Approximate filtering of conditional intensity process for Poisson count data: Application to urban crime

Naratip Santitissadeekorn [a],[*], David J.B. Lloyd [a], Martin B. Short [b], Sylvain Delahaies [a]

[a] *Department of Mathematics, University of Surrey, Guildford, Surrey, UK*
[b] *Department of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA*

A R T I C L E   I N F O

A B S T R A C T

The primary focus is a sequential data assimilation method for count data modelled by an inhomogeneous Poisson process. In particular, a quadratic approximation technique similar to the extended Kalman filter is applied to develop a sub-optimal, discrete-time, filtering algorithm, called the extended Poisson–Kalman filter (ExPKF), where only the mean and covariance are sequentially updated using count data via the Poisson likelihood function. The performance of ExPKF is investigated in several synthetic experiments where the true solution is known. In numerical examples, ExPKF provides a good estimate of the "true" posterior mean, which can be well-approximated by the particle filter (PF) algorithm in the very large sample size limit. In addition, the experiments demonstrate that the ExPKF algorithm can be conveniently used to track parameter changes; on the other hand, a non-filtering framework such as a maximum likelihood estimation (MLE) would require a statistical test for change points or implement time-varying parameters. Finally, to demonstrate the model on real-world data, the ExPKF is used to approximate the uncertainty of urban crime intensity and parameters for self-exciting crime models. The Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system data is used as a case study for both univariate and multivariate Hawkes models. An improved goodness of fit measured by the Kolmogorov–Smirnov (KS) statistics is achieved by the filtered intensity. The potential of using filtered intensity to improve police patrolling prioritisation is also tested. By comparing with the prioritisation based on MLE-derived intensity and historical frequency, the result suggests an insignificant difference between them. While the filter is developed and tested in the context of urban crime, it has the potential to make a contribution to data assimilation in other application areas.

1. Introduction

A conditional intensity process (as defined in Cox and Lewis (1966)) model is a powerful tool for investigating and predicting count data (either in the form of the number of events in a given time period or a time-series of occurrence times) commonly found in a wide variety of applications such as seismology (Ogata, 1988; Ogata and Zhuang, 2006), epidemic disease outbreaks (Unkel et al., 2012; Corberán, 2012), and urban crime (Mohler et al., 2015). In the context

of urban crime, a typical intensity process is the Hawkes process (Hawkes and Oakes, 1974) where the crime rate is split into a spatially-dependent baseline rate and a self-excitation rate. In particular, the Hawkes process allows the past events to increase the rate according to the recency of the events. In the short term absence of the events, the rate can eventually decline to the baseline (a positive lower bound of the rate) which can be interpreted as the rate of events generated by exogenous factors. The main reason the Hawkes process is well-suited to describe urban crime is that its self-excitation feature can be used to quantify the near-repeat victimisation often carried out by certain perpetrators, which is a concept that is well-established theory in criminology; see for instance (Johnson et al., 1997). A commonly used procedure to identify optimal parameters for the Hawkes process, especially in the application of crime prediction, is maximum likelihood estimation (MLE) via the expectation maximisation (EM) algorithm (Lewis and Mohler, 2011; Halpin, 2013; Olson and Kathleen, 2013), which employs the fact that the Hawkes process can be considered as a branching process with immigration (Hawkes and Oakes, 1974). Once the MLE parameters are obtained, the intensity at any time t can be deterministically computed given the history of event times up until t while the projection of the intensity into the future must be stochastically simulated via one of several possible techniques (Ogata, 1981; Dassios and Zhao, 2013). In contrast to the MLE approach, a fully Bayesian uncertainty quantification technique for parameter estimation of the Hawkes process has also been developed in Rasmussen (2013).

Our work develops an approximate filtering algorithm – i.e., an algorithm to recursively estimate and track unobserved state variables at time t given the history of observations up to and including time t – for the Hawkes process or other conditional intensity processes and investigates its application to urban crime data analysis and prediction. Developing filtering algorithms to help track and quantify uncertainty in real-time for intensity process models of count data could have impacts on a range of applications such as influenza forecasting. It also offers distinct advantages over the typical (non-filtering) MLE approach, where a large batch of data is used all at once instead of sequentially assimilating one observation at a time. One such advantage is the ability of a filtering algorithm to dynamically track model parameters over time. Nonetheless, it was shown in Lewis and Mohler (2011) that the EM algorithm can be used to track a time-varying baseline parameter within a non-parametric approach for Hawkes process, but, unlike filtering approach, it still has to work with a sufficiently large amount of prior dataset. As for most filtering algorithms, Bayesian principle lies at the core of algorithm development. Hence, the parameters have to be treated as stochastic while the true parameters may be fixed in time. Our algorithm, however, does not actually impose that the "true" parameter must evolve stochastically (e.g. via a random walk process). The MLE approach generally makes the assumption of quasi-static model parameters. However, for real-data, the intensity process model used could be very far from perfect and we show in this work that the model parameters would have to adapt to compensate for such imperfection and to provide a better goodness of fit. On the other hand, if the intensity model happens to be the "true" model for the given data, an effective filtering algorithm should be able to not only converge to the true parameters in the case of static parameters but also automatically track parameter changes if they occur. At the same time, at least in the urban crime context, the vast amount of spatio-temporal data being collected (in the context of Chicago Police Department's CLEAR – Citizen Law Enforcement Analysis and Reporting – system approximately 1.5 million records over a 4 year period are collected) also requires that any filtering algorithm be highly computationally efficient.

A step in this direction was recently carried out by Santitissadeekorn et al. (2018) where an ensemble-based filtering algorithm for a time-series of count data was developed and applied to assimilate crime data into the 1D Hawkes process. This algorithm updates the conditional intensity for each ensemble member in a way that the update is statistically consistent with the Poisson–Gamma conjugate pair. This provides the "innovation" in terms of intensity analogous to the innovation of observation used in the Kalman filter. The innovation in terms of the conditional intensity is then used to regress other relevant parameters through their sample covariance. Therefore, the above algorithm does not require the computation of gradient or Hessian terms but its extension to a general spatio-temporal model for a history-dependent intensity process looks to be very difficult due to its use of the univariate Poisson–Gamma pair to obtain the innovation.

In this work, we develop an approximate filtering of the Poisson intensity process where the parameter vector is considered as a Markov process and is allowed to vary stochastically with time. We believe that this consideration may be more useful to some real-world data than MLE, which only estimates static parameters. An extended Poisson–Kalman filter (ExPKF) is derived following a similar idea of the extended Kalman filtering framework applied to the Poisson likelihood function and requires the computation of the gradient and Hessian of the underlying intensity models but can be applied to multivariate intensity models.

We demonstrate the ExPKF on synthetic experiments where the Hawkes process is the true model and on real Chicago crime data. With the real-world data, by allowing time-varying parameters, the filtered intensity may be tuned to fit the data better than the model with static parameters; again, this can happen if the model cannot simulate observed behaviour in the data well enough when using static parameters. Nonetheless, the filtering algorithm below does not aim to detect a change point (i.e. the time point where the parameter has a step change). Instead, we only demonstrate below that when the parameter undergoes a step change, the filtered parameter will be en-route to the new parameter at some "learning rate". For the change-point detection problem, there are several dedicated algorithms that may be used; see for instance (Caron et al., 2012; Stephens, 1994).

One of the key goals for modelling crime and analysing crime data is to give an insight into an effective strategy of reducing crimes. Several data-driven approaches to crime prediction have attempted to incorporate predictive policing into their patrolling strategies so that a limited number of officers can be most efficiently allocated at the right place and

time. For example, recent randomised field-trial conducted with police departments has shown positive results for some regions in UK and United States (Mohler et al., 2015) in helping patrol officers to identify "at the moment" hotspots. Here the term hotspot roughly refers to the region in which the crime rate is relatively high compared to its neighbours. The model used by Mohler et al. (2015) is the Epidemic Type Aftershock Sequence (ETAS) model (Mohler et al., 2015, 2011) where the model parameters are estimated by MLE via the EM algorithm. Motivated by this application, we also carry out an experimental study below to compare the goodness-of-fit as well as predictive skill of the ExPKF and MLE in the context of patrol prioritisation in a neighbourhood of Chicago.

2. Discrete-time filtering for counting process

Consider a Poisson counting process $\{N_t^j, t > 0\}$, which represents the total number of (crime) events occurring in an interval $(0, t)$ at the jth location for $j = 1, \ldots, m$. We assume that $N_0^j = 0$ a.s. and $\{N_t^j, t > 0\}$ is a counting process with independent increments conditional on the intensity λ_t^j, where $\{\lambda_t^j, t \geq 0\}$ is a non-negative function. We will assume that for all j, N_t^j are conditionally independent given λ_t^j. Following the concept of the conditional intensity process (or doubly stochastic process introduced in Cox and Lewis (1966)), for sufficiently small δt, we may assume the existence of the limit

$$\lambda_t^j := \lim_{\delta t \to 0} \frac{Pr(\Delta N_t^j = 1 | \lambda_\tau^j, t < \tau < t + \delta t)}{\delta t}, \tag{1}$$

where $\Delta N_t^j := N_{t+\delta t}^j - N_t^j$. Furthermore, the intensity is assumed to be "modulated" by an $n-$vector Markov process, $\theta_t := [\theta_t^1, \ldots, \theta_t^n]$. Thus, we consider $\lambda_t(\theta_t)$, where $\lambda_t := [\lambda_t^1, \ldots, \lambda_t^m]$.

We will only study the discrete time (as well as discrete space) setting. The count data is assumed to arrive within a time interval of length δt in which the intensity λ_τ, for $t < \tau < t + \delta t$, is assumed to be constant for each location; hence we also assume δt to be sufficiently small. Therefore, we consider the data in the form of ΔN_k, the number of events observed during the time interval (t_{k-1}, t_k) for $k = 1, 2, \ldots$. We will use the notation λ_t and θ_t for continuous-time and λ_k and θ_k discrete time. Given the sequence of data $\Delta N_k := [\Delta N_k^1, \ldots, \Delta N_k^m]$ and underlying process of θ_k, we will approximate the evolution of the posterior statistics of θ_k as well as λ_k by developing a real-time tracking algorithm.

For a sufficiently small δt, the Bernoulli probability in (1) is well-approximated by the Poisson probability. Therefore, under the conditional independence assumption, the probability of observing ΔN_k is given by

$$Pr(\Delta N_k | \lambda_k) \propto \prod_{j=1}^m (\lambda_k^j \delta t)^{\Delta N_k^j} \exp(-\lambda_k^j \delta t). \tag{2}$$

Suppose that the prior distribution of θ_k is modelled by a multivariate normal distribution with mean $\bar{\theta}_{k|k-1}$ and covariance $\mathbf{P}_{k|k-1}$, denoted by $\theta_k \sim \mathcal{N}(\bar{\theta}_{k|k-1}, \mathbf{P}_{k|k-1})$. Following the Bayes rule and omitting the terms irrelevant to θ_k, the posterior distribution is proportional to

$$p(\theta_k | \Delta N_{1:k}) \propto \prod_{j=1}^m (\lambda_k^j \delta t)^{\Delta N_k^j} \exp(-\lambda_k^j \delta t) \times \exp\left(-\frac{1}{2}(\theta_k - \bar{\theta}_{k|k-1})^T \mathbf{P}_{k|k-1}^{-1}(\theta_k - \bar{\theta}_{k|k-1})\right), \tag{3}$$

where $\Delta N_{1:k} := [\Delta N_1, \ldots, \Delta N_k]$. In order to have a recursive formula, we decide to make a normal approximation to the posterior distribution. In other words, we require the posterior distribution to have the form:

$$p(\theta_k | \Delta N_{1:k}) \propto \exp\left(-\frac{1}{2}(\theta_k - \bar{\theta}_k)^T \mathbf{P}_k^{-1}(\theta_k - \bar{\theta}_k)\right). \tag{4}$$

To this end, we expand the likelihood function (2) up to the second-order term, which will lead to a normal posterior distribution after combining with the normal prior distribution. Taking log of both sides of (3) gives

$$-\frac{1}{2}(\theta_k^T \mathbf{P}_k^{-1} \theta_k - 2\theta_k^T \mathbf{P}_k^{-1} \bar{\theta}_k) = \sum_{j=1}^m \left[\Delta N_k^j \log(\lambda_k^j \delta t) - \lambda_k^j \delta t \right]$$
$$-\frac{1}{2}(\theta_k - \bar{\theta}_{k|k-1})^T \mathbf{P}_{k|k-1}^{-1}(\theta_k - \bar{\theta}_{k|k-1}) + C, \tag{5}$$

where C is a constant collecting all terms irrelevant to θ_k. Now we carry out a Taylor expansion around $\bar{\theta}_{k|k-1}$ to the right-hand side of (5) up to the quadratic term so that we may compare both sides term by term. It is straightforward to check that the quadratic term is given by:

$$\frac{1}{2}\theta_k^T \left(-\mathbf{P}_{k|k-1}^{-1} - \sum_{j=1}^m \left[\left(\frac{\partial \log \lambda_k^j}{\partial \theta_k}\right)^T \left(\frac{\partial \log \lambda_k^j}{\partial \theta_k}\right) \lambda_k^j \delta t - (\Delta N_k^j - \lambda_k \delta t)\frac{\partial^2 \log \lambda_k^j}{\partial \theta_k^2} \right] \right) \theta_k. \tag{6}$$

The linear term is given by:

$$\theta_k^T \left(\mathbf{P}_{k|k-1}^{-1} \bar{\theta}_{k|k-1} + \sum_{j=1}^{m} \left[\left(\frac{\partial \log \lambda_k^j}{\partial \theta_k} \right)^T (\Delta N_k^j - \lambda_k^j \delta t) \right] \right). \tag{7}$$

By comparing (6) to the first term on the left-hand side of (5) and likewise (7) to the second term, we obtain the mean and covariance update equation (8). The posterior mean $\bar{\theta}_k$ and covariance $\mathbf{P}_k$ at time t_k are given by:

$$\mathbf{P}_k^{-1} = \mathbf{P}_{k|k-1}^{-1} + \sum_{j=1}^{m} \left[\left(\frac{\partial \log \lambda_k^j}{\partial \theta_k} \right)^T \left(\frac{\partial \log \lambda_k^j}{\partial \theta_k} \right) \lambda_k^j \delta t - (\Delta N_k^j - \lambda_k \delta t) \frac{\partial^2 \log \lambda_k^j}{\partial \theta_k^2} \right]$$

$$\bar{\theta}_k = \bar{\theta}_{k|k-1} + \mathbf{P}_k \sum_{j=1}^{m} \left[\left(\frac{\partial \log \lambda_k^j}{\partial \theta_k} \right)^T (\Delta N_k^j - \lambda_k^j \delta t) \right], \tag{8}$$

where all derivatives are evaluated at $\theta_k = \bar{\theta}_{k|k-1}$. The update equation (8) will be referred to as the extended Poisson–Kalman filter (ExPKF). In the case that we have $\Delta N_k^j = 0$ for all j, the covariance update equation is greatly reduced to $\mathbf{P}_k = \mathbf{P}_{k|k-1}$. Also, if we have that

$$\left(\frac{\partial \log \lambda_k^j}{\partial \theta_k} \right)^T \left(\frac{\partial \log \lambda_k^j}{\partial \theta_k} \right) = - \frac{\partial^2 \log \lambda_k^j}{\partial \theta_k^2}, \tag{9}$$

an efficient computation of the covariance matrix in (8) can be achieved without a matrix inversion by the rank-1 update algorithm. To be more specific, we can rewrite the covariance update equation in (8) to

$$\mathbf{P}_k^{-1} = \mathbf{P}_{k|k-1}^{-1} + \sum_{j=1}^{m} h_j h_j^T, \tag{10}$$

where $h_j = \sqrt{\Delta N_k^j} \left(\frac{\partial \log \lambda_k^j}{\partial \theta_k} \right)^T$. Recall that if $\mathbf{B} = \mathbf{A} + hh^T$, where $\mathbf{B}$ and $\mathbf{A}$ are matrices and h is a vector, then we can find $\mathbf{B}^{-1}$ based on $\mathbf{A}^{-1}$ by

$$\mathbf{B}^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1} hh^T \mathbf{A}^{-1}}{1 + h^T \mathbf{A}^{-1} h}. \tag{11}$$

Therefore, by setting $\mathbf{H}_0 = \mathbf{P}_{k|k-1}^{-1}$, $\mathbf{H}_1 = \mathbf{H}_0 + h_1 h_1^T$, $\mathbf{H}_2 = \mathbf{H}_1 + h_2 h_2^T, \ldots, \mathbf{H}_m = \mathbf{P}_k^{-1}$, we can efficiently compute $\mathbf{P}_k$ by recursively applying (11) to (10).

As mentioned earlier, δt should be small enough to ensure that λ is approximately constant within the interval. Thus, if δt is fixed while the occurrence of events is rare (i.e. $\Delta N_k^j = 0$ for most of j and k), the computational cost of the filtering algorithm can be inefficient in general. Nevertheless, if the covariance update is reduced to just a rank-one update, having a small δt does not significantly add computational cost since only the mean requires the update but the covariance is unchanged. If the rank-one update is not the case, a very small δt can significantly increase the computational cost, especially if one wants to compare with that of MLE. It is not easy, however, to compare computational cost between a (non-filtering) MLE and filtering algorithm since they are operating in different ways of incorporating the data, each of which has their own advantages and disadvantage in different situations.

3. Numerical experiment: Toy model

Consider a toy model for an inhomogeneous Poisson process in which the conditional intensity function is given by

$$\lambda(t) = \alpha \exp(-\beta t), \tag{12}$$

where α and β are unknown parameters. In this case, we have $m = 1$, $n = 2$ and $\theta_k := [\alpha_k, \beta_k]$. We assume that θ_k evolves by a random walk process $\theta_{k+1} = \theta_k + \eta$ where $\eta \sim \mathcal{N}(0, \mathbf{Q})$; hence, normality is preserved under this linear dynamic. In particular, $\bar{\theta}_{k|k-1} = \bar{\theta}_{k-1}$ and $\mathbf{P}_{k|k-1} = \mathbf{P}_{k-1} + \mathbf{Q}$. The ExPKF equation (8) becomes

$$\mathbf{P}_k^{-1} = \mathbf{P}_{k|k-1}^{-1} + \begin{pmatrix} 0 & -t \exp(-\beta t) \\ -t \exp(-\beta t) & t^2 \alpha \exp(-\beta t) \end{pmatrix} \delta t + \begin{pmatrix} -1/\alpha^2 & 0 \\ 0 & 0 \end{pmatrix} \Delta N_k$$

$$\bar{\theta}_k = \bar{\theta}_{k|k-1} + \mathbf{P}_k \cdot [\alpha^{-1} - t]^T (\Delta N_k - \alpha_k \exp(-\beta_k t) \delta t). \tag{13}$$

In the experiment, the covariance matrix $\mathbf{Q}$ is a diagonal matrix $\mathbf{Q} = \begin{pmatrix} 0.04 & 0 \\ 0 & 10^{-6} \end{pmatrix}$. The ExPKF equations (8) can be readily implemented for this problem. Note that too large a value of $\mathbf{Q}$ can result in too large noise that can greatly divert the estimate from the true value, while too small a value of $\mathbf{Q}$ would make the estimate evolve too slowly and probably fail
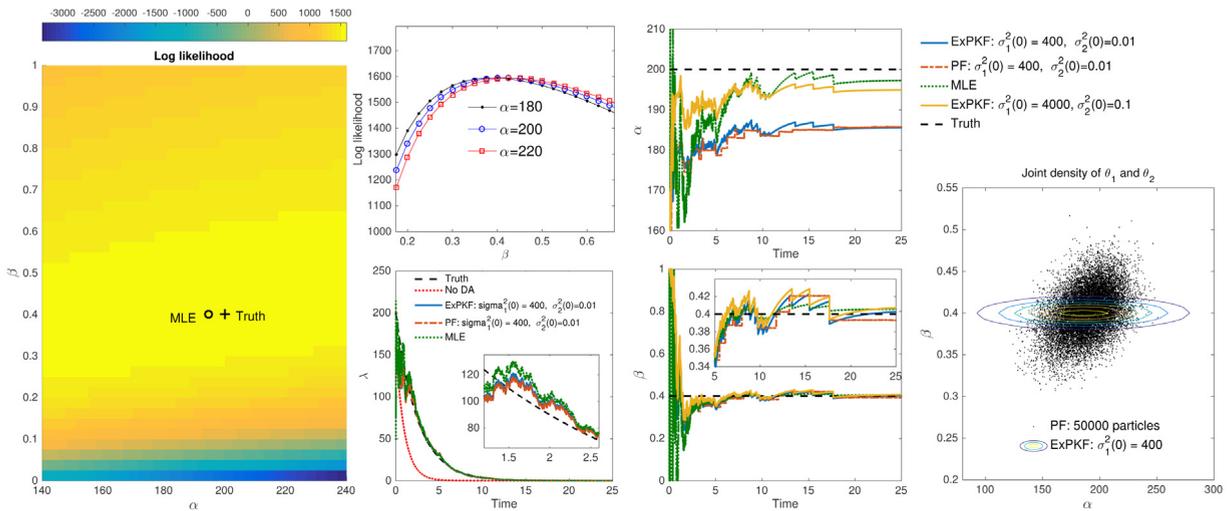
Fig. 1. [Leftmost] Likelihood function of α and β for the case of fixed parameters. The MLE using all data is shown in a circle mark. [Second, top] The slices along the direction of α emphasise the fact that the likelihood is very "flat" for α. [Second, bottom] Comparing the evolution of the mean of λ for different algorithms, all of which agree well. [Third, top] Comparison of the evolution of the mean of α for different algorithms and prior statistics. Notice the change of ExPKF toward MLE as the prior becomes more diffusive. [Third, bottom] Comparison of the evolution of the mean of β for different algorithms and prior statistics. [Rightmost] PF-based sample distribution and posterior statistics after assimilating all the data.

to track any parameter switch. We choose **Q** by trial and error and find similar results for other values of **Q** in the same orders of magnitude.

We consider two cases: (1) fixed parameters and (2) step change in a parameter. In the first case, the true parameters are $\theta_k = [200, 0.4]$ for all k and the data is generated in the time interval $[0, 25]$ with $\delta t = 0.0005$; hence, 50,000 data assimilation steps are carried out. A value $\delta t = 0.0005$ is chosen to be small enough so that the true intensity is approximately constant over the interval. In the second case, we use $\theta_k = [200, 0.4]$ during the time interval $[0, 12.5]$, after which it is changed to $\theta_k = [200, 0.2]$ in the interval $(12.5, 25]$. In order to validate the inference quality of the ExPKF, we will compare the results with the particle filter (PF), which is known to converge to the true posterior distribution in a large sample size limit (Crisan and Doucet, 2002). In addition, we also compare the result with MLE, which is run at every δt step as done in ExPKF and PF, but the MLE algorithm uses all historical data up to the current time while ExPKF and PF use only the current data. The computational cost for PF is in general prohibitive since it suffers from the curse of dimensionality, but for this two-dimensional inference problem the sample size can be taken large enough to ensure convergence to the true parameters; we use 50 000 particles since the result does not significantly change for larger numbers of particles. As for MLE, under some regularity conditions and assumption of fixed parameters, it is known to converge to the true parameter of the model as the number of data points becomes sufficiently large.

The results for case (1) are shown in Fig. 1. We first investigate the likelihood function of this problem and find that while it concentrates around the true β, it is very flat for α. The initial prior for θ at $t = 0$ is a normal distribution with mean $\bar{\theta} = [160, 1]$ and covariance matrix $P_0 = \begin{pmatrix} 400 & 0 \\ 0 & 0.01 \end{pmatrix}$. It is clear that the mean estimate of β for all algorithms are nearly the same and converges to the true parameter. However, the mean estimate of α is clearly affected by the flat likelihood in α. The MLE for α is closer to the true parameter than the PF and ExPKF, which is expected since the initial prior is relatively small compared to the spread of the likelihood function; hence, the estimate is pulled toward the initial guess. Thus, when the initial covariance is changed to $P_0 = \begin{pmatrix} 4000 & 0 \\ 0 & 0.1 \end{pmatrix}$, which is more diffusive than before, the estimates obtained from MLE and ExPKF become very close. This is a common feature of Bayesian inference, in which the mean estimate tends toward the MLE for a very diffusive prior. We also investigate the Bayesian quality of the ExPKF by comparing the sample distribution of the gold-standard PF with the posterior statistics at the final time step. It is clear that the PF-based sample mean agrees very well with the normal posterior mean of ExPKF. However, the sample spread in the direction of β is much greater than ExPKF and contains a positive correlation between α and β while the normal posterior of ExPKF is nearly zero.

For the second case where there is a step change for β at $t = 12.5$, the results are shown in Fig. 2. For both ExPKF and PF, we use a diffusive prior statistics $P_0 = \begin{pmatrix} 4000 & 0 \\ 0 & 0.1 \end{pmatrix}$ for θ. The three methods agree well until the step change of β. Both PF and ExPKF can quickly track the step change of β but MLE is unable to do so at a fast enough speed. The evolution of the ExPKF mean of α is also similar to that of PF even after the step change. However, the MLE drops quickly from the region near the true α. Most importantly, unlike ExPKF and PF, MLE fails to create a jump of the intensity rate λ at the change point. As for the uncertainty analysis, the sample distribution of PF presents more spreading in the direction of β than the normal posterior distribution of ExPKF. However, it is worth pointing out that the MLE can actually be implemented for a pre-specified sliding window and combined with an online change-point detection (Caron et al., 2012; Stephens,
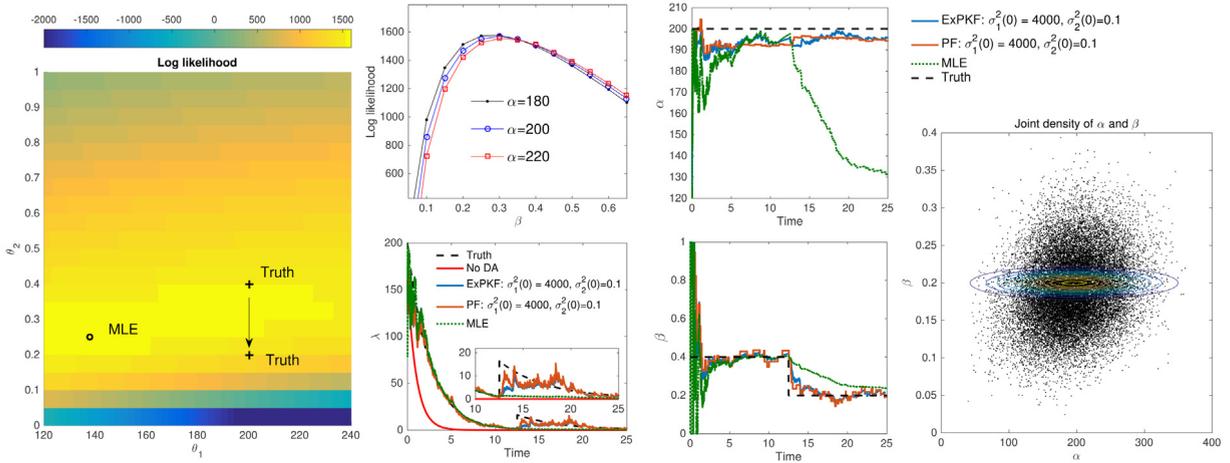
Fig. 2. [Leftmost] Likelihood function of α and β for the case of a step change of β. The MLE using all data is shown in a circle mark and the arrow indicates the step change of true parameter. [Second, top] The slices along the direction of α emphasise the fact that the likelihood is very "flat" for α. [Second, bottom] Comparing the evolution of the mean of λ for different algorithms, all of which agree well. [Third, top] Comparison of the evolution of the mean of α for different algorithms and prior statistics. [Third, bottom] Comparison of the evolution of the mean of β for different algorithms and prior statistics. [Rightmost] PF-based sample distribution and posterior statistics after assimilating all the data.

1994) to tests statistically for an abrupt parameter change in a real-time manner, which will provide a more suitable implementation of MLE for an online algorithm. In the above comparison, only the standard (non-sequential) version of MLE is concerned.

4. Numerical experiment: multivariate Hawkes process

In this section, we consider a crime model on a discrete space represented by m lattice nodes. Assuming that the crime intensity is constant within a given small time step δt, a discrete-time crime intensity rate $\lambda_j(k)$ at a node $j \in \{1, \ldots, m\}$ at a time step k (i.e. within the time interval $[k\delta t, (k+1)\delta t)$) follows a linear dynamical system:

$$\lambda_j^{k+1} = \mu_j + (1 - \beta_j \delta t)(\lambda_j^k - \mu_j) + \alpha_j \Delta N_j^k + \sum_{j' \in \mathcal{N}(j)} \alpha_{j,j'} \Delta N_{j'}^k, \qquad (14)$$

where $\lambda_j^0 = \mu_j$, $N_j^0 = 0$ for all j, and $\mathcal{N}(j)$ is the set of all nodes that are neighbours of node j. The above model assumes that at each location j, the intensity rate decays to the baseline $\mu_j \geq 0$ at the rate given by the parameter $\beta_j > 0$ and is self-excited by the number of events at the step k based on parameter $\alpha_j > 0$ as well as "cross-node" excitation induced by the number of events in the neighbourhood of j, $\Delta N_{j'}^k$, based on the parameter $\alpha_{j,j'} > 0$. Thus the above 2D model is nothing but a discrete-time version of the multivariate Hawkes process.

For simplicity, we will assume that $\alpha_{j,j'} = \alpha_c$ for all (j, j') and $\beta_j = \beta$ for all j. The integrated form of the above dynamic is then given by

$$\lambda_j^k = \mu_j + \sum_{\ell=1}^{k-1} \gamma^{k-\ell-1} \alpha_j \Delta N_j^k + \sum_{\ell=1}^{k-1} \sum_{j' \in \mathcal{N}(j)} \gamma^{k-\ell-1} \alpha_c \Delta N_{j'}^k, \qquad (15)$$

where $\gamma \equiv (1 - \beta \delta t)$. To test the ExPKF for the above intensity process, we generate a synthetic data according to (14) for $m = 5$ nodes, with $\alpha_c = 0.25$, all $\mu_j = 1$, and all $\alpha_j = 1$, except that μ_3 and α_4 jump to 2 and 1.5, respectively, at the midpoint of the simulation. The neighbourhood of a node is given by $\mathcal{N}(1) = 2$, $\mathcal{N}(j) = \{j - 1, j + 1\}$ for $j = 2, 3, 4$, and $\mathcal{N}(5) = 4$. The decay parameter $\beta = 2$ is assumed to be known; hence $\theta = [\mu_1, \ldots, \mu_5, \alpha_1, \ldots, \alpha_5, \alpha_c] \in \mathbb{R}^{11}$ in (8). Under this assumption the condition (9) for the rank-1 update is satisfied and computational cost to find P_k is greatly reduced. We use $\delta t = 0.01$, $P_0 = 0.01 I_{11}$ and $\mathbf{Q} = 10^{-6} I_{11}$ (i.e. the parameters are evolved in time by a random walk). The result obtained from ExPKF is shown in Fig. 3. It is clear that the ExPKF algorithm is able to converge to the vicinity of the true parameters and adapt to the step change. However, while the estimate of α quickly adapts to the step change, the estimate of μ takes a longer time to adapt to the change. The errors in the tracked intensity due to incorrect initial parameters are also reduced and become stable, but we notice a slight increase in this error due to the step change at the middle of the time window. In order to gain some understanding for how the algorithm adapts to the parameter jump, we also investigate the convergence of the covariance. The result shows that the analysed covariance eventually becomes stable and is quite robust to the change of initial covariance P_0, see Fig. 4. The structure of the covariance highlights the negative correlation between μ and α, especially at the intermediary nodes, and shows only negligible parameter
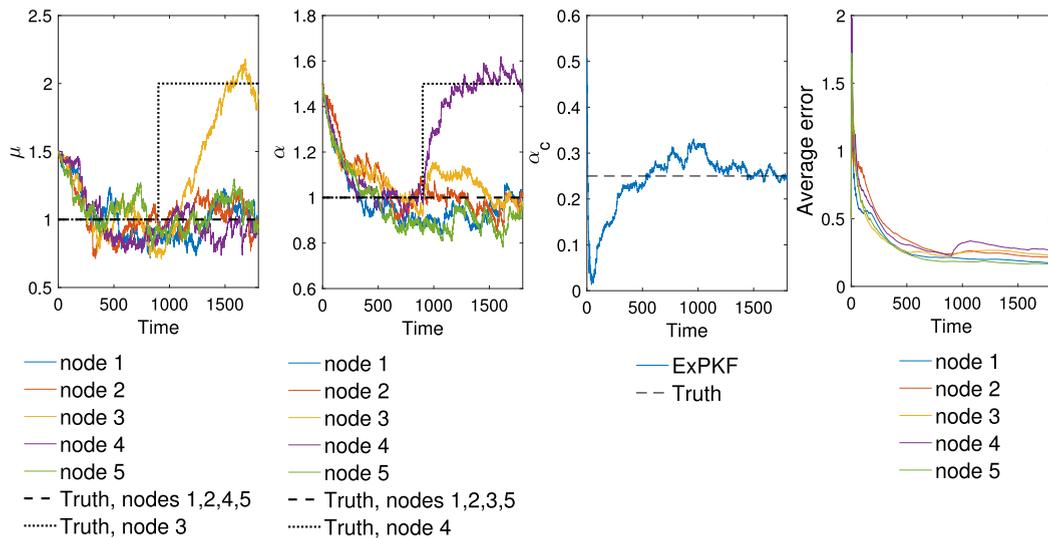
Fig. 3. Estimation of parameters for the model (15) given that $\beta = 2$ is known.
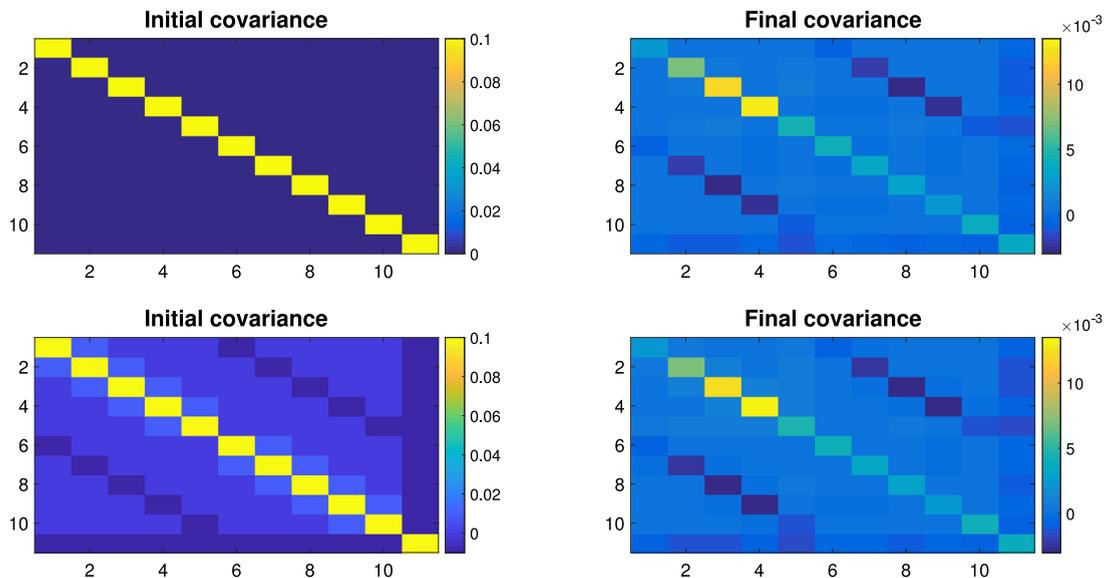
Fig. 4. The covariance after the final data assimilation step for different initial covariance P_0.

correlation among the nodes. Thus, if α is increased during a data assimilation step, the algorithm will tend to decrease μ at the next DA step.

We now assume that the parameter β is incorrectly identified. In particular, we assume $\beta = 3$ and repeat the above experiment. The result in Fig. 5 shows that the step change can still be detected and since $\beta = 3$ is larger than the true value, the parameter estimate compensates for this misidentification by decreasing both μ and α in a way that a reduction of intensity error over time is still achieved. In fact, a numerical investigation of LA gang violence data in Santitissadeekorn et al. (2018) showed, in the setting of maximum likelihood method, that a wide range of β actually gives a large likelihood close to the maximum while the optimal value of the other parameters slightly changes within this range; in particular, the hypersurface of parameters with the P-value of Kolmogorov–Smirnov statistics above 0.2 contains a range of β varying from 0.5 to 16, which is shrunk to $\beta \in (0.5, 8)$ for the threshold P-value of 0.1. For this experiment, the relative errors (i.e. the absolute error normalised by the true value) averaged over time and all cells are reported in Table 1 using the same initial setting for all values of β. Nevertheless, if the estimation of β is highly desired, the Hessian term for (14) as required in (8) can still be analytically derived. So, the main computation cost at each iteration is the matrix inversion, which costs $o(n^3)$ for n unknown parameters, while the rank-one update would cost only $o(n^2)$ due mainly to matrix multiplication without β.
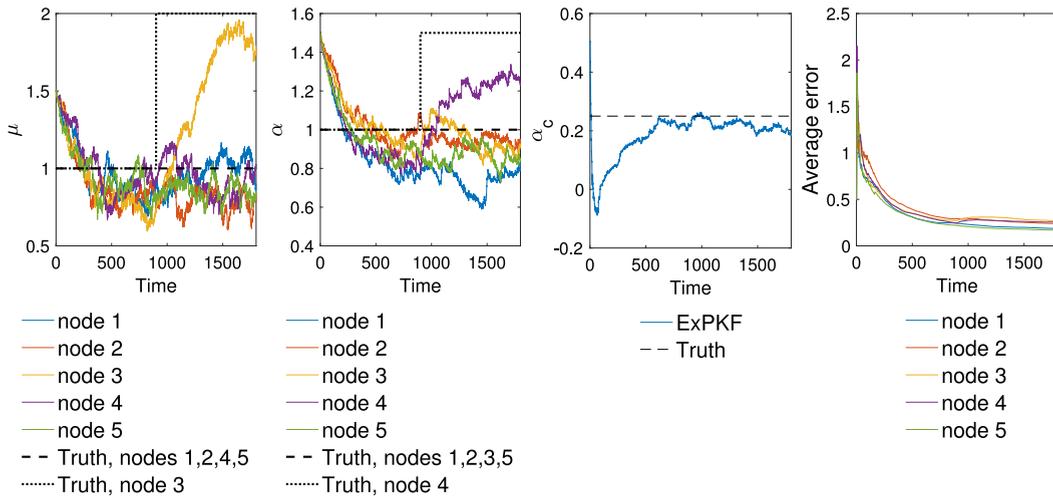
Fig. 5. Estimation of parameters for the model (15) given that β is misidentified as $\beta = 3$.

Table 1
The mean relative error when using a misspecified value of β. The true value is $\beta = 2$.

β	1	2	3	4	8	12	16	20
Error	0.12	0.05	0.07	0.11	0.19	0.24	0.26	0.28

5. Chicago crime data

We apply ExPKF to estimate the dynamic of crime intensity for a crime dataset for Chicago extracted from the Chicago data portal.[1] The dataset consists of robberies, burglaries, criminal damages and thefts from 1 January 2012 till 31 December 2017. We project the data on a regular grid with grid cells of dimension 400 m × 400 m. We restrict our analysis to the 10 × 10 subgrid displayed in Fig. 6 where the dataset shows a stationary hot spot, with the end result being a subset of the data in $m = 100$ uniform rectangular cells. We use $\delta t = 0.1$ day for the computational time step. We first consider the univariate Hawkes process, which is (14) without the cross-excitation in the last term, and use MLE to determine all three parameters for each cell in isolation. The results in Fig. 7 show that the variation of the likelihood is quite insensitive to the decay rate parameter β_j. For example, Fig. 7 shows the result for cell $j = 22$ and the 99% MLE volume (i.e. those with the likelihood no less than 99% of the maximum likelihood) sliced for different values of β_{22}, where the high-likelihood regions on each slice slowly expand as the decay rate increases, but their "core" regions vary only slightly. With this observation in mind, it is reasonable to fix the parameter β_j and use ExPKF to track only the baseline μ_j and the self-excitation parameter α_j; hence we can use the rank-1 update. Fig. 7 also shows the parameters tracked by ExPKF for a fixed value of β_j on each slice, which converge to the high likelihood region on the slice.

We evaluate the goodness-of-fit of the intensity estimated by filtering and non-filtering schemes. The first 800 days of the data (from 01/01/2012 00:00 to 21/03/2014 00:00) is used by MLE to estimate all three parameters for all cells. The MLE-fitted intensity is then computed based on the MLE parameters over the whole period of 1827 days. For the filtering scheme, the MLE-based parameters are used only as the initial mean for ExPKF but we allow the algorithm to track μ_j and α_j for all $j = 1, \ldots, m$. Hence, we obtain filtered intensity $\lambda_j(t + 1)$ simulated from the filtered parameters $\mu_j(t)$ and $\alpha_j(t)$. The one-sample Kolmogorov–Smirnov (KS) statistic is used to test the goodness-of-fit between MLE-based and ExPKF-based intensities for the period of day 801–1827. In particular, a variable transformation

$$z_i = 1 - \exp(-\int_{t_{i-1}}^{t_i} \lambda(t)dt),$$

where t_i is the time of the ith event, is used to compare against the uniform distribution under the KS test; hence, if the estimated intensity $\lambda(t)$ is a good fit, a uniform quantile plot of z_i would lie close to the 45 degree line. The scatter plot in Fig. 9 compares the p-values (under the null hypothesis that z_i comes from a uniform distribution) between MLE and ExPKF. It is clear that the goodness of fit is noticeably improved in most of the cells when using ExPKF compared to MLE. We also observe that there are 16 cells where MLE has the null hypothesis rejected at the 1% significant level but accepted for ExPKF; on the other hand, there are only 3 cells in reverse. In Fig. 9, we also show two examples of those

---

[1] https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present-Map/c4ep-ee5m.
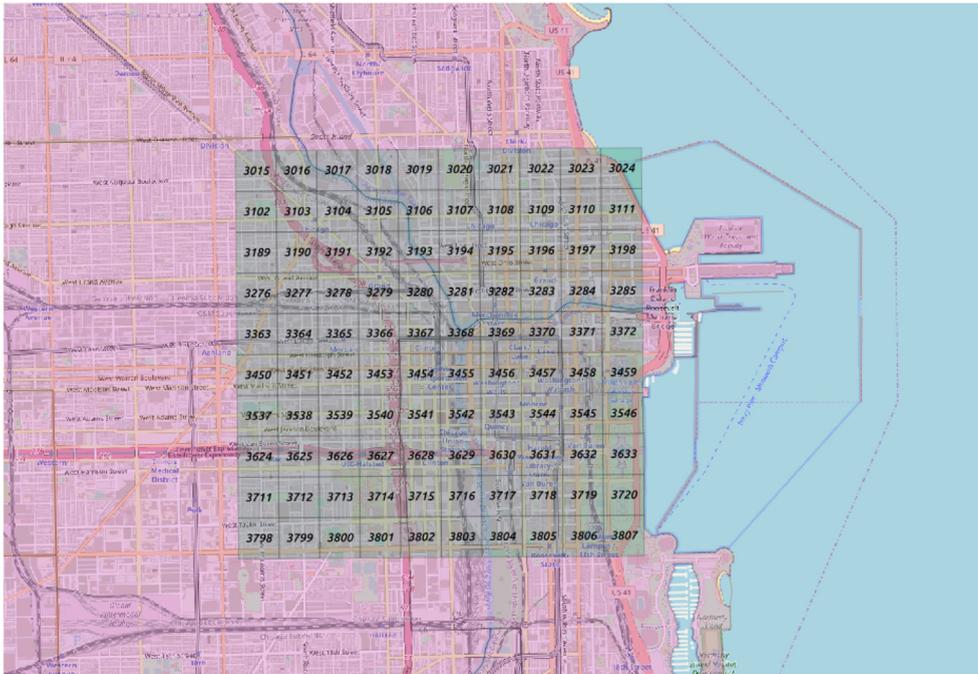
Fig. 6. The Chicago area for the data analysis in this work is shown along with the grid cell number.
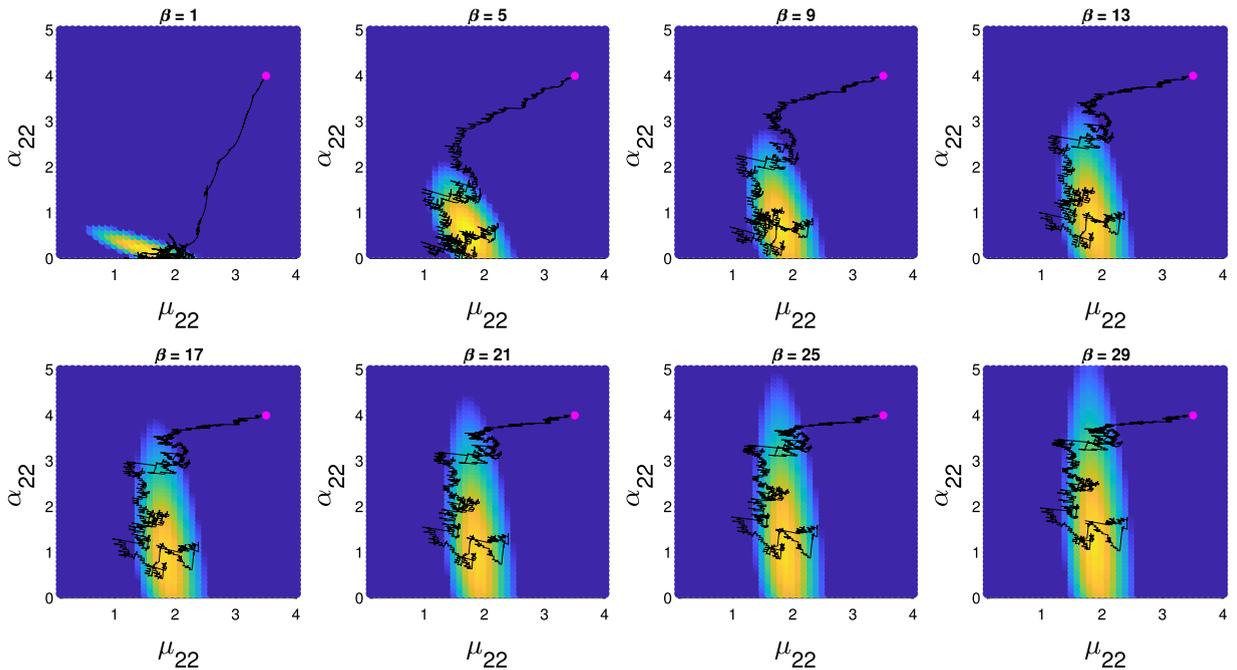
Fig. 7. The MLE region with likelihood greater than 99% of the maximum likelihood at different slices of β_{22}. The trajectories of tracked parameters for μ_{22} and α_{22} in each plot is obtained from ExPKF.

16 cells and find that the variation of the baseline parameter does agree with the variation of average number of events, which highlights the tracking ability of ExPKF.

To demonstrate the potential of ExPKF in dealing with a large number of parameters, we consider the multivariate Hawkes process (15). The 2D neighbourhood centred at the cell j is assumed to be the 3×3 adjacent cells (or the relevant subset of this near the boundaries). Again, we use ExPKF to track only μ_j, α_j and α_c in (15) and assume that the decay
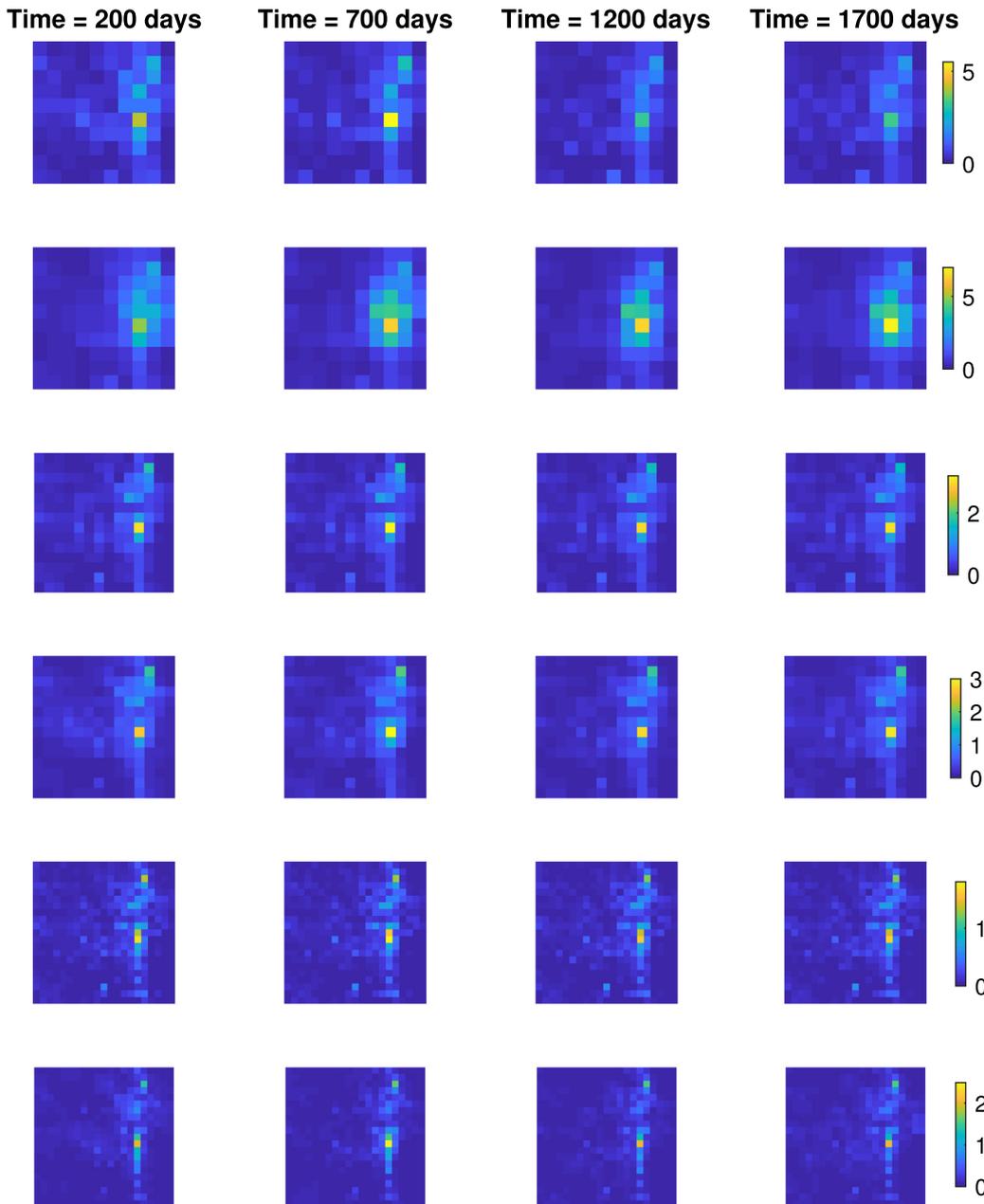
Fig. 8. The variation of the intensity estimated by ExPKF for the univariate model (odd row) and the multivariate model (even row) with cross-cell excitation at the resolution of $m = 100$ (row 1 and 2), $m = 196$ (row 3 and 4) and $m = 441$ (row 5 and 6).

rate is fixed to the MLE parameter obtained from the univariate case. The comparison of the intensity dynamics between univariate and multivariate cases is shown in Fig. 8 at three different spatial resolutions; $m = 100$, 196 and 441 cells ($m = 196$ and $m = 441$ correspond to projections on 300 m × 300 m and 200 m × 200 m grids, respectively), again showing the Cell 22. The smoothing effect of the cross-self excitation in the model (15) can be clearly seen.

5.1. Patrolling prioritisation

In this section we demonstrate an application of ExPKF to Chicago crime data in the context of patrolling efficiency. We assume that the forecast system will prioritise the top *K%* of the cells where the event is most likely to occur within a pre-specified time horizon, which is typically less than 1 day for a practical patrolling application. We test three different approaches for the forecast system: (1) "Climatology" system, which uses an average of long-run historical record to
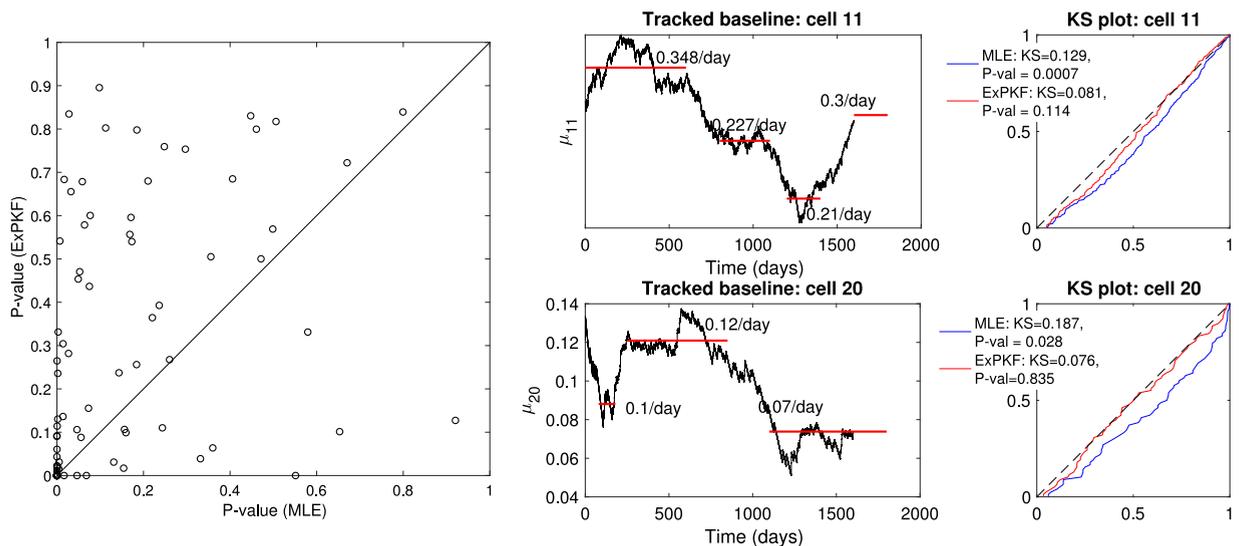
Fig. 9. (Left) Scatter plot between the *p*-value of ExPKF and MLE. (Middle) The tracked baseline parameters for cell 11 and 20 obtained from ExPKF. The average number of events over the period shown in the vertical line agrees with the variation of tracked baseline. (Right) The KS plot of the transformed variable for cell 11 and 20 compared to the cumulative distribution function of the uniform distribution.

determine a time-independent prioritisation, (2) Non-filtering approach, where the parameters of the univariate Hawkes process are fitted using a historical record based on MLE and the model is used to update the crime intensity with fixed MLE parameters and (3) Filtering approach, where the parameters of the Hawkes process will be adaptively tuned at every data assimilation time step. In this experiment, we apply ExPKF to both univariate and multivariate Hawkes processes as described in the preceding section. The tracked crime intensity at the forecast time is then used to rank the cells.

According to the data, some cells (at around $4 - 6\%$) are consistently "hotspots", meaning that these cells are always present in the set of prioritised cells for the values of K used here. However, the top 20% coverage varies and the prediction is more challenging. The climatology system is used as a baseline performance against which we expect the filtering and non-filtering systems to perform at least as well. To construct this baseline, we simply rank the cells in terms of the total amount of crime each contained within the first 800 days, and use that ranking for the entire prediction period. If the Hawkes process accurately represents the dynamic of real-world crime intensity, the non-filtering scheme is appropriate since all one has to do is identify the constant parameters, which in the most extreme scenario may in fact result in intensities that are stationary in time, yielding the climate prediction. On the other hand, if the Hawkes process fails to accurately simulate the real-world crime dynamic, the filtering scheme would attempt to recursively adjust the model parameters so that the predicted crime intensity statistically agrees with the latest data.

For MLE, the first 800 days in the data is used to fit all three parameters of the Hawkes process for each cell in isolation. The MLE parameters are then used to simulate the time-dependent intensity. The ExPKF is applied as done in the preceding section to generate intensity. To measure the forecast performance in the context of patrolling efficiency, we use the Prediction Efficiency Index (PEI), which is the ratio of the number of crime events occurring in the actual predicted cells to the highest possible number of crime events that could have occurred in prediction cells under $K\%$ coverage. We calculate the PEI for $K = 10\%$ and $K = 20\%$ coverage for the prediction time horizons of 0.2, 0.5, and 1 day; see Fig. 10. The result shows that PEI is roughly the same for all forecast methods in our study. That PEI has an increasing trend for a large forecast horizon can be explained by the existence of several persistent hotspots, which renders the (true) ranking less random as the forecast time horizon increases.

6. Conclusion

We adopt the idea of the extended Kalman filter to develop the extended Poisson–Kalman filter (ExPKF), which is an approximated algorithm to filter a history-dependent Poisson likelihood process. We demonstrate the validity of this algorithm based on toy examples for which the true parameters can be used to measure the accuracy of the filter. As can be seen from our experiments, ExPKF is able to converge to the true value in the case of static parameters as well as to track step changes of parameters, but the learning rate for which the filter progresses from the old parameter to the new one could be better in some cases. Note that the filter is not aimed to detect the change point of the parameter, but rather to track the change of the parameter's value. We also point out that the computational cost of ExPKF can be greatly reduced if the Hessian term has the outer-product form. For our focused application of data assimilation for the Hawkes process, this computational reduction can be achieved by fixing the decay rate parameter (i.e. estimating
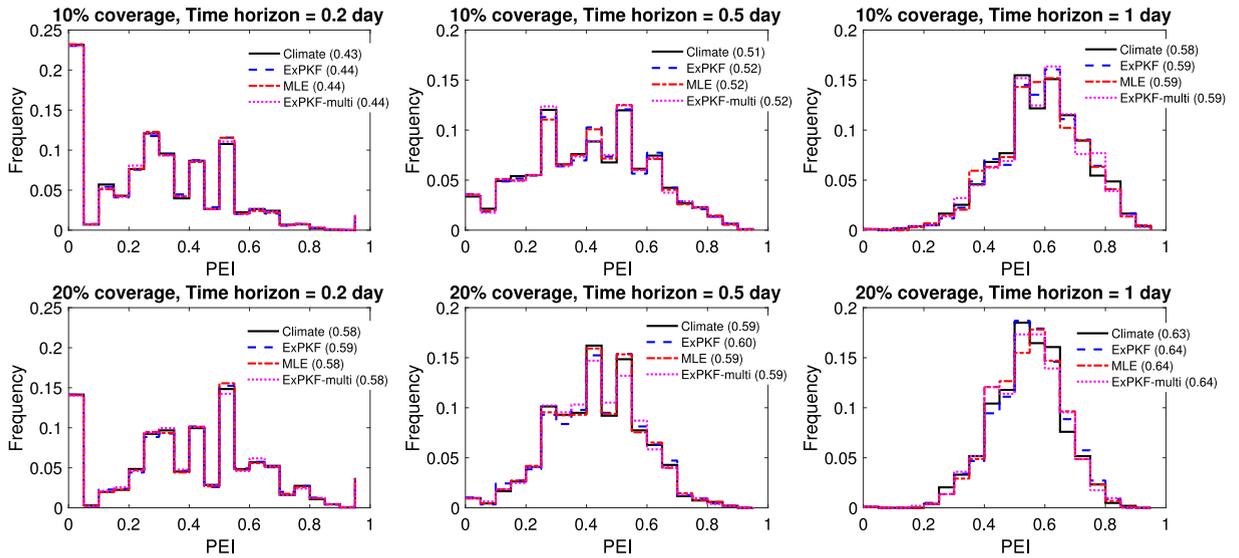
Fig. 10. The PEI is computed for three different forecast horizons (0.2, 0.5, and 1 day) at two different area coverages (10% in the first row and 20% in the second row) using Chicago data during the period of $t = 801 - 1827$ days. The histograms show the frequency of PEI for these cases. In the parentheses, we report the total number of successfully predicted crimes over all prediction periods divided by the total possible number of crimes that could have been predicted over all of these periods.

only the baseline and self-excitation parameters). However, robustness to mis-specification of the pre-specified decay rate may not be guaranteed in general cases. An investigation of likelihood function will shed light on the parameter regimes where the decay rate of Hawkes model is relatively insensitive. This has been experimentally investigated to some extent in Santitissadeekorn et al. (2018). In future work, we aim to develop a fast approximation technique to update the covariance matrix when the outer-product approximation is unavailable. Apart from the computational load to update the covariance matrix, the application of ExPKF is obviously limited to problems where the gradient or Hessian terms are simple to obtain either analytically or numerically. Our recent work (Santitissadeekorn et al., 2018) addressed this issue by developing an ensemble-based filtering method, which does not require the computation of gradient or Hessian terms, but its extension to high-dimensional problems is still waiting for pragmatic algorithm designs.

This work also investigates the application of ExPKF to real-world crime data, using data from Chicago over the years 2012–2017. In particular, we demonstrate that using a filtering scheme to keep track on the baseline and self-excitation parameters can provide estimated intensity with a better goodness of fit than those obtained using fixed MLE parameters. Significant changes in the baseline parameter in some regions are also identified by the tracking skill of ExPKF. These changes found by the algorithm (e.g. cell 20) could be a result of the Chicago Alternative Police Strategy (CAPS) started in July 2014. The evaluation of the CAPS started in 2014 has not been studied but the same program that was run during 1998–2002 demonstrated a decrease in crime in 24% of regions received the CAPS, while those without CAPS intervention saw increases in crime (Young and Skogan, 2003). We believe that the ability of the filtering scheme to automatically follow the parameter change in crime models may have practical advantages such as being able to track and assess policing strategies in real-time.

We also investigate the potential of data assimilation for improving efficient police patrolling. We investigate the Chicago data 2012–2017 as a case study, which displays both persistent and evolving hotspots. Our results based on PEI analysis demonstrate that all the prediction systems (including MLE-based, ExPKF-based, and historical frequency) provide more or less the same predictive skills of prioritising the crime-prone areas over a given forecast time horizon. It should be noted that our results do not necessarily contradict the outcomes found in the study of Mohler et al. (2015), where the prioritisation based on ETAS model showed a significant improvement over those of "crime analysts", who were responsible for placing the prediction boxes in specified time windows. Lacking the information from crime analysts in this case, we compared our results with the historical frequency as described above. We also note that in practice the police cannot simply implement a policy of only patrolling the same locations every day. In addition, it has to be kept in mind that the PEI does not give credit to the neighbourhood aspect of the prediction, in the sense that predicting a no-event cell adjacent to high-crime cells receives the same PEI as predicting another no-event cell far way from it. In practice, one of the ways to improve patrolling efficiency is by reducing the dispatch time of officers to crime locations; hence, having patrol officers near the high-crime cells can actually be helpful, even if the cells they are in do not themselves contain any crimes. Further research to develop a new score taking into account the neighbourhood aspect of the predictions will certainly be useful for the assessment of crime prediction methods in the future.

A future direction of research that has potential to improve efficiency of patrolling allocation (better than what is done in this work) is the joint estimation of parameters and hidden "network" structure in the crime data. A recent study has shown that a complex structure of street network has a significant impact on how crime events would occur and how police would patrol (Rosser et al., 2017). To our best knowledge, incorporating such a complex street network into a Hawkes process has not been reported. Also, a non-parametric approach in Flaxman et al. (2019) has shown that the grid or cell structure is the key element needed to obtain a significant improvement in crime prediction. Other types of non-physical networks can also be considered as well. For example, the network could be defined by the excitation rate in which an event at one cell excites the conditional intensities in other cells. This type of excitation network can be readily incorporated into the Hawkes model (14) but leading to an estimation of N^2 excitation parameters if no prior knowledge of the network is assumed. A follow-up investigation that incorporates such an excitation network structure may yield a major improvement in the predictive skill of the algorithms. A similar application of a point-process on an email network has recently been investigated in Fox et al. (2016) where the email network data is used to infer the leadership. Since ExPKF also approximates the variance–covariance structure of the parameters, it may be used for an extended study in which an uncertainty of leadership can be quantified. The ExPKF algorithm also has a potential for other applications related to a spatial effect of count data. For example, it may be applied to estimate the disease-specific overall risk and the spatial effect in the relative risk across both areas and diseases for a model similar to the multivariate spatial disease model used in Corberán (2012). In particular, the disease-specific risk and spatial effect can be modelled by the baseline term and cross-cells excitation terms in Hawkes model (14), respectively.

Acknowledgements

Appendix

The Appendix provides a brief description of the MATLAB codes (version R2018a) used for Sections 3–5 in this work as well as the (anonymised) data. The step-by-step explanation can be found in the scripts that run these codes.

1. These are codes that implement ExPKF for (13) in Section 3.

   - run_ExPKF_toy1.m is the main script that generates the test data, runs ExPKF, and plots the filtered estimate of parameters. It provides an example and description of how to call the function ExPKF_toy1.m.
   - ExPKF_toy1.m is a function that is called by run_ExPKF_toy1.m to obtain the filtered estimate of parameters.
   - MLE_toy1.m is a script that calls likelihood_toy1.m to run MLE for a given data.
   - likelihood_toy1.m is function that computes the likelihood of the data for a given model parameter value.
   - run_PF_toy1.m is a script that runs the particle filtering (PF) for the same data.
   - PF_toy1.m is a function that is called by run_PF_toy1.m to implement PF.
   - ResampSimp.m is a function that implements the re-sampling step in PF

2. These are codes that implement ExPKF for the multiple-cell Hawkes process (14).

   - run_NHP_hawkes5nodes.m is the main script for the experiment in Section 4. It provides an example and description of how to call the function ExPKF_hawkes5node_muk.m to produced the filtered parameters for the above experiment.
   - NHP_hawkes5nodes.m is the function that simulates a test data for the experiment
   - ExPKF_hawkes5node_muk.m is the function that runs ExPKF for (14) with 5 nodes cell.
   - R1update_Hawkes5nodes.m is the function that implements the rank-1 update for (14) with 5 nodes cell and it is called by ExPKF_hawkes5node_muk.m.

3. These are codes and Chicago crime data used for Section 5.

   - Run_MLE_Chicago1DHawke.m is a script to run MLE for Chicago crime data.
   - run_Chicago_hawkes_Knodes.m is a script to run ExPKF for Chicago crime data.
   - ExPKF_hawkes_Knodes.m is a function to implement ExPKF for Chicago crime data based on the model (14).
   - R1update_Hawkes_Knodes.m is a function that runs the rank-1 update for Chicago crime data based on the model (14). It is called by ExPKF_hawkes_Knodes.m
   - ts_subgrid.mat, ts_subgrid200.mat, ts_subgrid300.mat are the (anonymised) crime data at the scales of each cell at 100-by-100, 200-by-200 and 300-by-300 cells, respectively.

- `incidence_mat_subgrid.mat,inc_mat_200.mat,inc_mat_300.mat` are the incidence matrices indicating the neighbours of each cell for the data `ts_subgrid.mat,ts_subgrid200.m,ts_subgrid300.m, ts_subgrid300.m`, respectively.
- `KSTEST_1DHawkes.m` is a script that computes the Kolmogorov–Smirnov (KS) test from the output of either MLE or ExPKF.
- `PEI_ExPKF.m` is a script that computes PEI for filtered intensity obtained from ExPKF.
- `PEI_mle.m` is a script that computes PEI for the MLE-fitted intensity.

References

Caron, F., Doucet, A., Gottardo, R., 2012. On-line changepoint detection and parameter estimation with application to genomic data. Stat. Comput. 22 (2), 579–595.

Corberán, A., 2012. Prospective surveillance of multivariate spatial disease data. Stat. Methods Med. Res. 21 (5), 457–477.

Cox, D.R., Lewis, P.A., 1966. The Statistical Analysis of Series of Events. Methuen, London.

Crisan, D., Doucet, A., 2002. A survey of convergence results on particle filtering for practitioners. IEEE Trans. Signal Process. 50 (3), 736–746.

Dassios, A., Zhao, H., 2013. Exact simulation of hawkes prcess with exponentially decaying intensity. Electron. Commun. Prob. 18, 1–23.

Flaxman, S., Chirico, M., Pereira, P., Loeffler, P., 2019. Scalable high-resolution forecasting of sparse spatiotemporal events with kernel methods: a winning solution to the NIJ "Real-Time Crime Forecasting Challenge", arXiv:1801:02858v3 [stat.ML].

Fox, E.W., Short, M.B., Schoenberg, F.P., Coronges, K.D., Bertozzi, A.L., 2016. Modeling e-mail networks and inferring leadership using self-exciting point processes. J. Amer. Stat. Assoc. 111 (514), 564–584.

Halpin, P.F., 2013. A scalable EM algorithm for hawkes processes. In: New Developments in Quantitative Psychology. Springer New York, pp. 403–414.

Hawkes, A.G., Oakes, D., 1974. A cluster process representation of a self-exciting process. J. Appl. Probab. 11 (3), 493–503.

Johnson, S.D., Bowers, K., Hirschfield, A., 1997. New insights into the spatial and temporal distribution of repeat victimisation. Br. J. Criminol. 37, 224–244.

Lewis, E., Mohler, G., 2011. A nonparametric EM algorithm for multiscale hawkes process. J. Nonparamet. Stat. 1 (1), 1–20.

Mohler, G.O., Short, M.B., Malinowski, S., Johnson, M., Tita, G.E., Bertozzi, A.L., Brantingham, P.J., 2015. Randomized controlled field trials of predictive policing. J. Amer. Statist. Assoc. 110 (512), 1399–1411.

Mohler, M.B., Brantingham, P.J., Schoenberg, F.P., Tita, G.E., 2011. Self-exciting point process modeling of crime. J. Amer. Statist. Assoc. 106 (493), 100–108.

Ogata, Y., 1981. On lewis' simulation method for point process. IEEE Trans. Inform. Theory IT-27 (1), 23–31.

Ogata, Y., 1988. Statistical models for earthquake occurences and residual analysis for point process. J. Amer. Statist. Assoc. 83, 9–27.

Ogata, Y., Zhuang, J., 2006. Space-time ETAS models and an improved extension. Tectonophysics 413, 13–23.

Olson, J.F., Kathleen, M.C., 2013. Exact and approximate EM estimation of mutually exciting hawkes processes. Stat. Inference Stoch. Proc. 16 (1), 63–80.

Rasmussen, J.G., 2013. Bayesian Inference for hawkes processes. Methodol. Comput. Appl. Prob. 15 (3), 623–642.

Rosser, G., Davides, T., Bowers, K., Johnson, S., Cheng, T., 2017. Predictive crime mapping: arbitrary grids or street network?. J. Quant. Crim. 33 (3), 569–594.

Santitissadeekorn, N., Short, M.B., Lloyd, D.J.B., 2018. Sequential data assimilation for 1d self-exciting processes with application to urban crime data. Comput. Statist. Data Anal. 128, 163–183.

Stephens, D.A., 1994. Bayesian retrospective multiple-changepoint identification. J. Roy. Stat. Soc. C 43 (1), 159–178.

Unkel, S., Farrington, C.P., Gathwaite, P.H., 2012. Statistical methods for the prospective detection of infectious disease outbreaks: a review. J. Roy. Stat. Soc. A 175 (1), 49–82.

Young, K.S., Skogan, G., 2003. Statistical Analysis of Time Serires Data on Problem Solving, Technical Report Illnois Criminal Justice Information Authority.