the likelihood of future crimes at sites neighbouring s as well; this is sometimes referred to as the ‘near-repeat’ effect. So, in 1D, the dynamic attractiveness B at each site is given by

$$B_s(t + \delta t) = \left[B_s(t) + \frac{\eta l^2}{2} \Delta B_s(t) \right] (1 - \omega \delta t) + \Theta E_s(t), \quad (2.2)$$

where $\eta \in [0, 1]$ measures the relative strength of neighbourhood effects, l is the lattice spacing, ω is the dynamic attractiveness decay rate (so that elevated crime risk lasts only a finite amount of time), Θ is the increase of the attractiveness of s due to one burglary event there, Δ is the discrete spatial Laplacian operator, and $E_s(t)$ is the number of burglaries that occurred at site s over the timestep δt .

To model the criminal agents, the LA model assumes that criminals are performing a random walk biased toward areas of higher attractiveness and occasionally committing crimes, which ends their walk. So, during any given timestep δt , a criminal agent at location s may either commit a crime there, thus ending his movement and removing him from the lattice, or choose a new location to move to among the neighbouring sites of s . The probability of committing a crime is given by

$$p_s = 1 - e^{-A_s(t)\delta t}, \quad (2.3)$$

and, assuming no crime is committed, the probability of moving to site s' that is a neighbor of s (let the notation $s' \sim s$ signify this) is

$$q_{s \rightarrow s'} = \frac{A_{s'}}{\sum_{s'' \sim s} A_{s''}}. \quad (2.4)$$

Finally, criminals are also introduced at each site via a stationary Poisson process with rate Γ .

The agent-based LA model thus contains seven parameters: A^0 , η , ω , θ , Γ , ℓ , and δt . In [29], it is shown that the behaviour of the model varies substantially across parameter regimes, but that three basic forms of behaviour are present: no hotspots, transitory hotspots, and stationary hotspots. In the no hotspot case, crime levels are roughly uniform over space and time; in the transitory hotspot case, crime levels are not uniform over short time intervals, in which spatial hotspots exist but tend to move around and die out over longer periods of time; and in the stationary hotspot case, spatial hotspots form and remain indefinitely, leading to very non-uniform crime levels.

2.2 PDE model

The above ABM can be converted into a pair of PDEs by first replacing all stochastic quantities with their expectation values and then taking the limit as $\ell \rightarrow 0$ and $\delta t \rightarrow 0$, with $\ell^2/2\delta t = D$ and $2\theta\delta t = \theta$ fixed. Upon doing so, and assuming that A^0 is uniform in space, one obtains

$$A_t = \eta D A_{xx} - \omega(A - A^0) + \theta D \rho A, \quad (2.5a)$$

$$\rho_t = D \left[\rho_x - \frac{2\rho}{A} A_x \right]_x - \rho A + \gamma, \quad (2.5b)$$

where $A = A(x, t)$ is the continuum attractiveness field; $\rho = \rho(x, t)$ is the criminal density; η , ω , and A^0 are the same parameters as in the ABM; D (the diffusion coefficient) and θ are as indicated above; and γ is the rate of criminal generation per unit area.

Note in particular the quantity ρA that appears in both equations. This is the average crime rate density, akin to the quantity λ from the ETAS model (1.1). Within the equation for A , this term serves as a generator of attractiveness due to the self-excitation present in the model. In the equation for ρ , this term acts as a sink for criminals, since they are removed from the lattice when they commit crimes. This term also represents the quantity closest to actual data that we may want to assimilate into the model, which are times and locations of criminal events. However, ρA is a stochastic intensity, and any actual events are but one realisation of this intensity. At the same time, though, the self-excitation of the model demands that the system evolve in such a way as to respond to this realisation, rather than to the underlying intensity. It is these considerations that we explore below when determining how best to assimilate ‘data’ into this PDE model.

We will now just provide a brief overview of some standard dynamical systems results for the PDE system (2.5). Steady spatially homogeneous states, $(A, \rho) = (\bar{A}, \bar{\rho})$ satisfy

$$\bar{A} = \frac{\theta D \gamma + A^0 \omega}{\omega}, \quad \bar{\rho} = \frac{\gamma \omega}{\theta D \gamma + A^0 \omega},$$

and the crime rate $\bar{\rho} \bar{A} = \gamma$ (and hence only depends on the criminal generation per unit area). This state is linearly stable to spatially periodic perturbations provided

$$\eta > \frac{3\bar{\rho} + 1 - \sqrt{12\bar{\rho}}}{\bar{A}},$$

otherwise stationary spatially periodic hotspots form; see [26, 29]. In this paper, we will

be looking at the case that $\eta \ll 1$. A good description of the stationary hotspots in the case when $\eta \ll 1$, can be found by using singular perturbation theory; see [3, 19, 21]. Employing the rescaling

$$\tilde{A} = \frac{A}{\omega}, \quad \tilde{\rho} = \frac{\theta D}{\omega} \rho, \quad \tilde{x} = \sqrt{\omega} x, \quad \tilde{t} = \omega t, \quad \alpha = \frac{A^0}{\omega}, \quad \beta = \frac{\gamma \theta D}{\omega^2}, \quad \epsilon^2 = \eta D,$$

yields the PDE system studied by Kolokolnikov *et al.* [19] and Berestycki *et al.* [3]

$$\tilde{A}_t = \epsilon^2 \tilde{A}_{\tilde{x}\tilde{x}} - A + \tilde{\rho} \tilde{A} + \alpha, \quad (2.6)$$

$$\tilde{\rho}_t = D \left[\tilde{\rho}_{\tilde{x}} - \frac{2\tilde{\rho}}{\tilde{A}} \tilde{A}_{\tilde{x}} \right]_{\tilde{x}} - \tilde{\rho} \tilde{A} + \beta. \quad (2.7)$$

They showed that in the case that $D \gg 1$ and $\epsilon \ll 1$, the PDE system possesses a stationary hotspot of the form

$$\tilde{A} \sim \left(\frac{2L\beta}{\pi\epsilon} - \alpha \right) \operatorname{sech} \left(\frac{\tilde{x}}{\epsilon} \right) + \alpha, \quad \tilde{\rho} \sim 2 \operatorname{sech}^2 \left(\frac{\tilde{x}}{\epsilon} \right),$$

where $\tilde{x} \in [-L, L]$ periodic. For our purposes, the crime rate of the stationary hotspot is of interest and is given by (in original variables)

$$\rho A \sim \frac{2\omega^2}{\theta D} \operatorname{sech}^2 \left(\frac{x}{\sqrt{\omega}\epsilon} \right) \left[\left(\frac{2L\gamma\theta D}{\pi\epsilon\omega^2} - \frac{A^0}{\omega} \right) \operatorname{sech} \left(\frac{x}{\sqrt{\omega}\epsilon} \right) + \frac{A^0}{\omega} \right],$$

on $x \in [-L, L]/\sqrt{\omega}$. Some basic observations of this crime rate is that $\sqrt{\omega}\epsilon$ governs the width of the hotspot with it being most sensitive to ϵ i.e., η and D . The maximum crime rate in this situation is given by $4L\gamma/\pi\epsilon$ and is just governed by the parameters γ , the criminal generation per unit area, the diffusion coefficient D and η , the size of the neighbourhood effects.

3 Set-up

3.1 Truth and model runs

In order to explore the data assimilation issues directly without getting into the technical difficulties of dealing with real crime data, we use the 1D version of the agent-based LA model to construct ‘ground truth’ data that we will then proceed to assimilate into the PDE model. We simulate the ABM to a time T with the attractiveness at each point initially set to $A^0 + \theta D \gamma / \omega$ (i.e., the spatially homogeneous steady state) and the number of criminals at each location initially set to one. The simulated attack times t_k and locations x_k , which we denote as $y(x_k, t_k)$ where $k = 1, \dots, N$ and N is the total number of attacks in the simulation, are then collated.

Bearing in mind the points raised in Section 2.2 concerning the term ρA in the PDEs (2.5), we proceed with the continuum model as follows. First, we note that in (2.5a), the ρA term acts as a source of attractiveness, which corresponds to the local increase in attractiveness Θ that occurs in the ABM whenever an event occurs. So, to allow the attractiveness field to evolve in such a way that the locations of actual events are

respected, we remove the ρA term from (2.5a). This leaves the PDE

$$A_t = \eta D A_{xx} - \omega(A - A^0), \quad (3.1)$$

which is completely uncoupled from ρ , and in fact is linear with an exact solution available, given initial and boundary conditions. The events are then introduced back into this field's evolution in the following way. Let time t' be a moment at which no event occurs, with $A(x, t')$ fully specified, and the next actual event from the data sequence y that will occur subsequent to t' be event k . We evolve A using (3.1) until time t_k . Then, the new field $A(x, t_k)$ is instantaneously modified via

$$A(x, t_k) \rightarrow A(x, t_k) + \theta D \delta(x - x_k), \quad (3.2)$$

where $\delta(x)$ is the Dirac delta function. Hence, the event k has caused a sudden increase in attractiveness, of magnitude θD , localised at x_k at time t_k ; precisely what should occur given the ABM. This process can then be repeated as necessary until time T is reached, giving a full solution for $A(x, t)$ over the time-frame of interest.

Given the modification made to (2.5a) in order to incorporate the event data, one must also consider how (2.5b) should be modified. Here, though, the answer is not so simple. The ρA term in (2.5b) serves as a sink for criminals, corresponding to the removal of each criminal from the ABM once he commits a crime. However, it would be incorrect to simply drop the ρA term from (2.5b) and instead locally decrease the density of criminals ρ after each criminal event, as would be the analogue to how the dynamics of A were modified. This is because ρ is really a linear combination of probability distributions on where criminals are located in space at a given time. Hence, in order to remove the offending criminal properly, we would first have to determine for each offender in our superposition the probability that he was the offender that committed the crime at x_k at time t_k , then alter ρ *globally* by subtracting from it the probability distribution for each individual offender weighted by the probability that he was the one who committed the crime in question.

To better illustrate this, let us consider a simple scenario in which $\gamma = 0$, so that no new offenders are introduced, and for which $A(x, 0) = A^0$ and $\rho(x, 0) = \delta(x - x_1) + \delta(x - x_2)$, such that we initially have two criminals, which we will refer to as criminal 1 and criminal 2, starting out located precisely at points x_1 and x_2 , respectively. Further, suppose the first event is at $x = 0$ at time $t = t_1$ and that our spatial domain is infinite in extent. Since no events are occurring between times 0 and t_1 , $A(x, t) = A^0$ for all times up to t_1 , and no criminals ought to be removed over this time, so that ρ will evolve according to

$$\rho_t = D \left[\rho_x - \frac{2\rho}{A} A_x \right]_x. \quad (3.3)$$

However, since $A_x = 0$, we have ρ solving the standard heat equation, such that the solution at time $0 < t \leq t_1$ is simply

$$\rho(x, t) = \frac{1}{\sqrt{4\pi Dt}} \left[e^{-(x-x_1)^2/4Dt} + e^{-(x-x_2)^2/4Dt} \right]. \quad (3.4)$$

When the first event happens, then, we must first determine for each of our criminals the probability that they are the one committing the crime. Since each criminal is equally likely to commit the crime, conditional on the fact that they are present at $x = 0$ at time $t = t_1$, these probabilities are given by the relative proportion of ρ that each criminal contributes at the origin at t_1 . Hence,

$$p_1 = \frac{e^{-(x_1)^2/4Dt_1}}{e^{-(x_1)^2/4Dt_1} + e^{-(x_2)^2/4Dt_1}}, \quad p_2 = \frac{e^{-(x_2)^2/4Dt_1}}{e^{-(x_1)^2/4Dt_1} + e^{-(x_2)^2/4Dt_1}}. \quad (3.5)$$

Once the event occurs, we must instantaneously remove a criminal. To do so, we subtract from ρ each individual's own current probability distribution weighted by the probability that they committed the crime, so that $\rho(x, t_1)$ is modified by the criminal event to be

$$\rho(x, t_1) = \frac{1}{\sqrt{4\pi Dt_1}} \left[(1 - p_1)e^{-(x-x_1)^2/4Dt_1} + (1 - p_2)e^{-(x-x_2)^2/4Dt_1} \right]. \quad (3.6)$$

In principle, one could extend this technique to the more general case where $\gamma \neq 0$ and initial conditions are arbitrary. The first task of determining the probability that each criminal is the one committing the next event can be recast into a geographic profiling problem [23], where one attempts to determine where the criminal who committed an event originated. Once this has been determined, the criminal density can then be altered as described above in response to the event. However, given that this process would likely increase the computational time for our method significantly with questionable benefits to accuracy, we leave its implementation to further work. Instead, we take a more practical approach here, and simply leave the equation that we use to solve for ρ unaltered from the form (2.5b), so that the discrete events are only directly used to modify the attractiveness field (though they indirectly alter ρ through its coupling with A). Looking back to the original PDE system (2.5), if the approximation for the attractiveness field A is reasonable, then the approximation of the expected crime rate, ρA , should also be reasonable since

$$\rho A = \frac{1}{\theta D} [A_t - \eta D A_{xx} + \omega(A - A^0)],$$

and the right-hand side solely depends on A . We justify this choice *ex post facto* by noting that this method yields results that appear quite reasonable, both in terms of parameter estimates and in direct comparisons between ρ and the true agent density.

In summary, then, the events are assimilated into the PDE model by simulating the equations

$$A_t = \eta D A_{xx} - \omega(A - A^0), \quad (3.7a)$$

$$\rho_t = D \left[\rho_x - \frac{2\rho}{A} A_x \right]_x - \rho A + \gamma, \quad (3.7b)$$

between each attack. In order to initialise the PDE system, we set $A(x, 0) = A^0 + \theta D \gamma / \omega$, $\rho(x, 0) = 1/l^2$. We then simulate the PDE system to the first attack time, t_k , happening at x_k . At t_k , we stop the PDE simulation and re-start the simulation with an updated A -function where we have increased the attractiveness at x_k by the number of attacks that occurred at x_k times $\theta D / \delta x$, where δx is the computational spatial grid

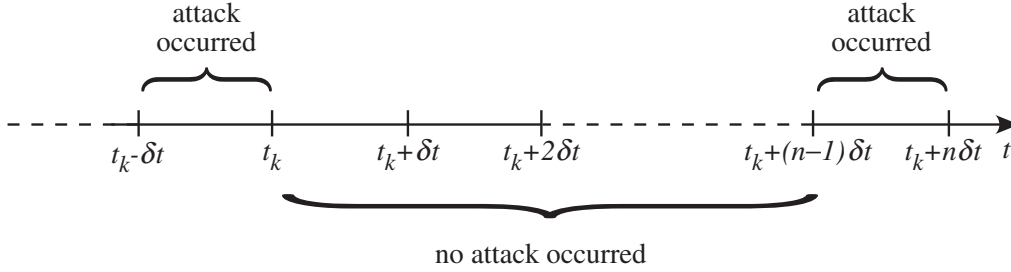


FIGURE 1. A schematic sketch of when attacks occur in one point in space.

size; the $1/\delta x$ term represents our computational approximation of the Dirac delta. The ρ function is re-started with the same value before the simulation was terminated. We repeat this re-start at every attack time to yield $A(x, t)$ and $\rho(x, t)$.

We discretise the PDE system (3.7) as follows. For the A equation, we discretise space using an equi-spaced mesh and use a second-order finite-difference (or pseudo-spectral Fourier method) to evaluate the spatial derivative, A_{xx} . To time-step the PDE, we use a first-order semi-implicit method where the spatial derivative is solved for in the advanced time-step; see [9]. The discretisation of the ρ equation is more difficult. We found that a similar space and time discretisation to that for the A -equation was numerically very unstable due to the delta functions being introduced in A at each attack time and location. Hence, we employ a time-step method inspired by that of the ABM; see [29, equation (3.3)]. First, we split the linear operator in (3.7b) such that we treat the $-\rho A$ term first, leaving an equation with an exact solution that will behave well even for the high A values that accompany the delta function additions. Accordingly, we let $\tilde{\rho}_i(t) = \rho_i e^{-A_i \delta t}$, where $A_i(t) \approx A(x_i, t)$, $\rho_i(t) \approx \rho(x_i, t)$, the space mesh is given by $x_i = i\delta x$, $i = 0, 1, \dots, N_x$, δx is the computational grid spacing, and δt is the time-step. The remaining linear operator acting on $\tilde{\rho}$ is then approximated by the following scheme, which leaves us with our updated value of ρ :

$$\rho_i(t + \delta t) = \left(1 - \frac{2D\delta t}{(\delta x)^2}\right) \tilde{\rho}_i + \frac{2D\delta t}{(\delta x)^2} A_i \left[\frac{\tilde{\rho}_{i+1}}{A_i + A_{i+2}} + \frac{\tilde{\rho}_{i-1}}{A_i + A_{i-2}} \right] + \delta t \gamma. \quad (3.8)$$

Note that this scheme avoids any direct calculations of A_x , which would tend to give very large values near the delta functions, and is thus better behaved. The relationship of this scheme to (3.7b) is detailed in Appendix A.

3.2 Log-likelihood function of observing the attacks

Using the simulated PDE described in the previous section, we now describe a function that measures the goodness of the PDE model fit to the discrete data. Between any attacks, we have the following situation as shown in Figure 1, where we have n time intervals of size δt and space size δx with no attacks and then a final time interval during which we know an attack occurred. Our aim is to calculate a probability that this situation occurs that we can then maximise over the set \mathbf{q} of model parameters.

We assume that the probability of a criminal attacking a location region $x \in [x_j - \delta x/2, x_j + \delta x/2]$ in the time interval $t \in [t_i, t_i + \delta t]$ is governed by a Poisson process with rate

$$\lambda_{ij} = \rho(x_j, t_i)A(x_j, t_i)\delta t\delta x,$$

where ρ and A are calculated from the simulation of (3.7). Hence, the probability that a criminal does not attack in a time interval $t \in [t_i, t_i + \delta t]$ and space interval $x \in [x_j - \delta x/2, x_j + \delta x/2]$ is given by

$$\mathbb{P}(\text{no attack in } t \in [t_i, t_i + \delta t]|\mathbf{q}) = e^{-\lambda_{ij}}, \quad (3.9)$$

where \mathbf{q} is the vector of parameters to be fitted. Similarly, the probability that an attack occurs in a space interval $x \in [x_j - \delta x/2, x_j + \delta x/2]$ and time interval $t \in [t_i, t_i + \delta t]$ is given by

$$\mathbb{P}(\text{one attack in } t \in [t_i, t_i + \delta t]|\mathbf{q}) = \lambda_{ij}e^{-\lambda_{ij}}. \quad (3.10)$$

Assuming that the events in each interval are independent, we compute the total probability of no attacks in the time interval $t \in [t_k, t_{k+1}]$ in a space interval $x \in [x_j - \delta x/2, x_j + \delta x/2]$ followed by an event at t_{k+1} as

$$\begin{aligned} & \mathbb{P}(\text{no attack in } t \in [t_k, t_{k+1} - \delta t] \wedge \text{attack in } t \in [t_{k+1} - \delta t, t_{k+1}]|\mathbf{q}) \\ &= \lambda_{(k+1),j}e^{-\lambda_{(k+1),j}} \cdot e^{-\sum_{i=1}^{n-1} \lambda_{ij}}, \end{aligned} \quad (3.11)$$

Computing the probability for all the N number of attacks from the agent-based simulation to time T , summing over all space and taking the continuum limit as $\delta x, \delta t \rightarrow 0$, yields the likelihood function of observing the attacks,

$$\mathbb{P}(\text{observing the attacks}|\mathbf{q}) \propto \left(\prod_{k=1}^N \rho(x_k, t_k)A(x_k, t_k) \right) \cdot \exp\left(-\int_0^T \int_0^{L_x} \rho A dx dt\right). \quad (3.12)$$

We wish to maximise this likelihood function over the set of unknown parameters \mathbf{q} . It is often easier to deal with the natural logarithm of the likelihood function and so for computational reasons we will use the log-likelihood function, \mathbb{L}

$$\mathbb{L}(\text{attacks}; \mathbf{q}) = \sum_{k=1}^N \log(\rho(x_k, t_k)A(x_k, t_k)) - \int_0^T \int_0^{L_x} \rho A dx dt. \quad (3.13)$$

Note that this is the same log-likelihood function that is maximised in other methods (such as ETAS) of crime density estimation. The essential interpretation is that we would like our simulated ρ and A to have a large product at the times and locations where events occurred, but we do not want the spatio-temporal integral of their product (which is the expected number of crimes) to be arbitrarily large to achieve this.

3.3 Kolmogorov–Smirnov (KS) Statistic

Once the Maximum Likelihood Estimate (MLE) method has been used to carry out the fitting, it is desirable to also assess the quality of the fit without prior knowledge

of the ‘truth’. Solely using point estimates for the forecasting can lead to a false sense of confidence in the scenario analysis and hence it is crucial for the forecasting to have some knowledge of the confidence intervals for the fit. Here, we construct a standard non-parametric KS statistic to help assess the goodness-of-fit from the MLE that does not require knowledge of the ‘truth’. This test comes with the standard assumption of a perfect model and should be used only as a “rule-of-thumb” to assess the quality of the fit.

Given a time-dependent attack rate $\lambda(x_j, t)$ at a location x_j and $\int_0^t \lambda(x_j, u) du < \infty$ for all $t \in [0, T]$, we can define the transformed time scale

$$s_k^j = \int_0^{t_k} \lambda(x_j, u) du, \quad (3.14)$$

where the t_k is the k th attack time. For convenience, we suppress the j -superscript and subscript. It is well known that the series $\{s_k\}$ is a Poisson process with constant unit rate [6]. Therefore, the elapsed time between the $(k-1)$ st and the k th attacks, denoted by

$$\tau_k = s_k - s_{k-1}, \quad (3.15)$$

has an exponential distribution with mean 1. It follows immediately that $z_k := 1 - \exp(-\tau_k)$ has a uniform distribution on the interval $(0, 1)$. This allows us to apply the KS test for z_k as the following:

$$\mathcal{D}_N = \sup_z |F_N(z) - z|, \quad (3.16)$$

where N is the number of data points and $F_N(z)$ is the empirical distribution function of the series z_1, \dots, z_N . If the estimate $\lambda(x, t)$ statistically agrees with the actual series of the attack times t_k , \mathcal{D}_N should be small and we should expect the optimal parameter estimate to produce the attack rate that minimises \mathcal{D}_N over the parameter space. In addition, if the estimate of the intensity function is correct, the points z_k should lie on the 45-degree line. For a sufficiently large N , the 95% confidence intervals are approximated as $b_k \pm 1.36/N^{1/2}$, where $b_k = (k - 1/2)/N$ for $k = 1, \dots, N$ [14].

We note that one could use the KS statistic instead of the MLE method to carry out the fitting. In this case, we have found that it yields similar results (not shown) to the MLE but with the disadvantage that the KS statistic can no longer be used to compute confidence intervals.

4 Results

4.1 Implementation and initial comparison between ABM and PDE simulations

We use the ABM to generate a ‘truth’ run. To this end, we solve the ABM with 66 spatial grid points and $\delta t = 0.01$ and grid spacing $\ell = 1$. The true observation is the attack times and locations over the time period $t \in [0, T]$ for $T = 70$. We discretise the PDE on the same temporal and spatial mesh for simplicity, but note that this is not necessary.

We will focus on two sets of parameters:

- *Stationary hotspots*: We set the model parameters to $\omega = 1/15$, $\eta = 0.03$, $\Theta = 0.056$, $\Gamma = 0.19$, $A_0 = 1/300$.

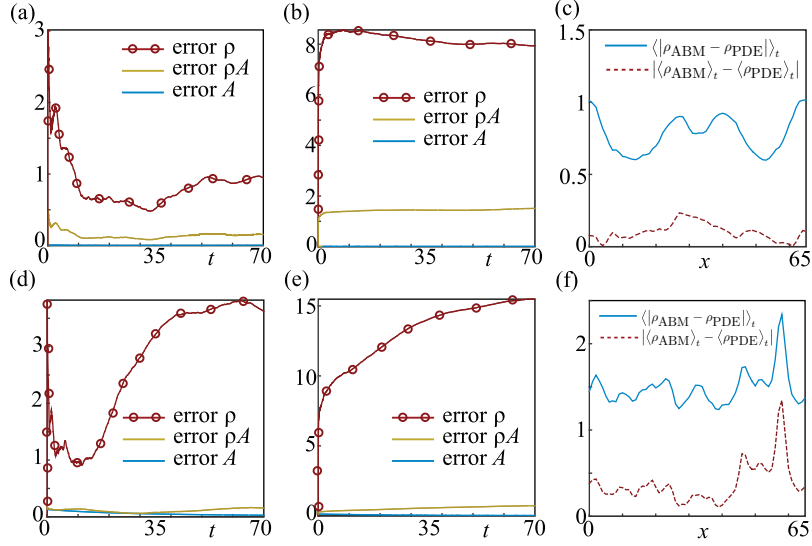


FIGURE 2. Comparison between simulated ABM and PDE. For the stationary hotspot case, we plot error_w in panel (a) and error_s in panel (b). For the non-stationary case, we plot in panel (d) error_w and panel (e) error_s . In panels (c) and (f), we plot the weak and strong point-wise difference (the mean, $\langle \cdot \rangle_t$, taken over the time interval $t \in [0, 70]$) of the criminal density for the stationary and non-stationary hotspot case, respectively.

- *Non-stationary hotspots*: We set the model parameters to $\omega = 1/15, \eta = 0.2, \Theta = 0.0056, \Gamma = 0.19, A_0 = 1/30$.

For the stationary hotspot case, we imagine it will be easier to carry out data assimilation as one just needs enough attack data for the expected crime rate to converge. However, it is unlikely that one has stationary hotspots in reality and so we also look at the second case where the hotspots are not stationary.

Before carrying out any data assimilation, we first investigate how the ABM and PDE compare when we know the true parameter values. We find for the stationary hotspot case, there are roughly 1,200 attacks in the time interval $t \in [0, 70]$ and approximately 700 attacks for the non-stationary hotspot case. In Figure 2, we plot for both sets of parameters two different ‘errors’:

- *Weak difference*: $\text{error}_w(u(t)) = \|\langle u(t)_{\text{ABM}} \rangle_t - \langle u(t)_{\text{PDE}} \rangle_t\|_2$, where $\langle \cdot \rangle_t$ is the temporal mean of $u(t)$ from $t = 0$ to time t .
- *Strong difference*: $\text{error}_s(u(t)) = \langle \|u(t)_{\text{ABM}} - u(t)_{\text{PDE}}\|_2 \rangle_t$, where $\langle \cdot \rangle_t$ is the temporal mean of $u(t)$ from $t = 0$ to time t .

We see in Figure 2 that the attractiveness field, A , is well approximated even point-wise indicating that the dynamics between the attacks is governed by the first equation in (3.7). Despite the issue with not evolving the criminal density between attacks correctly, as discussed in Section 3, we see in Figures 2(c) and (f) that on average we fail in certain locations to correctly predict the number of criminals point-wise by one for the stationary hotspot case and by approximately two in the non-stationary case when

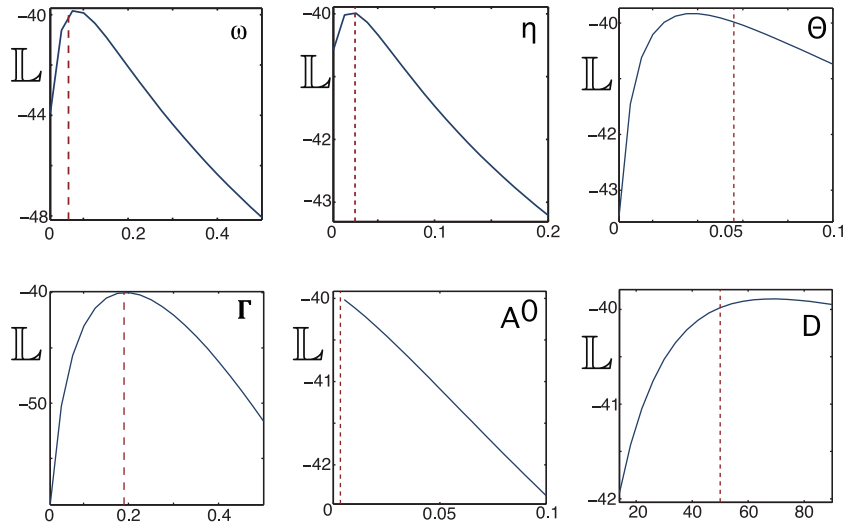


FIGURE 3. Log-likelihood function plots for the stationary hotspot case with all parameters fixed at “truth” and just one parameter varied. The vertical dashed lines denote the ‘true’ parameter values.

looking at the strong difference, errors. This is an exceedingly good approximation given that we are comparing individual realisations of a stochastic ABM with the PDE model approximation: given the discrete nature of the number of criminals in the ABM one would expect to get the precise locations of individual criminals wrong due to rounding alone. However, if we look at the weak difference between the ABM and PDE criminal densities the results are even better. Hence, we believe the PDE (3.7) simulation between attacks to sufficiently accurately represent the ABM dynamics. For the computation of the log-likelihood function, we use the fields A and ρ generated from the PDE model. Since the log-likelihood function (3.13) involves a log-average of *all* the crime-rate intensities over all the attacks and a space-time average over the crime-rate fields only the weak error matters since we just need the average crime rate at each spatial location. Hence, we expect the log-likelihood function (3.13) to be a good measure of how close the PDE is to the truth.

4.2 Data assimilation

In this section, we will look at how to maximise (3.13) over the set \mathbf{q} of model parameters; specifically, $\mathbf{q} = \{\omega, \eta, \Theta, \Gamma, A^0, D\}$ in the special case where we know in advance the initial conditions for the PDE simulations.

We will first look at the stationary hotspot case. We assimilated approximately 800 attacks to construct the log-likelihood function. In Figure 3, we plot the log-likelihood function as we vary just one parameter keeping all the others fixed at “truth”. Here, we see that the log-likelihood function does a good job of estimating the parameters with the maxima of \mathbb{L} being close to the true parameter values. However, it should be noted that due to the discrepancy between the agent-based and PDE models, the optimal parameter values for the PDE model are not necessarily the same as the true parameter values

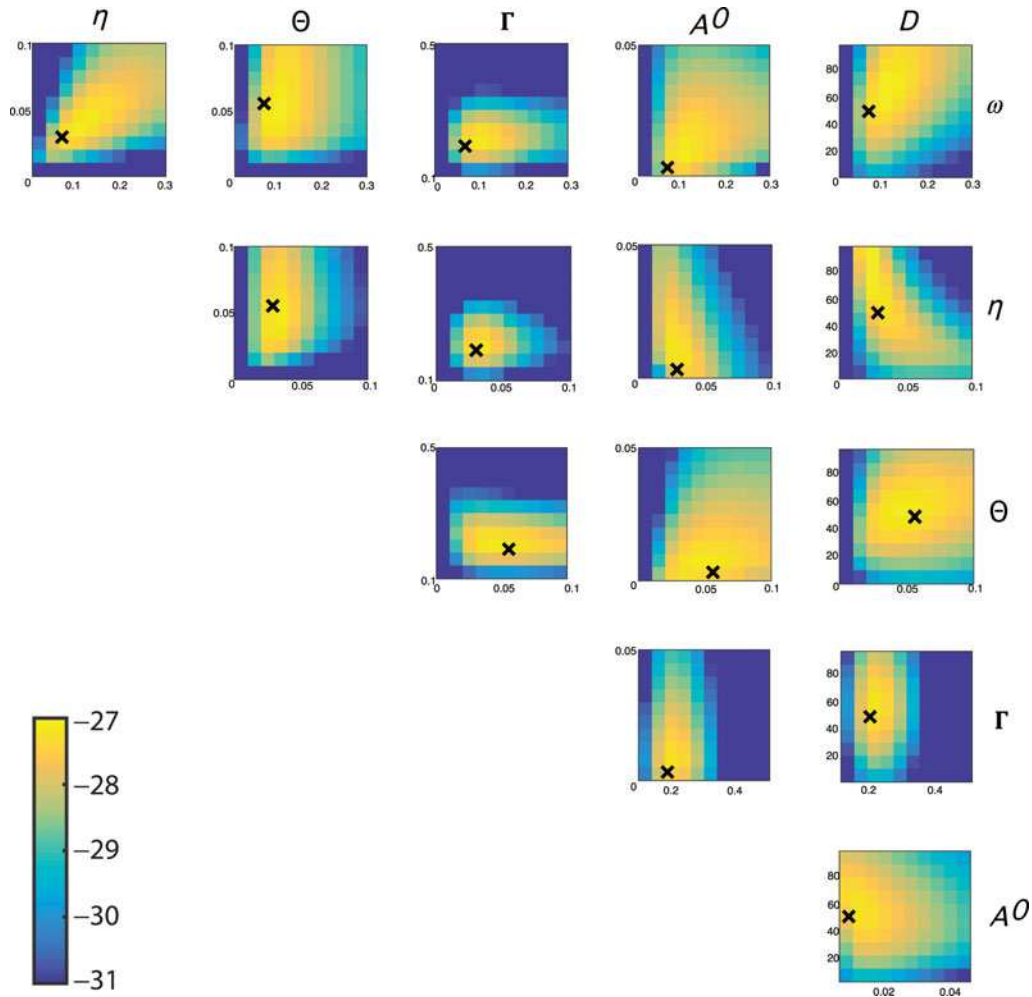


FIGURE 4. Log-likelihood function plots for the stationary hotspot case with all parameters fixed at “truth” and varying two parameters. The black crosses denote the ‘true’ parameter values.

used in the truth run. We have found that the maxima of the log-likelihood function do vary slightly for independent runs of the simulated attack data. In particular, if there are multiple crime hotspots, then the log-likelihood function is maximal closer to the true parameter values.

Figure 3 shows the values of the likelihood function evaluated for the i th element of \mathbf{q} , denoted by \mathbf{q}_i , while setting \mathbf{q}_j , $j \neq i$, to the true parameter values. In Figure 4, the (i, j) th off-diagonal blocks for $i < j$ show the values of the likelihood function evaluated on the 11×11 grid points for the parameters \mathbf{q}_i and \mathbf{q}_j , while setting the others to the true parameter values (i.e. two parameters are co-varied).

The advantage of plotting the likelihood over \mathbf{q} is that one can begin to understand the parameter sensitivity or uncertainty of the model fit and how parameters are correlated. If two parameters were strongly correlated to each other, then the likelihood function

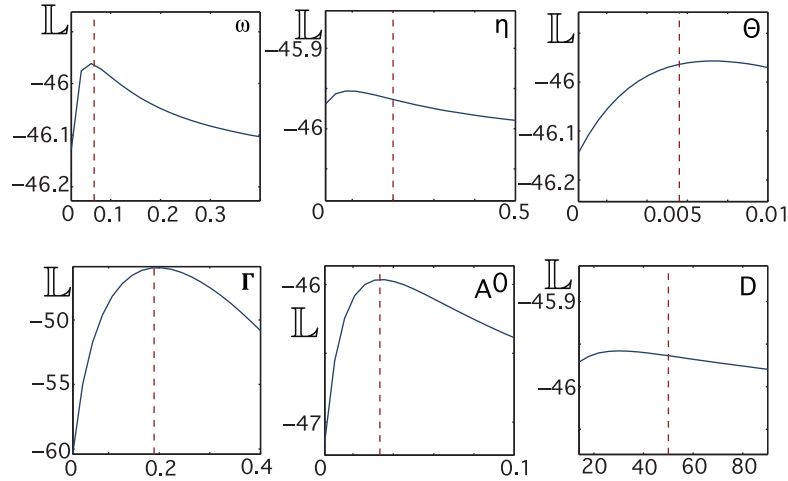


FIGURE 5. Log-likelihood function plots for the non-stationary hotspot case with all parameters fixed at “truth” and just one parameter varied. The vertical dashed lines denote the ‘true’ parameter values.

would be maximal along a diagonal line. A flat likelihood indicates that a large range of parameters could fit the data equally well. We see that the neighbourhood effect parameter, η , and the decay rate, ω , are the most sensitive parameters. As discussed in Section 2.2, from the dynamical systems analysis, we expect the maximum height of the crime rate and the width of the hotspot to be most sensitive to η and ω , and this appears to be borne out in the data assimilation. In addition, Figure 4 shows very low (pair-wise) correlations, indicating no linear dependency among these parameters. Figures 3 and 4 suggests that just taking an optimal parameter fit may not capture the full range of future scenarios and hence one needs to understand the ranges parameters can take. For the purposes of forecasting, we define the optimal parameters to be those that maximise the likelihood function in Figure 3, which is motivated by our speculation about their low correlation. One simple observation is that the log-likelihood function is convex and so a standard optimisation routine would be able to maximise the log-likelihood function over \mathbf{q} , but we do not do this here.

In Figures 5 and 6, we plot the values of the likelihood function for the non-stationary parameter region similar to that for Figures 3 and 4, respectively. There were approximately 700 attacks assimilated. In this case, the crime rate ρA is approximately equal to Γ point-wise and we see even more clearly (than for the stationary hotspot case) that large ranges of parameters would equally fit the data well. For most parameters except Γ , the likelihood function is very flat suggesting that the model (3.7) is very insensitive to these parameters. We would expect this behaviour for two reasons: (1) since Θ is much smaller than A^0 , criminal events are mostly due to the background and are therefore uncorrelated, and (2) since the PDE predicts no hotspots forming and there is a large range of parameters for which this is true. The most sensitive parameter is thus Γ .

We note that we have also tested this data assimilation in the case where we use a different initial condition for the PDE that is not identical to the ABM. In this case, we see for the large times that we assimilate for, i.e. $t \in [0, 70]$, this makes little difference as

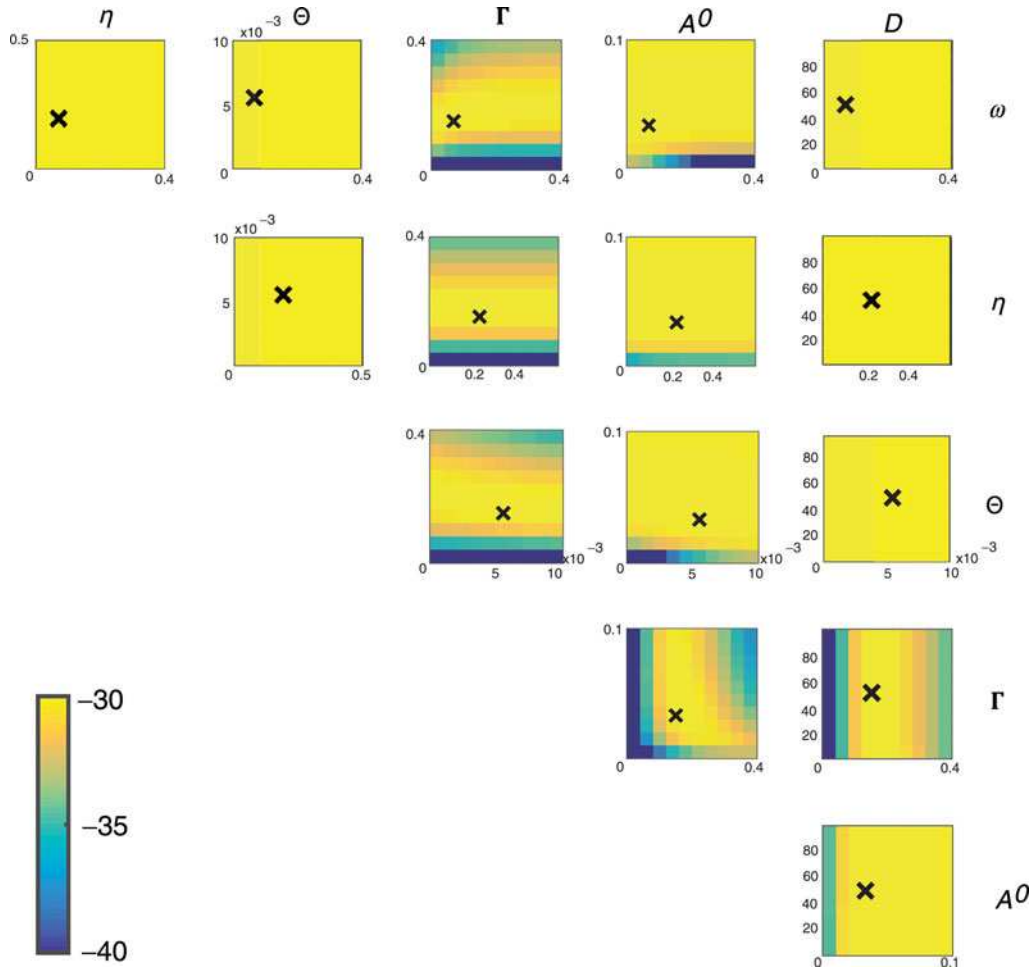


FIGURE 6. Log-likelihood function plots for the non-stationary hotspot case with all parameters fixed at “truth” and varying two parameters. The black crosses denote the ‘true’ parameter values.

the effect of the initial transient is minimal. However, for short data assimilation times, one would also have to maximise the log-likelihood function over the initial conditions as well as the parameters.

4.3 Goodness-of-fit

We vary one parameter at a time and observe the goodness-of-fit for various parameter values analogous to Figure 3. Figure 7 compares \mathcal{D}_N , averaged over all x_j , as specific parameters vary and for various values of T . For some parameters, the valleys (corresponding to better fitting parameter values) around the optimal parameter values are more noticeable for a greater value of T , hence a higher number of data points, and some of the valleys are closer to the true parameter values. In Figure 8, we show KS plots with the 95% confidence band, calculated at the node x_i where the total number of attacks up

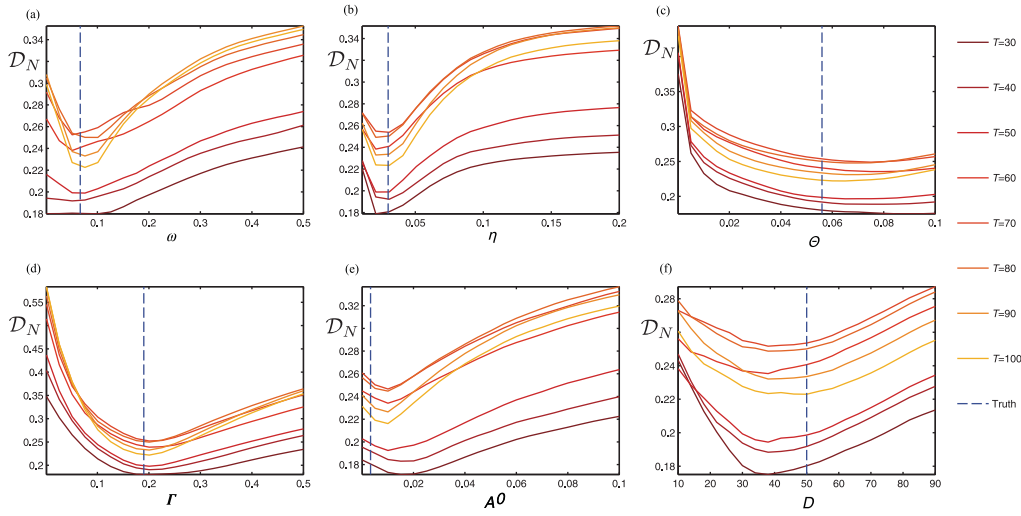


FIGURE 7. Comparing the KS statistics of each parameter, averaged over all nodes x_j and for various values of T , for the stationary hotspot case. In each plot, the vertical dashed line represents the true value of the parameter in question.

to $T = 70$ is highest. We see that all parameter values for D are within the confidence band, and the optimal value and those close to it have curves that agree very well with the 45-degree line. This is unsurprising given that the KS curve for D in Figure 7 is relatively flat for the entire parameter range. It is surprising that almost all parameter values of ω have KS plots that lie within the 95% confidence band, as its KS curve is not particularly flat. For other parameters, the KS plots corresponding to parameter values close to the optimal values are within the band, whereas those far from the optimal value are outside the band.

Similar plots for the case of the non-stationary hotspot are shown in Figures 9 and 10. In Figure 10, we show the KS plots at the node where the number of attacks up to $T = 70$ is highest. As shown in Figure 9, only parameter Γ develops a valley around the true parameter for a large enough T . This can be justified by the fact that a broad range of parameter values, except Γ , would be able to produce similar flat intensity profiles in this case, but the parameter Γ , which is equal to the overall attack rate, has to be identified correctly to fit the data well. Furthermore, since the attacks are evenly spread out over all nodes in this case, the highest number of attack at a node is smaller than the case of stationary hotspots. Therefore, the confidence band is larger than the stationary hotspot case, which is reasonable as less data should have higher uncertainty in the estimate. Thus, almost all of the KS plots, except for Γ , lie within the 95% confidence band, although they are not very close to the 45-degree line. For Γ , however, only those near the optimal values lie within the confidence band.

As with the MLE, we also plot in Figures 11 and 12 the KS statistic as we vary two parameters while keeping the other parameters fixed at their ‘true’ values. In Figure 11, we see that Γ and θ have a significant effect on the goodness-of-fit whereas the influence of

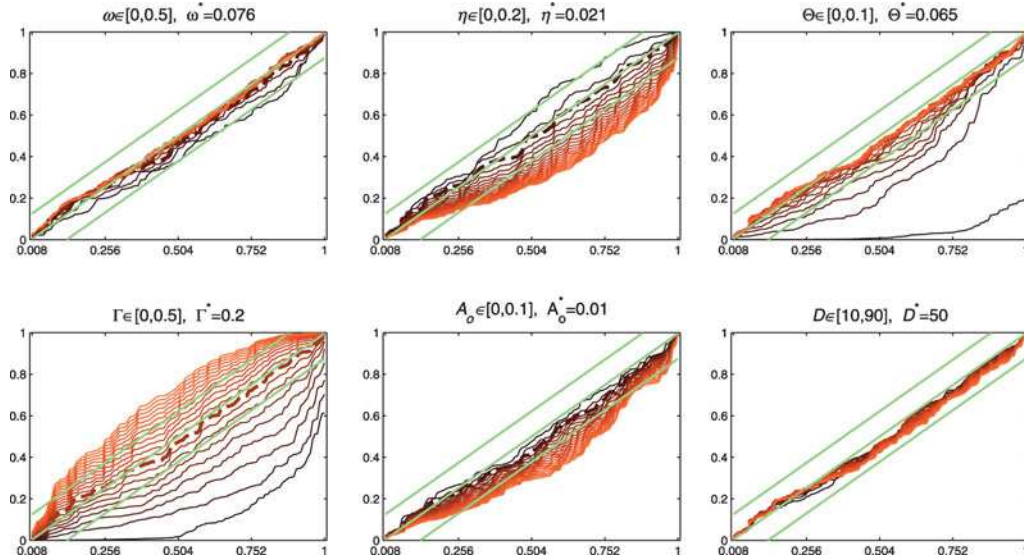


FIGURE 8. KS plots for the stationary hotspot case. We show only the plots at the node x_i where the total number of attacks up to $T = 100$ is highest. The solid 45-degree line represents the true cumulative distribution, which is bounded by the 95% confidence bounds. The thick dashed line shows the KS plot for the optimal parameter value corresponding to Figure 7; the optimal parameter values are indicated by the asterisk above each plot.

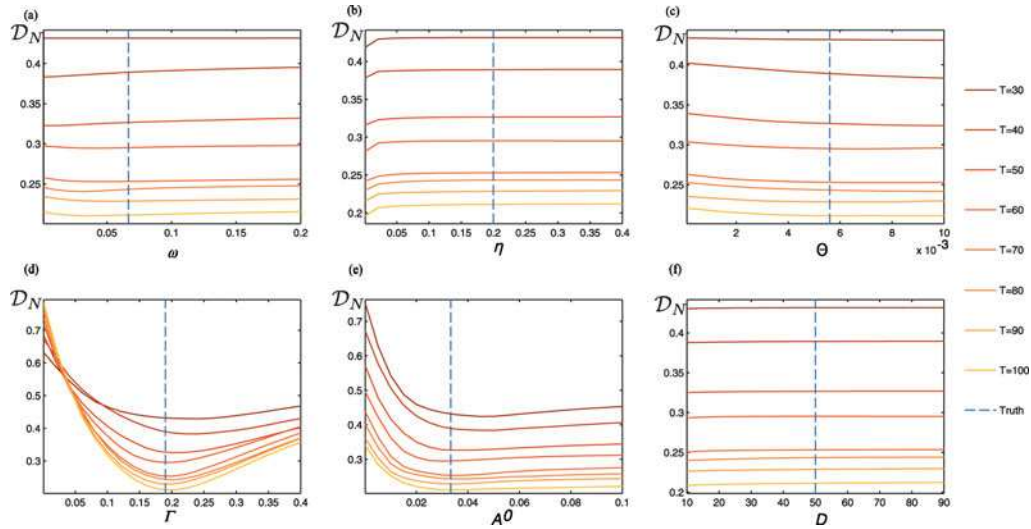


FIGURE 9. Comparing the KS statistics of each parameter, averaged over all nodes x_j and for various values of T , for the non-stationary hotspot case. In each plot, the vertical dashed line represents the true value of the parameter in question.

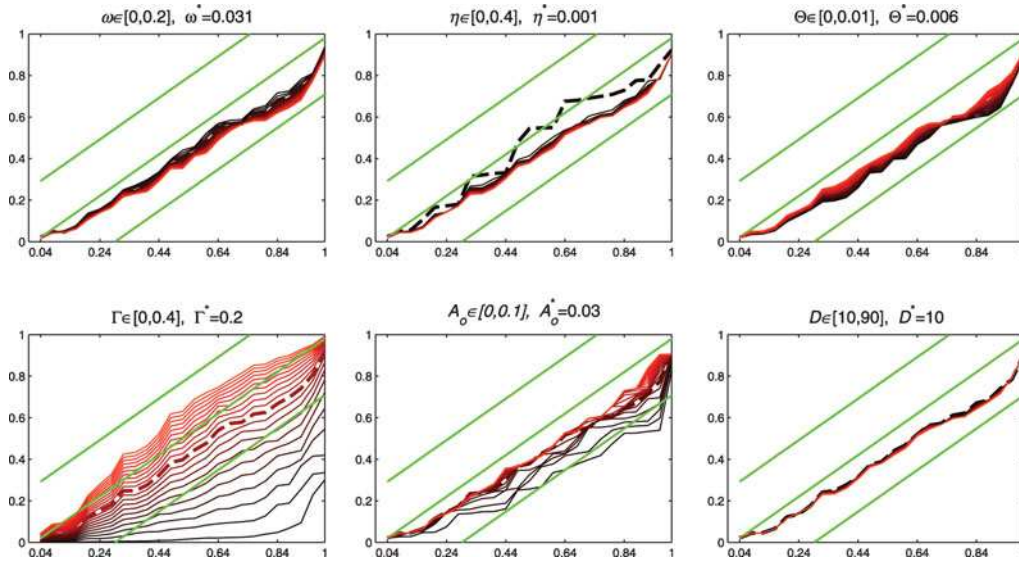


FIGURE 10. KS plots for the non-stationary hotspot case. We show only the plots at the node x_i where the total number of attacks up to $T = 100$ is highest. The solid 45-degree line represents the true cumulative distribution, which is bounded by the 95% confidence bounds. The thick dashed line shows the KS plot for the optimal parameter value corresponding to Figure 9; the optimal parameter values are indicated by the asterisk above each plot.

the other parameters is less pronounced. In Figure 12, it is clear that only Γ (the criminal generation rate in space) has any effect on the goodness-of-fit.

4.4 Scenario inference

From the previous section, it is clear that relatively large ranges of parameters fit the data well, so one needs to take this into account when providing crime forecasts. In particular, one needs to understand the range of crime rates that could feasibly be observed. One could carry out many simulations of the agent-based crime model described in Section 2 with the appropriate ranges of parameters and look at the various distributions that could be observed. This is clearly a very computationally expensive approach, so we take a more efficient strategy to carry out forecasting and comparison with the attack data at the expense of knowing the variance of the possible future outcomes.

The dimensions of the fitted parameters will depend on the units of attack data supplied (minutes/hours and metres/kilometres). In both cases, the dimensions of the parameters will matter when it comes to forecasting since one needs to know over what time and space scales the predictions apply. Since we are fitting only simulated data (and hence have no time/space scales), we will train using data over 70 time units and then forecast over 30 time units and keep the space dimensions $x \in [0, 65]$ fixed.

To make forecasts, we simulate the PDE (2.5) starting from the last spatial profiles calculated from simulation of (3.7) and stopping at the end of the forecast window. One can then carry out a comparison of this forecast with the simulation of (3.7) that incorporates the attack data in the time interval $t \in [70, 100]$.

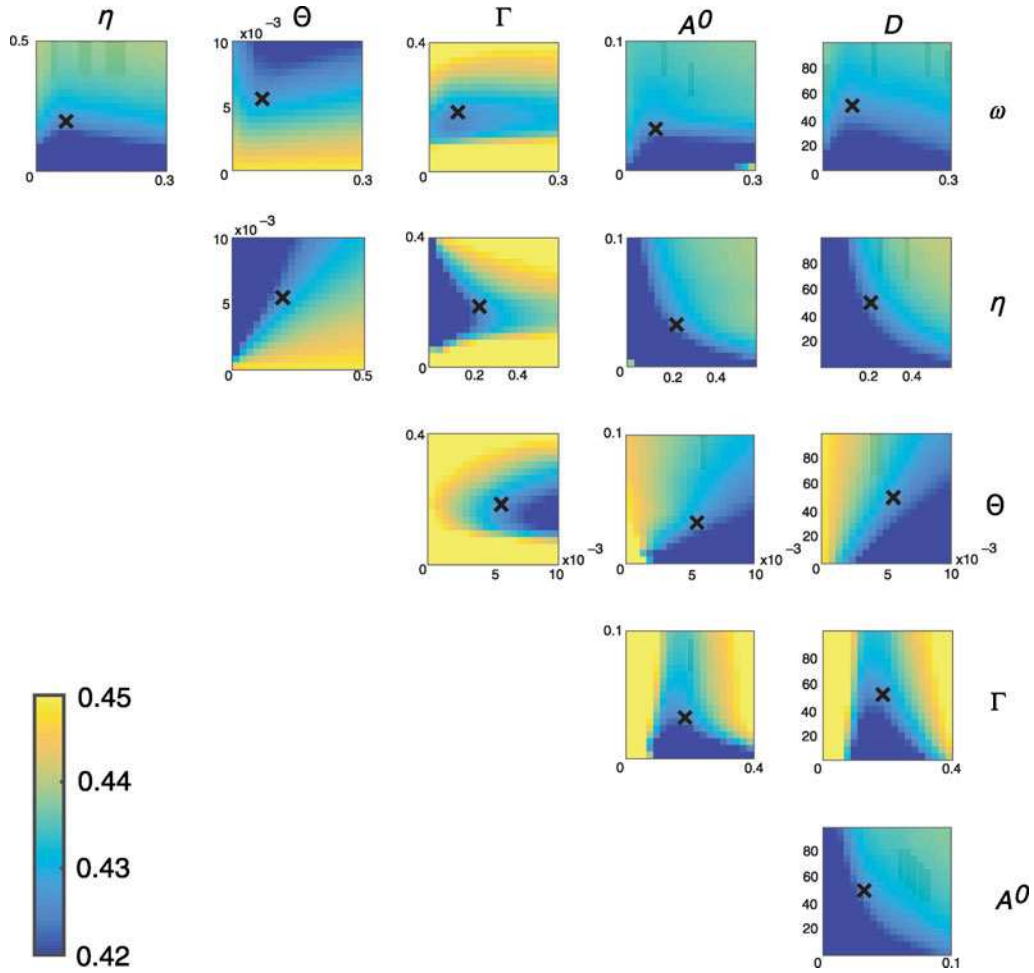


FIGURE 11. KS plots for the stationary hotspot case with all parameters fixed at “truth” and varying two parameters. The black crosses denote the ‘true’ parameter values.

In our case, we simulate the attack data from the ABM in the time range $[0, 100]$. The assimilation of the attack data in the time interval $[0, 70]$ is then done to find the appropriate parameter ranges. Using these parameter ranges, we then forecast and compare the crime rate in the time interval $[70, 100]$.

In Figure 13, we show the forecasts and a comparison between the “true” parameters and the “optimal” fitted parameters. We define the “optimal” fitted parameters to be those that maximise the log-likelihood function in Figures 3 and 5, and are as follows:

- *Stationary hotspots*: $\omega = 0.076, \eta = 0.03, \Theta = 0.0405, \Gamma = 0.2, A_0 = 1/300, D = 70$,
- *Non-stationary hotspots*: $\omega = 1/15, \eta = 0.1, \Theta = 0.007, \Gamma = 0.205, A_0 = 1/30, D = 30$.

Figures 13(a) and (d) plot the spatial profiles of the expected crime rate, ρA , at time $t = 100$. The forecasts from simulating (2.5) (i.e. where the ρA is added back into the evolution of the attractiveness field A), show that the optimal parameters predict a higher

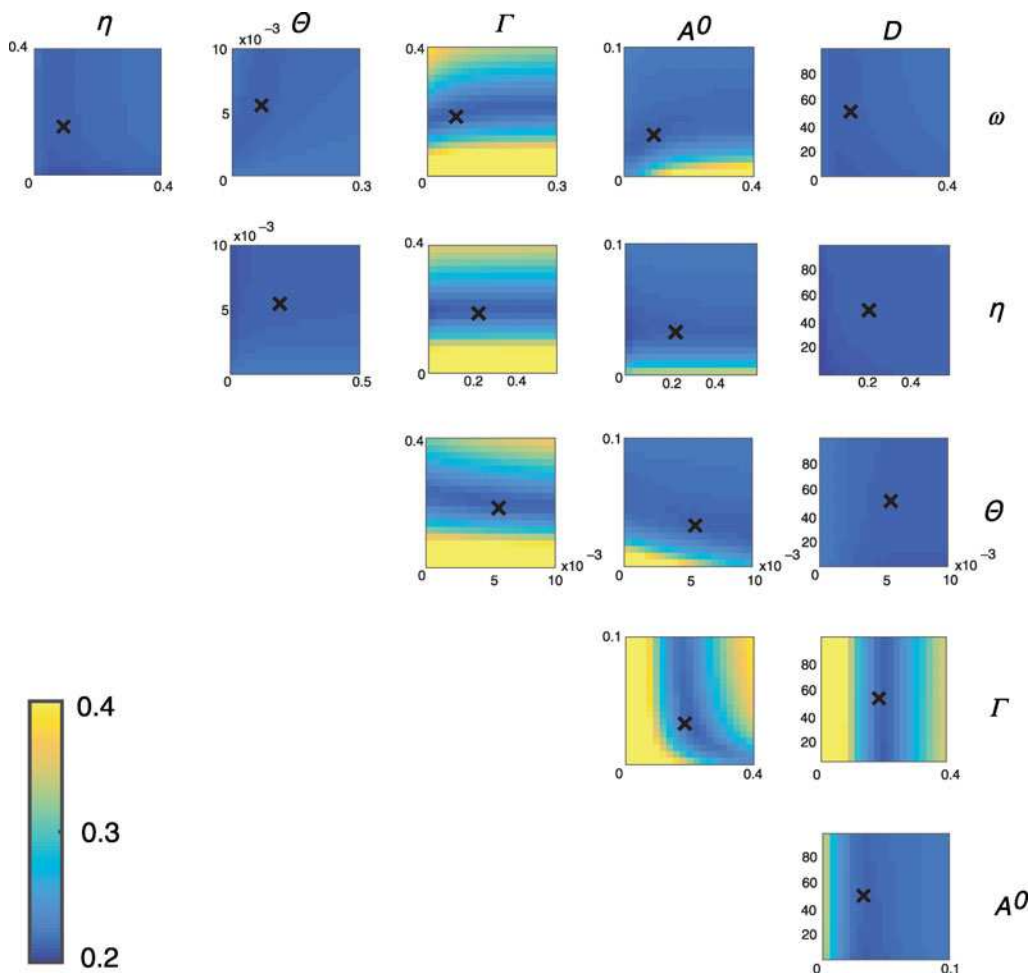


FIGURE 12. KS plots for the non-stationary hotspot case with all parameters fixed at “truth” and varying two parameters. The black crosses denote the ‘true’ parameter values.

maximum crime rate than the true parameters. We also see that the optimal parameters fit the spatial distribution of the crime rate in the stationary hotspot case better than the truth as seen in Figure 13(c) where we plot the vector 2-norm difference between the two PDE simulations (with and without attack data). We believe this to occur in the fitting since the likelihood function tries to minimise the space–time average of the crime rate between attacks and hence fitting the spatial distribution of crime rate better is likely to be optimal.

For the non-stationary case, looking at the spatially homogeneous steady states of (2.5) we find that the crime rate $\rho A = \gamma$. Hence, γ (and respectively Γ in the data assimilation step) is the only parameter that matters for long-time forecasts. The maximum crime rate is again observed to be slightly higher using “optimal” parameters than for the “truth”. But the vector 2-norm difference is slightly poorer when comparing the simulations of the modified PDE (3.7) with the “optimal” parameters versus the “true” ones (and it is

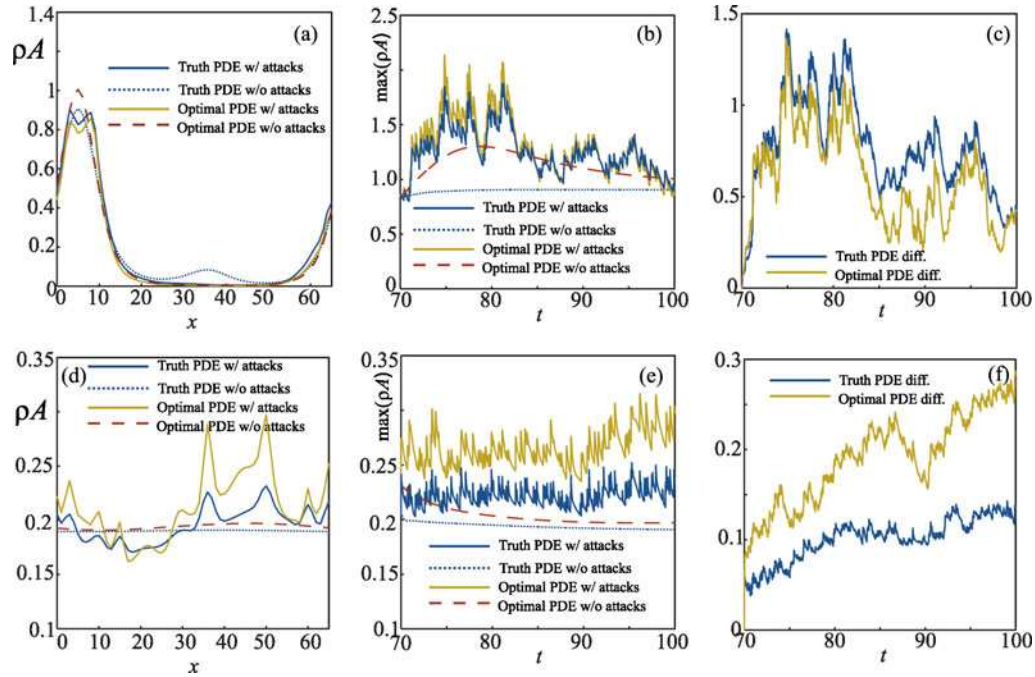


FIGURE 13. Forecast analysis for the stationary hotspot case panels (a)–(c) and the non-stationary hotspot case panels (d)–(f). In (a) and (d), we plot the spatial profiles of the expected crime rate ρA at $t = 100$ comparing both the true and fitted parameters. The blue/(yellow and red) lines denote the crime rate profiles for the true/fit parameters of the simulation of the PDEs (3.1) (i.e. with attacks in $t \in [70, 100]$ added) and (2.5) (i.e. without attacks added). In panels (b) and (e), we plot the respective maximum crime rate for both the true and fitted parameters. Panels (c) and (f) plot the vector 2-norm difference at the lattice sites between the simulations of the PDEs (3.1) (i.e. with attacks in $t \in [70, 100]$ added) and (2.5) (i.e. without attacks added).

clear the simulation of (3.7) is strongly affected by the other parameters leading to a poor vector 2-norm). The reason that the vector 2-norm is poorer for the ‘optimal’ parameters is that the simulation of the PDE (3.7) between the attacks is affected by the diffusion parameter D , and we see that the large crime rate points do not decay as fast with the ‘true’ parameters; see the solid red and blue lines in Figure 13(d)). We see from Figure 5 that Γ is the most significant parameter with the other parameters having little effect on the log-likelihood function. Hence, provided the parameters yield a stable spatially homogeneous steady state, one only needs to worry about how well Γ is fitted.

In the stationary hotspot case, one can go further by applying the singular limit analysis of Kolokolnikov *et al.* [19] described in Section 2.2 in order to understand how varying the parameters is likely to affect the shape of the crime rate forecasts. Since the maximum height of the crime rate and spatial width of the hotspots is mostly governed by ω, η, D and γ , one only needs to consider the sensitivity of the forecasts with respect to just these four parameters. The parameters that have the largest impact on the forecasts is η and D since the height of the crime rate is inversely proportional to these parameters, with the next sensitive parameter being γ (the crime rate is proportional to this parameter) and

then ω (this just governs the width of the crime rate). However, the likelihood function is flat for D and so one would expect a large range for the maximum crime rate.

5 Discussion and conclusion

The incorporation of crime attack data into dynamical systems models to provide a prediction for future crime in any sensible fashion is a highly challenging task. In this paper, we have shown how one might begin going about doing this in the special case of the urban crime models of Short *et al.* [29]. Just knowing the attack times and locations, we adapted the PDE model [29] in order to simulate an expected crime rate between attack times. Using this simulated crime rate, we then described a likelihood function that one could maximise to yield optimal parameter fits. We found that the likelihood function is rather flat for a large range of parameters, suggesting that various parameter scenarios need to be considered when forecasting crime rates. We show in the stationary hotspot case that the optimal parameters fit the spatial distribution of the crime rate better than the ‘true’ parameters. When it comes to forecasting various crime scenarios, the dynamical systems analysis of the PDE (2.5) (see for instance [3, 19, 21, 26, 29]) proves invaluable in understanding how the ranges of feasible parameter fits impact the crime rate. In the non-stationary hotspot case, the long-term forecasts for the crime rate is governed only by γ whereas in the stationary hotspot case the singular limit analysis of [3, 19] allows us to understand how the feasible parameter fits impact the crime rate distribution. It is clear that this is a very initial investigation into the fitting of dynamical systems models that really requires a more detailed analysis than that done here.

Ultimately, one would like to use this approach to yield optimal and robust (in the sense of yielding an outcome under noisy/uncertainty perturbations) policing strategies and it is here we believe that the combination of data assimilation, modelling, and dynamical systems analysis is invaluable. It is clear that parameter estimation is likely to yield large parameter regions which may be exacerbated by rapidly changing events and/or lack of data. Hence, understanding the qualitative dynamics of the region of parameter space the fitting yields will be paramount in determining how to act. For instance, if one does the fitting and finds that the parameters are in the non-stationary hotspot case then crime rate dynamics are on average equally spread over time and space, suggesting that evenly distributing one’s crime force may be a sensible strategy. However, even in this scenario one may be able to make shorter term predictions of future crimes if the estimated parameters are sufficiently accurate. On the other hand, if one finds the best region of parameter space occurs in the stationary hotspot region, then a policing strategy such as that investigated by Short *et al.* [27] or Zipkin *et al.* [33] may be more desirable.

While this has been a theoretical study using simulated attack data, our focus has been very much on highlighting and addressing the issues one would have in trying to carry out data assimilation of actual crime data to dynamical systems models. We outline several areas for further research in order to maximise the predictive power of the dynamical systems models.

Throughout this data assimilation, we have assumed that the times of the attacks are known to within a time window of the time step size of the PDE model (this could be say a day depending on the coarseness of the attack data). This is slightly less restrictive

than the usual ETAS method or Hawkes process models where the times need to also be fitted. If the attack times are not that well known, then one could add the time-points of the attacks to the optimisation problem of the MLE. This would create a significant computational overhead as many different simulations of the PDE model would be required. We highlight this as a major area for future development of efficient methods to cope with this issue.

In terms of the simulated attack data, there are several simple studies that one can do. For instance, one could look at what happens if you only know a proportion of the total number of attacks or when you add in attack data generated by another process not from the ABM. It would also be interesting to compare various different urban crime models against how they fit and predict attack data generated by different mechanisms.

Most crime models are likely to be ABMs and it would be highly desirable to use these ABMs in the data assimilation. However, there would be massive computational issues that one would need to overcome in order to use ABMs within the framework outlined in this paper. Our approach is to compute expected fields A and ρ between attacks. However, it is clear that our simulation of the criminal density between attacks needs to be greatly improved as mentioned in Section 3. The dynamical systems analysis of large-scale stochastic ABMs remains a major challenge requiring the development of new tools and techniques designed to deal with these sociological models such as stochastic bifurcation analysis; see for instance [18]. We note that it would also be interesting to investigate data assimilation for other PDE models as described in [1, 2].

It would be interesting to analyse the likelihood function based on the analysis of the partial differential equations (3.7). Since the equations (3.7) are linear between attacks, one could use the singular limit analysis approach of [3, 19] to yield estimates on the crime rate given some known/assumed properties of the attack data. It may then be possible to construct confidence intervals for the parameters under the assumption of a perfect model. This appears to be most tractable in the parameter regions investigated in this paper.

In this study, we have analysed the effect of varying parameters on the likelihood function and the KS statistic. The advantage of the KS statistic is that it does not require knowledge of the ‘truth’ in determining the goodness-of-fit. Hence, in practice one should use a likelihood function to fit the parameters and then use the KS statistic to assess the sensitivity of the fit to uncertainty in the parameters.

In theory, as the number of attack data grows, the MLE will typically converge to the optimal parameter value. Asymptotically, the shape of the likelihood function will be close to a symmetric “parabola” and have a small variance. Thus, the uncertainty of the MLE can be approximated by a normal distribution centred around the true parameter value with variance inversely proportional to Fisher’s information [30] and another method to compute confidence intervals can be constructed by a standard procedure. However, for our model, there are some issues in using Maximum Likelihood:

- The rate of an inhomogeneous Poisson process in our problem is, in a sense, “parameterised” by unknown parameter vector \mathbf{q} and the unknown initial (or current) distributions $A(x, t = 0)$ and $\rho(x, t = 0)$. In general, the unknown parameter vector \mathbf{q} may not be a constant over a long time interval. Therefore, constant parameter

values can be assumed only in a short time interval, in which case the data may be inadequate and, as a result, the likelihood may not be sharply peaked or may even fail to be concave. In such situation, the MLE may not be reliable.

- A long sequence of attack data is certainly preferred for the Maximum Likelihood method if the model accurately represents the true dynamic of the Poisson rate. In practice, there is certainly a discrepancy between the true driving forces and the model dynamic and a longer assimilation window may not yield good MLE.
- When $A(x, t = 0)$ and $\rho(x, t = 0)$ are also unknown, we have a high-dimensional problem. For example, for the experiment in this paper, the dimension of the problem will become $66 \times 2 + 6 = 138$, instead of 6. The likelihood function for a high-dimensional problem may have multiple local maxima and without a good prior knowledge of the region where parameter values with high probability lie, the optimisation problem can be very difficult. Therefore, if we have some prior knowledge about a “low-dimensional” structure where parameter values with high probability lie, we should then utilise it. The MLE, however, does not provide a good platform to incorporate such a prior information.

When it cannot be guaranteed that parameters are static over a long time interval, it is more reasonable to tune parameters sequentially, that is, when a new attack data is available, we immediately assimilate it to tune our current parameter estimates and incorporate prior knowledge of the parameters and initial state. A fully non-linear method such as particle filtering can be used for this purpose. However, particle filtering suffers from a curse of dimensionality where the computational cost becomes quickly prohibitive as the dimension grows [13]. Alternatively, the ensemble Kalman filter is less computationally demanding but only the first two moments of the uncertainty can be estimated [8]. We will investigate the applicability of both methods in future work. We highlight in this paper that developing a good forecasting analysis is a major area for future work. While we could have used the ABM to predict a crime rate and used this to compare our fits, we focused on the practical issue of how one would do a forecasting analysis when the ‘true’ crime rate is not known. This is of particular importance from a crime management perspective since one needs to know how reliable the forecasts are and perhaps, more crucially, when the forecasts are poor.

One of the major challenges facing police forces worldwide is in trying to determine if a certain policing strategy actually had an effect on crime e.g., if crime goes down was it due to something you did or just simply good luck? It is here that we believe the approach in this paper may prove most useful. The development of good models to analyse the crime data and understand various policing strategies will be crucial in answering this issue.

There is of course a very deep philosophical and societal issue in developing crime prevention and policing strategies based on modelling and data assimilation. The key objective of any criminal is likely to be to maximise their unpredictability so as to not get caught. If we develop a crime strategy based on known rules, then criminals can adapt their behaviour and hence make the strategy worthless; this is a classic problem of modelling reflexive social systems, see for instance [11]. There may be good reasons for calling for a public debate as to whether large amounts of public resources should be allocated based on this methodology. It is clear that a more iterative method of steering/managing complex

adaptive systems in the context of crime needs to be researched; see for instance [31] in the context of sustainable ecosystem development. We would argue that the work in this paper only forms a foundation for developing a small part of the steering/managing complex adaptive systems methodology.

Acknowledgements

DJBL and NS gratefully acknowledge the support of the UK Engineering and Physical Sciences Research Council for programme grant EP/H021779/1 (Evolution and Resilience of Industrial Ecosystems (ERIE)). MBS gratefully acknowledges support from the US ARO MURI Grant W911NF-11-1-0332.

The authors confirm that data underlying the findings are available without restriction. Details of the data and how to request access are available from the University of Surrey publications repository <http://epubs.surrey.ac.uk/809260/>

References

- [1] BERESTYCKI, H. & NADAL, J. P. (2010) Self-organised critical hot spots of criminal activity. *Euro. J. Appl. Math.* **21**(Special Double Issue 4–5), 371–399.
- [2] BERESTYCKI, H., RODRIGUEZ, N. & RYZHIK, L. (2013) Traveling wave solutions in a reaction-diffusion model for criminal activity. *SIAM Multiscale Model. Simul.* **11**(4), 1097–1126.
- [3] BERESTYCKI, H., WEI, J. & WINTER, M. (2014) Existence of symmetric and asymmetric spikes for a crime hotspot model. *SIAM J. Math. Anal.* **46**(1), 691–719.
- [4] BOWERS, K. J., JOHNSON, S. D. & PEASE, K. (2004) Prospective hot-spotting the future of crime mapping? *Br. J. Criminology* **44**(5), 641–658.
- [5] CHAINEY, S., TOMPSON, L. & UHLIG, S. (2008) The utility of hotspot mapping for predicting spatial patterns of crime. *Secur. J.* **21**(1), 4–28.
- [6] COX, D. R. & LEWIS, P. A. W. (1966) *The Statistical Analysis of Series of Events*, John Wiley & Sons, New York.
- [7] EVANS, A. J. (2011) *Agent-Based Models of Geographical Systems*, chapter Uncertainty and Error. Springer, Berlin.
- [8] EVENSEN, G. (2007) *Data Assimilation: The Ensemble Kalman Filter*, Springer, Berlin.
- [9] DOUGLAS FAIRES, J. & BURDEN, R. (1998) *Numerical Methods*, 2nd ed., Brooks/Cole Publishing Co., Pacific Grove, CA. With 1 IBM-PC floppy disk (3.5 inch; HD).
- [10] FIELDING, M. & JONES, V. (2012) ‘Disrupting the optimal forager’: Predictive risk mapping and domestic burglary reduction in Trafford, Greater manchester. *Int. J. Police Sci. Manage.* **14**(1), 30–41.
- [11] FLANAGAN, O. J. (1981) Psychology, progress, and the problem of reflexivity: A study in the epistemological foundations of psychology. *J. History Behav. Sci.* **17**(3), 375–386.
- [12] GILBERT, N. (2008) *Agent-Based Models*, SAGE Publications, CA.
- [13] GORDON, N. J., SALMOND, D. J. & SMITH, A. F. M. (1993) Novel approach to nonlinear/non-gaussian bayesian state estimation. *Radar Signal Process. IEE Proc. F* **140**(2), 107–113.
- [14] JOHNSON, A. & KOTZ, S. (1970) *Distributions in Statistics: Continuous Univariate Distributions*, 2nd ed., Wiley, New York.
- [15] JOHNSON, S. D. (2007) *Prospective Crime Mapping in Operational Context: Final report*. Home Office, UK.

- [16] JOHNSON, S. D., BOWERS, K. J., BIRKS, D. J. & PEASE, K. (2009) Predictive mapping of crime by promap: Accuracy, units of analysis, and the environmental backcloth. *Putting Crime in its Place*, Springer, Berlin, pp. 171–198.
- [17] KENNEDY, L. W., CAPLAN, J. M. & PIZA, E. (2011) Risk clusters, hotspots, and spatial intelligence: Risk terrain modeling as an algorithm for police resource allocation strategies. *J. Quant. Criminology* **27**(3), 339–362.
- [18] KEVREKIDIS, I. G. & SAMAEY, G. (2009) Equation-free multiscale computation: Algorithms and applications. *Annu. Rev. Phys. Chem.* **60**, 321–344.
- [19] KOLOKOLNIKOV, T., WARD, M. J. & WEL, J. (2012) The stability of steady-state hot-spot patterns for a reaction-diffusion model of Urban crime. *Discrete and Continuous Dyn. Syst. – Series B*, **19**(5), 1373–1410.
- [20] LIU, H. & BROWN, D. E. (2003) Criminal incident prediction using a point-pattern-based density model. *Int. J. Forecast.* **19**(4), 603–622.
- [21] LLOYD, D. J. B. & O’FARRELL, H. (2013) On localised hotspots of an urban crime model. *Phys. D* **253**, 23–39.
- [22] MITCHELL, L. & CATES, M. E. (2010) Hawkes process as a model of social interactions: A view on video dynamics. *J. Phys. A: Math. Theor.* **43**(4), 045101.
- [23] MOHLER, G. O. & SHORT, M. B. (2012) Geographic profiling from kinetic models of criminal behavior. *SIAM J. Appl. Math.* **72**(1), 163–180.
- [24] MOHLER, G. O., SHORT, M. B., JEFFREY BRANTINGHAM, P., PAIK SCHOENBERG, F. & TITA, G. E. (2011) Self-exciting point process modeling of crime. *J. Am. Stat. Assoc.* **106**(493), 100–108.
- [25] MOHLER, G. O., SHORT, M. B., MALINOWSKI, S., JOHNSON, M., TITA, G. E., BERTOZZI, A. L. & BRANTINGHAM, P. J. (2014) Randomized controlled field trials of predictive policing. Preprint 2015, *J. Am. Stat. Assoc.*
- [26] SHORT, M. B. & BERTOZZI, A. L. (2010) Nonlinear patterns in urban crime: Hotspots, bifurcations, and suppression. *SIAM J. Appl. Dyn. Syst.* **9**(2), 462–483.
- [27] SHORT, M. B., JEFFREY BRANTINGHAM, P., BERTOZZI, A. L. & TITA, G. E. (2010) Dissipation and displacement of hotspots in reaction-diffusion models of crime. *PNAS* **107**(9), 3961–3965.
- [28] SHORT, M. B., D’ORSOGNA, M. R., BRANTINGHAM, P. J. & TITA, G. E. (2009) Measuring and modeling repeat and near-repeat burglary effects. *J. Quant. Criminology* **25**(3), 325–339.
- [29] SHORT, M. B., D’ORSOGNA, M. R., PASOUR, V. B., TITA, G. E., JEFFREY BRANTINGHAM, P., BERTOZZI, A. L. & CHAYES, L. B. (2008) A statistical model of criminal behavior. *Math. Models Methods Appl. Sci.* **18**, 1249–1267.
- [30] WALKER, A. M. (1969) On the asymptotic behaviour of posterior distributions. *J. R. Stat. Soc. Series B (Methodological)* **31**(1), 80–88.
- [31] WALTNER-TOEWS, D. & KAY, J. (2005) The evolution of an ecosystem approach: The diamond schematic and an adaptive methodology for ecosystem sustainability and health. *Ecology Soc.* **10**(1), 38.
- [32] WANG, X. & BROWN, D. E. (2012) The spatio-temporal modeling for criminal incidents. *Secur. Inform.* **1**(1), 1–17.
- [33] ZIPKIN, J. R., SHORT, M. B. & BERTOZZI, A. L. (2014) Cops on the dots in a mathematical model of urban crime and police response. *Discrete Continuous Dyn. Syst. – Series B* **19**(5), 1479–1506.

Appendix A Connection between (3.8) and (3.7b)

Let $\kappa \equiv 2D\delta t/\delta x^2$. Then, expanding (3.8) in terms of Taylor Series up to orders δt and δx^2 , and dropping the subscript i , one obtains the following:

$$\rho + \rho_t \delta t = (1 - \kappa)\rho + \kappa A \left[\frac{\rho + \rho_x \delta x + \rho_{xx} \delta x^2 / 2}{2A + 2A_x \delta x + 2A_{xx} \delta x^2} + \frac{\rho - \rho_x \delta x + \rho_{xx} \delta x^2 / 2}{2A - 2A_x \delta x + 2A_{xx} \delta x^2} \right] + \gamma \delta t. \quad (\text{A } 1)$$

Cancelling the two ρ terms from left and right and factoring out $2A$ from the denominators, then approximating the denominators up to order δx^2 gives

$$\rho_t \delta t = -\kappa \rho + \frac{\kappa}{2} [(\rho + \rho_x \delta x + \rho_{xx} \delta x^2 / 2)(1 - A_x \delta x / A - A_{xx} \delta x^2 / A + A_x^2 \delta x^2 / A^2) + (\rho - \rho_x \delta x + \rho_{xx} \delta x^2 / 2)(1 + A_x \delta x / A - A_{xx} \delta x^2 / A + A_x^2 \delta x^2 / A^2)] + \gamma \delta t. \quad (\text{A } 2)$$

Expanding all terms and again keeping only up to order δx^2 , then dividing both sides by δt gives

$$\rho_t = \frac{\kappa \delta x^2}{2 \delta t} \left[\rho_{xx} - \frac{2\rho A_{xx}}{A} + \frac{2\rho A_x^2}{A^2} - \frac{2\rho_x A_x}{A} \right] + \gamma, \quad (\text{A } 3)$$

which, with our definition of κ above, is equivalent to (3.7b).