# Erlang mixture modeling for Poisson process intensities

Hyotae Kim and Athanasios Kottas *

**Abstract**

We develop a prior probability model for temporal Poisson process intensities through structured mixtures of Erlang densities with common scale parameter, mixing on the integer shape parameters. The mixture weights are constructed through increments of a cumulative intensity function which is modeled nonparametrically with a gamma process prior. Such model specification provides a novel extension of Erlang mixtures for density estimation to the intensity estimation setting. The prior model structure supports general shapes for the point process intensity function, and it also enables effective handling of the Poisson process likelihood normalizing constant resulting in efficient posterior simulation. The Erlang mixture modeling approach is further elaborated to develop an inference method for spatial Poisson processes. The methodology is illustrated with synthetic and real data examples.

KEY WORDS: Bayesian nonparametrics; Erlang mixtures; Gamma process; Markov chain Monte Carlo; Non-homogeneous Poisson process.

*Hyotae Kim (hkim153@ucsc.edu) is Ph.D. student and Athanasios Kottas (thanos@soe.ucsc.edu) is Professor, Department of Statistics, University of California, Santa Cruz. This research was supported in part by the National Science Foundation under award SES 1950902.

# 1 Introduction

Poisson processes play a key role in both theory and applications of point processes. They form a widely used class of stochastic models for point patterns that arise in biology, ecology, engineering and finance among many other disciplines. The relatively tractable form of the non-homogeneous Poisson process (NHPP) likelihood is one of the reasons for the popularity of NHPPs in applications involving point process data.

Theoretical background for the Poisson process can be found, for example, in Kingman (1993) and Daley and Vere-Jones (2003). Regarding Bayesian nonparametric modeling and inference, prior probability models have been developed for the NHPP mean measure (e.g., Lo, 1982, 1992), and mainly for the intensity function of NHPPs over time and/or space. Modeling methods for NHPP intensities include: mixtures of non-negative kernels with weighted gamma process priors for the mixing measure (e.g., Lo and Weng, 1989; Wolpert and Ickstadt, 1998; Ishwaran and James, 2004; Kang et al., 2014); piecewise constant functions driven by Voronoi tessellations with Markov random field priors (Heikkinen and Arjas, 1998, 1999); Gaussian process priors for logarithmic or logistic transformations of the intensity (e.g., Møller et al., 1998; Adams et al., 2009; Rodrigues and Diggle, 2012); and Dirichlet process mixtures for the NHPP density, i.e., the intensity function normalized in the observation window (e.g., Kottas and Sansó, 2007; Taddy and Kottas, 2012). Models based on priors for the NHPP intensity typically require complex computational methods for full Bayesian inference. The modeling approach that builds from the NHPP density utilizes well established posterior simulation methods for Dirichlet process mixtures, and it thus facilitates extensions to different types of hierarchical settings (e.g., Taddy, 2010; Xiao et al., 2015; Rodriguez et al., 2017). However, this approach relies on a prior structure that models separately the NHPP density and the total intensity over the observation window.

Here, we seek to develop a flexible and computationally efficient model for NHPP intensity functions over time or space. We focus on temporal intensities to motivate the modeling approach and to detail the methodological development, and then extend the model for spatial NHPPs. The NHPP intensity over time is represented as a weighted combination of Erlang densities indexed by their integer shape parameters and with a common scale parameter. Thus, different from existing mixture representations, the proposed mixture model is more structured with each Erlang density identified by the corresponding mixture weight. The non-negative mixture weights are defined through increments of a cumulative intensity on $\mathbb{R}^+$. Under certain conditions, the Erlang mixture intensity model can approximate in a pointwise sense general intensities on $\mathbb{R}^+$ (see Section 2.1). A gamma process prior is assigned to the primary model component, that is, the cumulative intensity that defines the mixture weights. Mixture weights driven by a gamma process prior result in flexible intensity function shapes, and, at the same time, ready prior-to-posterior updating given the observed point pattern. Indeed, a key feature of the model is that it can be implemented with an efficient Markov chain Monte Carlo (MCMC) algorithm that does not require approximations, complex computational methods, or restrictive prior modeling assumptions in order to handle the NHPP likelihood normalizing constant. The intensity model is extended to the two-dimensional setting through products of Erlang densities for the mixture components, with the weights built from a measure modeled again with a gamma process prior. The extension to spatial NHPPs retains the appealing aspect of computationally efficient MCMC posterior simulation.

The paper is organized as follows. Section 2 presents the modeling and inference methodology for NHPP intensities over time. The modeling approach for temporal NHPPs is illustrated through synthetic and real data in Section 3. Section 4 develops the model for spatial NHPP intensities, including two data examples. Finally, Section 5 concludes with a summary.

# 2 Methodology for temporal Poisson processes

The mixture model for NHPP intensities is developed in Section 2.1, including discussion of model properties and theoretical justification. Sections 2.2 and 2.3 present a prior specification approach and the posterior simulation method, respectively.

## 2.1 The mixture modeling approach

A NHPP on $\mathbb{R}^+$ can be defined through its intensity function, $\lambda(t)$, for $t \in \mathbb{R}^+$, a non-negative and locally integrable function such that: (a) for any bounded $B \subset \mathbb{R}^+$, the number of events in $B$, $N(B)$, is Poisson distributed with mean $\Lambda(B) = \int_B \lambda(u)\,\mathrm{d}u$; and (b) given $N(B) = n$, the times $t_i$, for $i = 1, ..., n$, that form the point pattern in $B$ arise independently and identically distributed (i.i.d.) according to density $\lambda(t)/\Lambda(B)$. Consequently, the likelihood for the NHPP intensity function, based on the point pattern $\{0 < t_1 < ... < t_n < T\}$ observed in time window $(0, T)$, is proportional to $\exp(-\int_0^T \lambda(u)\,\mathrm{d}u) \prod_{i=1}^n \lambda(t_i)$.

Our modeling target is the intensity function, $\lambda(t)$. Denote generically by gamma$(\alpha, \beta)$ the gamma distribution with mean $\alpha/\beta$, and by $\mathrm{ga}(t \mid \alpha, \beta)$, for $t \in \mathbb{R}^+$, the corresponding density. The proposed intensity model involves a structured mixture of Erlang densities, $\mathrm{ga}(t \mid j, \theta^{-1})$, mixing on the integer shape parameters, $j$, with a common scale parameter $\theta$. The non-negative mixture weights are defined through increments of a cumulative intensity function, $H$, on $\mathbb{R}^+$, which is assigned a gamma process prior. More specifically,

$$\begin{aligned} \lambda(t) \equiv \lambda(t \mid H, \theta) \; &= \; \sum_{j=1}^{J} \omega_j \, \mathrm{ga}(t \mid j, \theta^{-1}), \;\; t \in \mathbb{R}^+ \\ \omega_j = H(j\theta) - H((j-1)\theta), &\quad H \sim \mathcal{G}(H_0, c_0) \end{aligned} \tag{1}$$

where $\mathcal{G}(H_0, c_0)$ is a gamma process specified through $H_0$, a (parametric) cumulative intensity function, and $c_0$, a positive scalar parameter (Kalbfleisch 1978). For any $t \in$

$\mathbb{R}^+$, $\mathrm{E}(H(t)) = H_0(t)$ and $\mathrm{Var}(H(t)) = H_0(t)/c_0$, and thus $H_0$ plays the role of the centering cumulative intensity, whereas $c_0$ is a precision parameter. As an independent increments process, the $\mathcal{G}(H_0, c_0)$ prior for $H$ implies that, given $\theta$, the mixture weights $\omega_j$ are independent gamma$(c_0\,\omega_{0j}(\theta), c_0)$ distributed, where $\omega_{0j}(\theta) = H_0(j\theta) - H_0((j-1)\theta)$. As shown in Section 2.3, this is a key property of the prior model with respect to implementation of posterior inference.

The model in (1) is motivated by Erlang mixtures for density estimation, under which a density $g$ on $\mathbb{R}^+$ is represented as $g(t) \equiv g_{J,\theta}(t) = \sum_{j=1}^{J} p_j \,\mathrm{ga}(t \mid j, \theta^{-1})$, for $t \in \mathbb{R}^+$. Here, $p_j = G(j\theta) - G((j-1)\theta)$, where $G$ is a distribution function on $\mathbb{R}^+$; the last weight can be defined as $p_J = 1 - G((J-1)\theta)$ to ensure that $(p_1, ..., p_J)$ is a probability vector. Erlang mixtures can approximate general densities on the positive real line, in particular, as $\theta \to 0$ and $J \to \infty$, $g_{J,\theta}$ converges pointwise to the density of distribution function $G$ that defines the mixture weights. This convergence property can be obtained from more general results from the probability literature that studies Erlang mixtures as extensions of Bernstein polynomials to the positive real line (e.g., Butzer, 1954); a convergence proof specifically for the distribution function of $g_{J,\theta}$ can be found in Lee and Lin (2010). Density estimation on compact sets via Bernstein polynomials has been explored in the Bayesian nonparametrics literature following the work of Petrone (1999a,b). Regarding Bayesian nonparametric modeling with Erlang mixtures, we are only aware of Xiao et al. (2020) where renewal process inter-arrival distributions are modeled with mixtures of Erlang distributions, using a Dirichlet process prior (Ferguson 1973) for distribution function $G$. Venturini et al. (2008) study a parametric Erlang mixture model for density estimation on $\mathbb{R}^+$, working with a Dirichlet prior distribution for the mixture weights.

Therefore, the modeling approach in (1) exploits the structure of the Erlang mixture density model to extend to a model for NHPP intensities, using the density/distribution function and intensity/cumulative intensity function connection to define the prior model for the mixture weights. In this context, the gamma process prior for cumulative intensity $H$ is the natural analogue to the Dirichlet process prior for distribution function $G$; recall that the Dirichlet process can be defined through normalization of a gamma process (e.g., Ghosal and van der Vaart, 2017). To our knowledge, this is a novel construction for NHPP intensities that has not been explored for intensity estimation in either the classical or Bayesian nonparametrics literature. The following lemma, which can be obtained applying Theorem 2 from Butzer (1954), provides theoretical motivation and support for the mixture model.

**Lemma.** Let $h$ be the intensity function of a NHPP on $\mathbb{R}^+$, with cumulative intensity function $H(t) = \int_0^t h(u)\,\mathrm{d}u$, such that $H(t) = O(t^m)$, as $t \to \infty$, for some $m > 0$. Consider the mixture intensity model $\lambda_{J,\theta}(t) = \sum_{j=1}^{J} \{H(j\theta) - H((j-1)\theta)\}\,\mathrm{ga}(t \mid j, \theta^{-1})$, for $t \in \mathbb{R}^+$. Then, as $\theta \to 0$ and $J \to \infty$, $\lambda_{J,\theta}(t)$ converges to $h(t)$ at every point $t$ where $h(t) = \mathrm{d}H(t)/\mathrm{d}t$.

The form of the prior model for the intensity in (1) allows ready expressions for other NHPP functionals. For instance, the total intensity over the observation time window $(0,T)$ is given by $\int_0^T \lambda(u)\,\mathrm{d}u = \sum_{j=1}^{J} \omega_j K_{j,\theta}(T)$, where $K_{j,\theta}(T) = \int_0^T \mathrm{ga}(u \mid j, \theta^{-1})\,\mathrm{d}u$ is the $j$-th Erlang distribution function at $T$. In the context of the MCMC posterior simulation method, this form enables efficient handling of the NHPP likelihood normalizing constant. Moreover, the NHPP density on interval $(0,T)$ can be expressed as a mixture of truncated Erlang densities. More specifically, $f(t) = \lambda(t)/\int_0^T \lambda(u)\,\mathrm{d}u = \sum_{j=1}^{J} \omega_j^*\, k(t \mid j, \theta)$, for $t \in (0,T)$, where $\omega_j^* = \omega_j K_{j,\theta}(T)/\{\sum_{r=1}^{J} \omega_r K_{r,\theta}(T)\}$, and $k(t \mid j,\theta)$ is the $j$-th Erlang density truncated on $(0,T)$.

Regarding the role of the different model parameters, we reiterate that (1) corresponds to a structured mixture. The Erlang densities, $\mathrm{ga}(t \mid j, \theta^{-1})$, play the role of basis functions in the representation for the intensity. In this respect, of primary importance is the

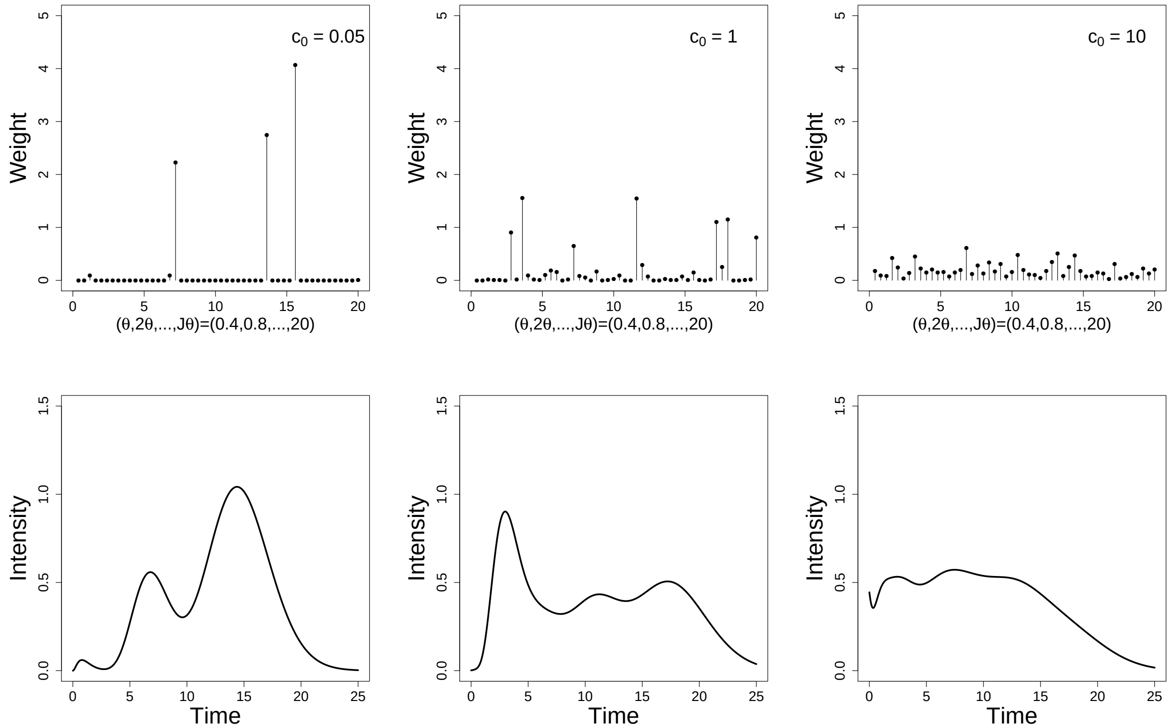


Figure 1: Prior realizations for the mixture weights (top panels) and the corresponding intensity function (bottom panels) for three different values of the gamma process precision parameter, $c_0 = 0.05, 1, 10$. In all cases, $J = 50$, $\theta = 0.4$, and $H_0(t) = t/2$.

flexibility of the nonparametric prior for the cumulative intensity function $H$ that defines the mixture weights. In particular, the gamma process prior provides realizations for $H$ with general shapes that can concentrate on different time intervals, thus favoring different subsets of the Erlang basis densities through the corresponding $\omega_j$. Here, the key parameter is the precision parameter $c_0$, which controls the variability of the gamma process prior around $H_0$, and thus the effective mixture weights. As an illustration, Figure 1 shows prior realizations for the weights $\omega_j$ (and the resulting intensity function) for different values of $c_0$, keeping all other model parameters the same. Note that as $c_0$ decreases, so does the number of practically non-zero weights.

The prior mean for $H$ is taken to be $H_0(t) = t/b$, i.e., the cumulative intensity (hazard) of an exponential distribution with scale parameter $b > 0$. Although it is possible to use more general centering functions, such as the Weibull $H_0(t) = (t/b)^a$, the exponential form is sufficiently flexible in practice, as demonstrated with the synthetic data examples of Section 3. Based on the role of $H$ in the intensity mixture model, we typically anticipate realizations for $H$ that are different from the centering function $H_0$, and thus, as discussed above, the more important gamma process parameter is $c_0$. Moreover, the exponential form for $H_0$ allows for an analytical result for the prior expectation of the Erlang mixture intensity model. Under $H_0(t) = t/b$, the prior expectation for the weights is given by $\mathrm{E}(\omega_j \mid \theta, b) = \theta/b$. Therefore, conditional on all model hyperparameters, the expectation of $\lambda(t)$ over the gamma process prior can be written as

$$\mathrm{E}(\lambda(t) \mid b, \theta) = \frac{\theta}{b} \sum_{j=1}^{J} \mathrm{ga}(t \mid j, \theta^{-1}) = \frac{\exp(-(t/\theta))}{b} \sum_{m=0}^{J-1} \frac{(t/\theta)^m}{m!}, \quad t \in \mathbb{R}^+$$

which converges to $b^{-1}$, as $J \to \infty$, for any $t \in \mathbb{R}^+$ (and regardless of the value of $\theta$ and $c_0$). In practice, the prior mean for the intensity function is essentially constant at $b^{-1}$ for $t \in (0, J\theta)$, which, as discussed below, is roughly the effective support of the NHPP intensity. This result is useful for prior specification as it distinguishes the role of $b$ from that of parameters $\theta$ and $c_0$.

Also key are the two remaining model parameters, the number of Erlang basis densities $J$, and their common scale parameter $\theta$. Parameters $\theta$ and $J$ interact to control both the effective support and shape of NHPP intensities arising under (1). Regarding intensity shapes, as the lemma suggests, smaller values of $\theta$ and larger values of $J$ generally result in more variable, typically multimodal intensities. Moreover, the representation for $\lambda(t)$ in (1) utilizes Erlang basis densities with increasing means $j\theta$, and thus $(0, J\theta)$ can be used as a proxy for the effective support of the NHPP intensity. Of course, the mean underestimates the effective support, a more accurate guess can be obtained using, say,

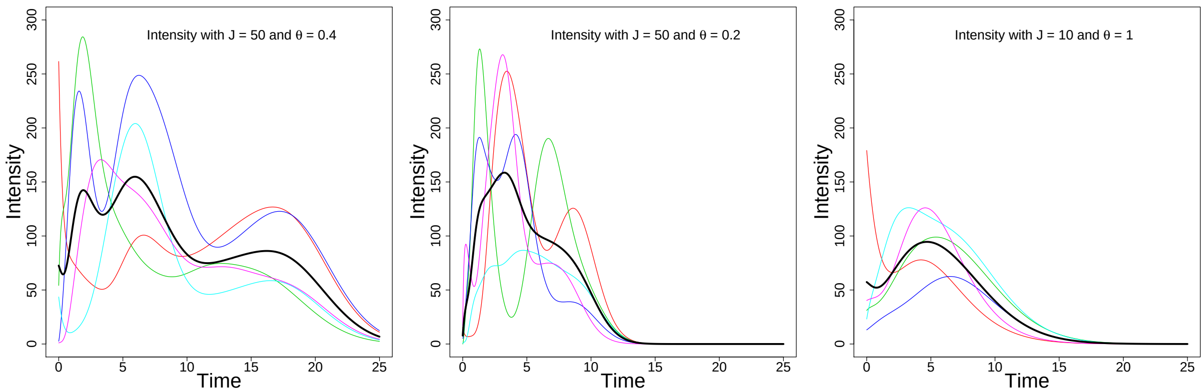


Figure 2: Prior realizations for the intensity under the Erlang mixture model in (1) with $(\theta, J) = (0.4, 50)$ (left panel), $(\theta, J) = (0.2, 50)$ (middle panel), and $(\theta, J) = (1, 10)$ (right panel). In all cases, the gamma process prior is specified with $c_0 = 0.01$ and $H_0(t) = t/0.01$. Each panel plots five realizations, the average of which is indicated by the black line.

the 95% percentile of the last Erlang density component. For an illustration, Figure 2 plots prior intensity realizations for three combinations of $(\theta, J)$ values, with $c_0 = 0.01$ and $b = 0.01$ in all cases. The left panel corresponds to the largest value for $J\theta$ and, consequently, to the widest effective support interval. The value of $J\theta$ is the same for the middle and right panels, resulting in similar effective support. However, the intensities in the middle panel show larger variability in their shapes, as expected since the value of $J$ is increased and the value of $\theta$ decreased relative to the ones in the right panel.

## 2.2 Prior specification

To complete the full Bayesian model, we place exponential prior distributions on the parameters $c_0$ and $b$ of the gamma process prior for $H$, and a Lomax prior on the common scale parameter $\theta$ of the Erlang basis densities. A fairly generic approach to specify these hyperpriors can be obtained using the observation time window $(0, T)$ as the effective support of the NHPP intensity to be estimated.

Using the prior mean for the intensity function, which as discussed in Section 2.1 is roughly constant at $b^{-1}$ within the time interval of interest, the total intensity in $(0, T)$ can be approximated by $T/b$. Therefore, taking the size $n$ of the observed point pattern, as a proxy for the total intensity in $(0, T)$, we can use $T/n$ to specify the mean of the exponential prior distribution for $b$. Given its role in the gamma process prior, we anticipate that small values of $c_0$ will be important to allow prior variability around $H_0$, as well as sparsity in the mixture weights. Experience from prior simulations, such as the ones shown in Figure 1, is useful to guide the range of "small" values. Note that the pattern observed in Figure 1 is not affected by the length of the observation window. In general, a value around 10 can be viewed as a conservative guess at a high percentile for $c_0$. For the data examples of Section 3, we assigned an exponential prior with mean 10 to $c_0$, observing substantial learning for this key model hyperparameter with its posterior distribution supported by values (much) smaller than 1.

Also given the key role of parameter $\theta$ in controlling the intensity shapes, we recommend favoring sufficiently small values in the prior for $\theta$, especially if prior information suggests a non-standard intensity shape. Recall that $\theta$, along with $J$, control the effective support of the intensity, and thus "small" values for $\theta$ should be assessed relative to the length of the observation window. Again, prior simulation, as in Figure 2, is a useful tool. A practical approach to specify the prior range of $\theta$ values involves reducing the Erlang mixture model to the first component. The corresponding (exponential) density has mean $\theta$, and we thus use $(0, T)$ as the effective prior range for $\theta$. Because $T$ is a fairly large upper bound, and since we wish to favor smaller $\theta$ values, rather than an exponential prior, we use a Lomax prior, $p(\theta) \propto (1 + d_\theta^{-1}\theta)^{-3}$, with shape parameter equal to 2 (thus implying infinite variance), and median $d_\theta(\sqrt{2} - 1)$. The value of the scale parameter, $d_\theta$, is specified such that $\Pr(0 < \theta < T) \approx 0.999$. This simple strategy is effective in practice in identifying a plausible range of $\theta$ values. For the synthetic data examples of Section 3, for which $T = 20$, we assigned a Lomax prior with scale parameter $d_\theta = 1$ to $\theta$, obtaining overall moderate prior-to-posterior learning for $\theta$.

Finally, we work with fixed $J$, the value of which can be specified exploiting the role of $\theta$ and $J$ in controlling the support of the NHPP intensity. In particular, $J$ can be set equal to the integer part of $T/\theta^*$, where $\theta^*$ is the prior median for $\theta$. More conservatively, this value can be used as a lower bound for values of $J$ to be studied in a sensitivity analysis, especially for applications where one expects non-standard shapes for the intensity function. In practice, we recommend conducting prior sensitivity analysis for all model parameters, as well as plotting prior realizations and prior uncertainty bands for the intensity function to graphically explore the implications of different prior choices.

The number of Erlang basis densities is the only model parameter which is not assigned a hyperprior. Placing a prior on $J$ complicates significantly the posterior simulation method, as it necessitates use of variable-dimension MCMC techniques, while offering relatively little from a practical point of view. The key observation is again that the Erlang densities play the role of basis functions rather than of kernel densities in traditional (less structured) finite mixture models. Also key is the nonparametric nature of the prior for function $H$ that defines the mixture weights which *select* the Erlang densities to be used in the representation of the intensity. This model feature effectively guards against over-fitting if one conservatively chooses a larger value for $J$ than may be necessary. In this respect, the flexibility afforded by random parameters $c_0$ and $\theta$ is particularly useful. Overall, we have found that fixing $J$ strikes a good balance between computational tractability and model flexibility in terms of the resulting inferences.

## 2.3 Posterior simulation

Denote as before by $\{0 < t_1 < ... < t_n < T\}$ the point pattern observed in time window $(0, T)$. Under the Erlang mixture model of Section 2.1, the NHPP likelihood is proportional to

$$\exp\left(-\int_0^T \lambda(u)\,\mathrm{d}u\right)\prod_{i=1}^n \lambda(t_i) = \exp\left(-\sum\nolimits_{j=1}^J \omega_j K_{j,\theta}(T)\right)\prod_{i=1}^n \left\{\sum_{j=1}^J \omega_j \,\mathrm{ga}(t_i \mid j, \theta^{-1})\right\}$$

$$= \prod_{j=1}^J \exp(-\omega_j K_{j,\theta}(T)) \prod_{i=1}^n \left\{\left(\sum\nolimits_{r=1}^J \omega_r\right)\sum_{j=1}^J \left(\frac{\omega_j}{\sum_{r=1}^J \omega_r}\right) \mathrm{ga}(t_i \mid j, \theta^{-1})\right\}$$

where $K_{j,\theta}(T) = \int_0^T \mathrm{ga}(u \mid j, \theta^{-1})\,\mathrm{d}u$ is the $j$-th Erlang distribution function at $T$.

For the posterior simulation approach, we augment the likelihood with auxiliary variables $\boldsymbol{\gamma} = \{\gamma_i : i = 1, \ldots, n\}$, where $\gamma_i$ identifies the Erlang basis density to which time event $t_i$ is assigned. Then, the augmented, hierarchical model for the data can be expressed as follows:

$$\begin{aligned}
\{t_1, ..., t_n\} \mid \boldsymbol{\gamma}, \boldsymbol{\omega}, \theta &\sim \prod_{j=1}^J \exp(-\omega_j K_{j,\theta}(T)) \prod_{i=1}^n \left\{\left(\sum\nolimits_{r=1}^J \omega_r\right) \mathrm{ga}(t_i \mid \gamma_i, \theta^{-1})\right\} \\
\gamma_i \mid \boldsymbol{\omega} &\overset{i.i.d.}{\sim} \sum_{j=1}^J \left(\frac{\omega_j}{\sum_{r=1}^J \omega_r}\right) \delta_j(\gamma_i), \quad i = 1, ..., n \\
\theta, c_0, b, \boldsymbol{\omega} &\sim p(\theta)\, p(c_0)\, p(b) \prod_{j=1}^J \mathrm{ga}(\omega_j \mid c_0\, \omega_{0j}(\theta), c_0)
\end{aligned} \tag{2}$$

where $\boldsymbol{\omega} = \{\omega_j : j = 1, ..., J\}$, and $p(\theta)$, $p(c_0)$, and $p(b)$ denote the priors for $\theta$, $c_0$, and $b$. Recall that, under the exponential distribution form for $H_0 = t/b$, we have $\omega_{0j}(\theta) = \theta/b$.

We utilize Gibbs sampling to explore the posterior distribution. The sampler involves ready updates for the auxiliary variables $\gamma_i$, and, importantly, also for the mixture weights $\omega_j$. More specifically, the posterior full conditional for each $\gamma_i$ is a discrete distribution on $\{1, ..., J\}$ such that $\Pr(\gamma_i = j \mid \theta, \boldsymbol{\omega}, \text{data}) \propto \omega_j \,\mathrm{ga}(t_i \mid j, \theta^{-1})$, for $j = 1, ..., J$.

Denote by $N_j = |\{t_i : \gamma_i = j\}|$, for $j = 1, ..., J$, that is, $N_j$ is the number of time points assigned to the $j$-th Erlang basis density. The posterior full conditional distribution for

$\omega$ is derived as follows:

$$
\begin{aligned}
p(\boldsymbol{\omega} \mid \theta, c_0, b, \boldsymbol{\gamma}, \text{data}) &\propto \left\{ \prod_{j=1}^{J} \exp(-\omega_j K_{j,\theta}(T)) \right\} \left( \sum\nolimits_{r=1}^{J} \omega_r \right)^{n} \\
&\times \left\{ \prod_{j=1}^{J} \omega_j^{N_j} \left( \sum\nolimits_{r=1}^{J} \omega_r \right)^{-N_j} \right\} \left\{ \prod_{j=1}^{J} \text{ga}(\omega_j \mid c_0\, \omega_{0j}(\theta), c_0) \right\} \\
&\propto \prod_{j=1}^{J} \exp(-\omega_j K_{j,\theta}(T))\, \omega_j^{N_j} \, \text{ga}(\omega_j \mid c_0\, \omega_{0j}(\theta), c_0) \\
&= \prod_{j=1}^{J} \text{ga}(\omega_j \mid N_j + c_0\, \omega_{0j}(\theta), K_{j,\theta}(T) + c_0)
\end{aligned}
$$

where we have used the fact that $\sum_{j=1}^{J} N_j = n$. Therefore, given the other parameters and the data, the mixture weights are independent, and each $\omega_j$ follows a gamma posterior full conditional distribution. This is a practically important feature of the model both in terms of convenient updates for the mixture weights, as well as with respect to efficiency of the posterior simulation algorithm in updating this key component of the model parameter vector.

Finally, parameter $\theta$ and the hyperparameters, $c_0$ and $b$, of the gamma process prior for $H$ are updated with Metropolis-Hastings (M-H) steps. A log-normal proposal distribution is employed for the M-H step used to update each of these parameters.

# 3 Data examples

To empirically investigate inference under the proposed model, we present three synthetic data examples corresponding to decreasing, increasing, and bimodal intensities. We also consider the coal-mining disasters data set, which is commonly used to illustrate NHPP intensity estimation.

Convergence and mixing of the MCMC algorithm was assessed graphically through

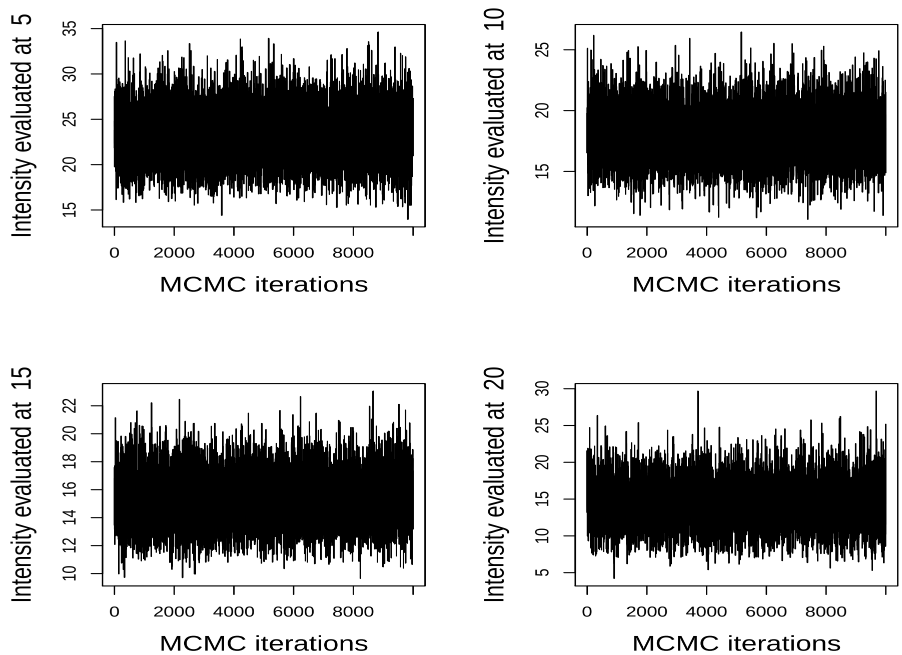


Figure 3: Synthetic data from temporal NHPP with decreasing intensity. Trace plots of posterior samples for the intensity function evaluated at time points $t = 5$, 10, 15, 20.

trace plots of the intensity function evaluated at specific time points within the observation window. An example is given in Figure 3 for the synthetic data of Section 3.1. The trace plots in Figure 3 are representative of the mixing observed in all other data examples. Regarding model parameters, the highest autocorrelation was observed in posterior samples for parameter $\theta$.

We used the approach of Section 2.2 to specify the priors for $c_0$, $b$ and $\theta$, and the value for $J$. In particular, we used the exponential prior for $c_0$ with mean 10 for all data examples. For the three synthetic data sets (for which $T = 20$), we used the Lomax prior for $\theta$ with shape parameter equal to 2 and scale parameter equal to 1.

## 3.1 Decreasing intensity synthetic point pattern

The first synthetic data set involves 491 time points generated in time window $(0, 20)$ from a NHPP with intensity function $\beta^{-1}\alpha(\beta^{-1}t)^{\alpha-1}$, where $(\alpha, \beta) = (0.5, 8 \times 10^{-5})$. This form corresponds to the hazard function of a Weibull distribution with shape parameter less than 1, thus resulting in a decreasing intensity function.

The Erlang mixture model was applied with $J = 50$, and an exponential prior for $b$ with mean 0.04. As can be seen in Figure 4, the model captures the decreasing pattern of the data generating intensity function. We note that there is significant prior-to-posterior learning in the intensity function estimation; the prior intensity mean is roughly constant at value about 25 with prior uncertainty bands that cover almost the entire top left panel in Figure 4. Prior uncertainty bands were similarly wide for all other data examples.

## 3.2 Increasing intensity synthetic point pattern

We consider again the form $\beta^{-1}\alpha(\beta^{-1}t)^{\alpha-1}$ for the NHPP intensity function, but here with $(\alpha, \beta) = (6, 7)$ such that the intensity is increasing. A point pattern comprising 565 points was generated in time window $(0, 20)$. The Erlang mixture model was applied with $J = 50$, and an exponential prior for $b$ with mean 0.035. Figure 5 reports inference results which demonstrate that the model captures successfully the underlying increasing intensity function.

## 3.3 Bimodal intensity synthetic point pattern

The data examples in Sections 3.1 and 3.2 illustrate the model's capacity to uncover monotonic intensity shapes, associated with a parametric distribution different from the Erlang distribution that forms the basis of the mixture intensity model. Here, we consider a point pattern generated from a NHPP with a more complex intensity function, $\lambda(t) =$

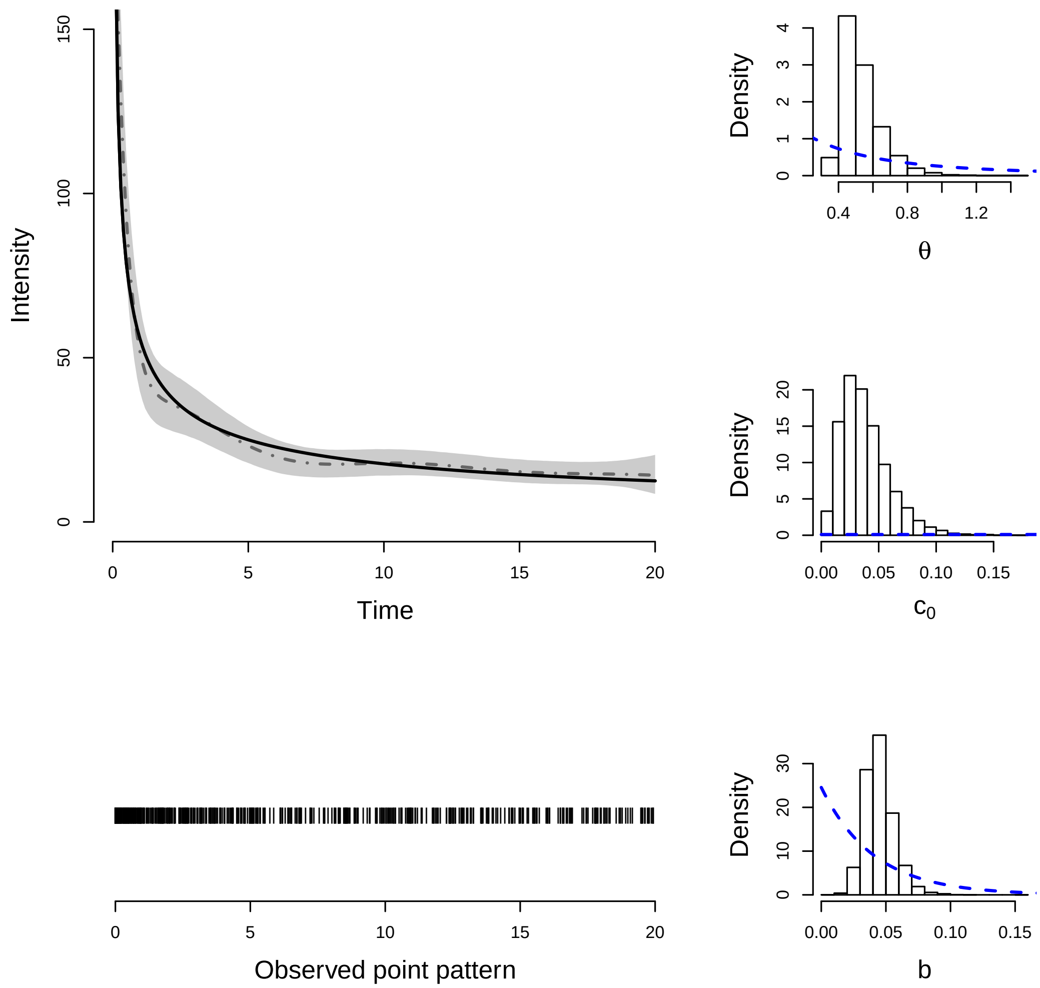


Figure 4: Synthetic data from temporal NHPP with decreasing intensity. The top left panel shows the posterior mean estimate (dashed-dotted line) and posterior 95% interval bands (shaded area) for the intensity function. The true intensity is denoted by the solid line. The point pattern is plotted in the bottom left panel. The three plots on the right panels display histograms of the posterior samples for the model hyperparameters, along with the corresponding prior densities (dashed lines).

$50\,\mathrm{We}(t \mid 3.5, 5) + 60\,\mathrm{We}(t \mid 6.5, 15)$, where $\mathrm{We}(t \mid \alpha, \beta)$ denotes the Weibull density with shape parameter $\alpha$ and mean $\beta\,\Gamma(1 + 1/\alpha)$. This specification results in a bimodal intensity within the observation window $(0, 20)$ where a synthetic point pattern of 112 time points is generated; see Figure 6.

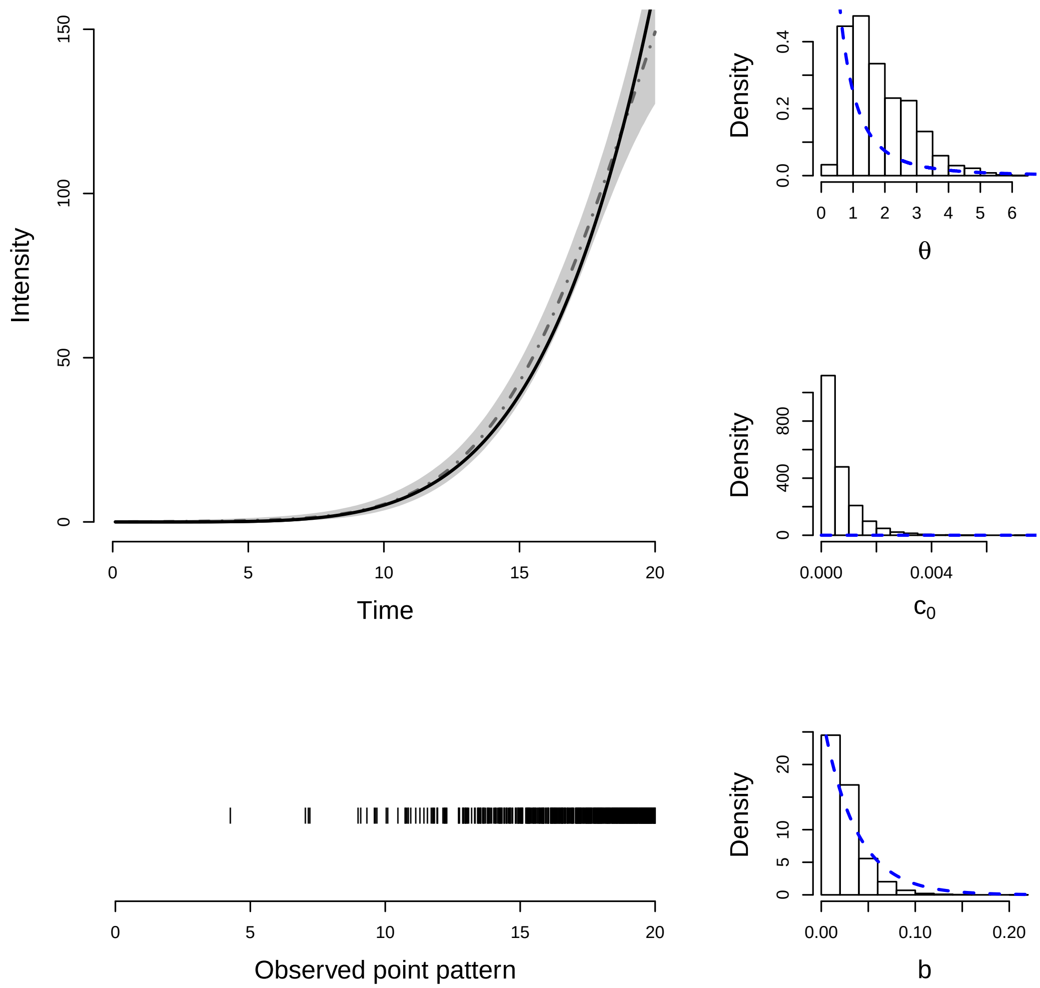


Figure 5: Synthetic data from temporal NHPP with increasing intensity. The top left panel shows the posterior mean estimate (dashed-dotted line) and posterior 95% interval bands (shaded area) for the intensity function. The true intensity is denoted by the solid line. The point pattern is plotted in the bottom left panel. The three plots on the right panels display histograms of the posterior samples for the model hyperparameters, along with the corresponding prior densities (dashed lines).

We used an exponential prior for $b$ with mean 0.179. Anticipating an underlying intensity with less standard shape than in the earlier examples, we compare inference results under $J = 50$ and $J = 100$; see Figure 6. The posterior point and interval estimates capture effectively the bimodal intensity shape, especially if one takes into account the

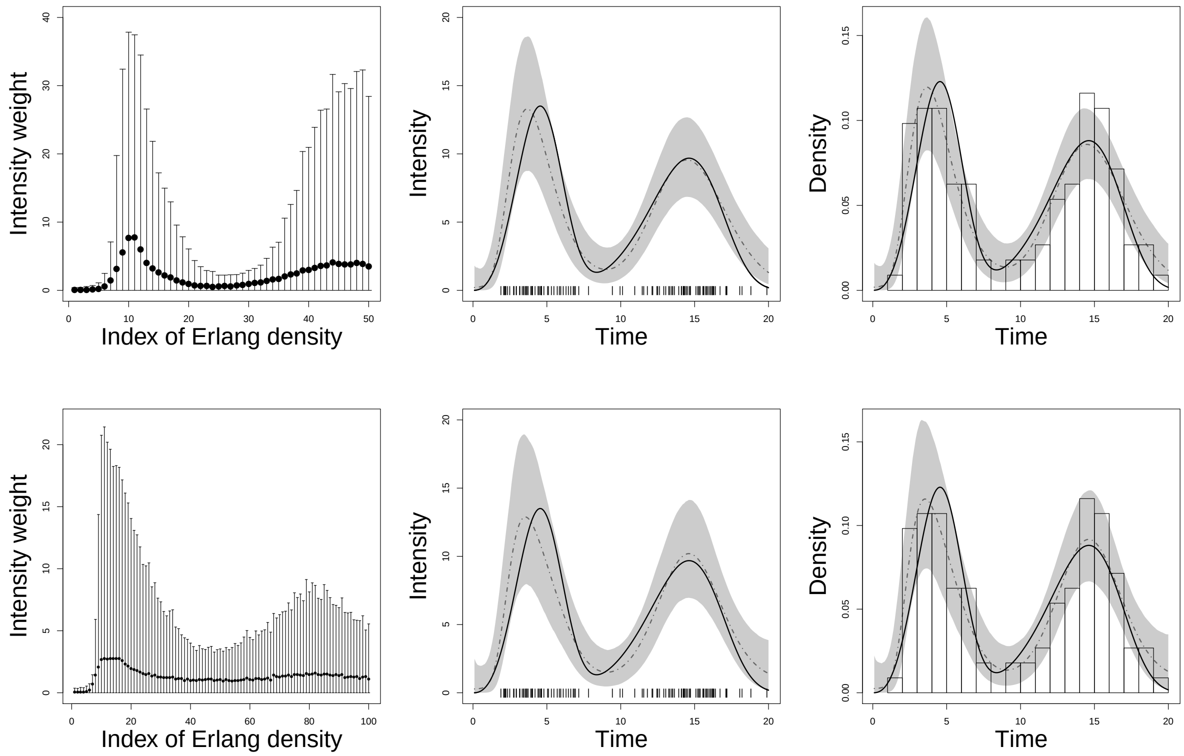


Figure 6: Synthetic data from temporal NHPP with bimodal intensity. Inference results are reported under $J = 50$ (top row) and $J = 100$ (bottom row). The left column plots the posterior means (circles) and 90% interval estimates (bars) of the weights for the Erlang basis densities. The middle column displays the posterior mean estimate (dashed-dotted line) and posterior 95% interval bands (shaded area) for the NHPP intensity function. The true intensity is denoted by the solid line. The bars on the horizontal axis indicate the point pattern. The right column plots the posterior mean estimate (dashed-dotted line) and posterior 95% interval bands (shaded area) for the NHPP density function on the observation window. The histogram corresponds to the simulated times that comprise the point pattern.

relatively small size of the point pattern. (In particular, the histogram of the simulated random time points indicates that they do not provide an entirely accurate depiction of the underlying NHPP density shape.) The estimates are somewhat more accurate under $J = 100$. The estimates for the mixture weights (left column of Figure 6) indicate the subsets of the Erlang basis densities that are utilized under the two different values for $J$. The posterior mean of $\theta$ was 0.366 under $J = 50$, and 0.258 under $J = 100$, that is, as expected, inference for $\theta$ adjusts to different values of $J$ such that $(0, J\theta)$ provides roughly the effective support of the intensity.

### 3.4 Coal-mining disasters data

Our real data example involves the "coal-mining disasters" data (e.g., Andrews and Herzberg 1985, p. 53-56), a standard dataset used in the literature to test NHPP intenstiy estimation methods. The point pattern comprises the times (in days) of $n = 191$ explosions of fire-damp or coal-dust in mines resulting in 10 or more casualties from the accident. The observation window consists of 40,550 days, from March 15, 1851 to March 22, 1962.

We fit the Erlang mixture model with $J = 50$, using a Lomax prior for $\theta$ with shape parameter 2 and scale parameter $2,000$, such that $\Pr(0 < \theta < 40,550) \approx 0.998$, and an exponential prior for $b$ with mean 213. We also implemented the model with $J = 130$, obtaining essentially the same inference results for the NHPP functionals with the ones reported in Figure 7.

The estimates for the point process intensity and density functions (Figure 7, top row) suggest that the model successfully captures the multimodal intensity shape suggested by the data. The estimates for the mixture weights (Figure 7, bottom left panel) indicate the Erlang basis densities that are more influential to the model fit.

The bottom right panel of Figure 7 reports results from graphical model checking, using the "time-rescaling" theorem (e.g., Daley and Vere-Jones 2003). If the point pattern $\{0 = t_0 < t_1 < ... < t_n < T\}$ is a realization from a NHPP with cumulative intensity function $\Lambda(t) = \int_0^t \lambda(u)\mathrm{d}u$, then the transformed point pattern $\{\Lambda(t_i) : i = 1, ..., n\}$ is a realization from a unit rate homogeneous Poisson process. Therefore, if we further transform to $U_i = 1 - \exp\{-(\Lambda(t_i) - \Lambda(t_{i-1}))\}$, where $\Lambda(0) \equiv 0$, then the $\{U_i : i = 1, ..., n\}$ are independent uniform$(0, 1)$ random variables. Hence, graphical model checking can be

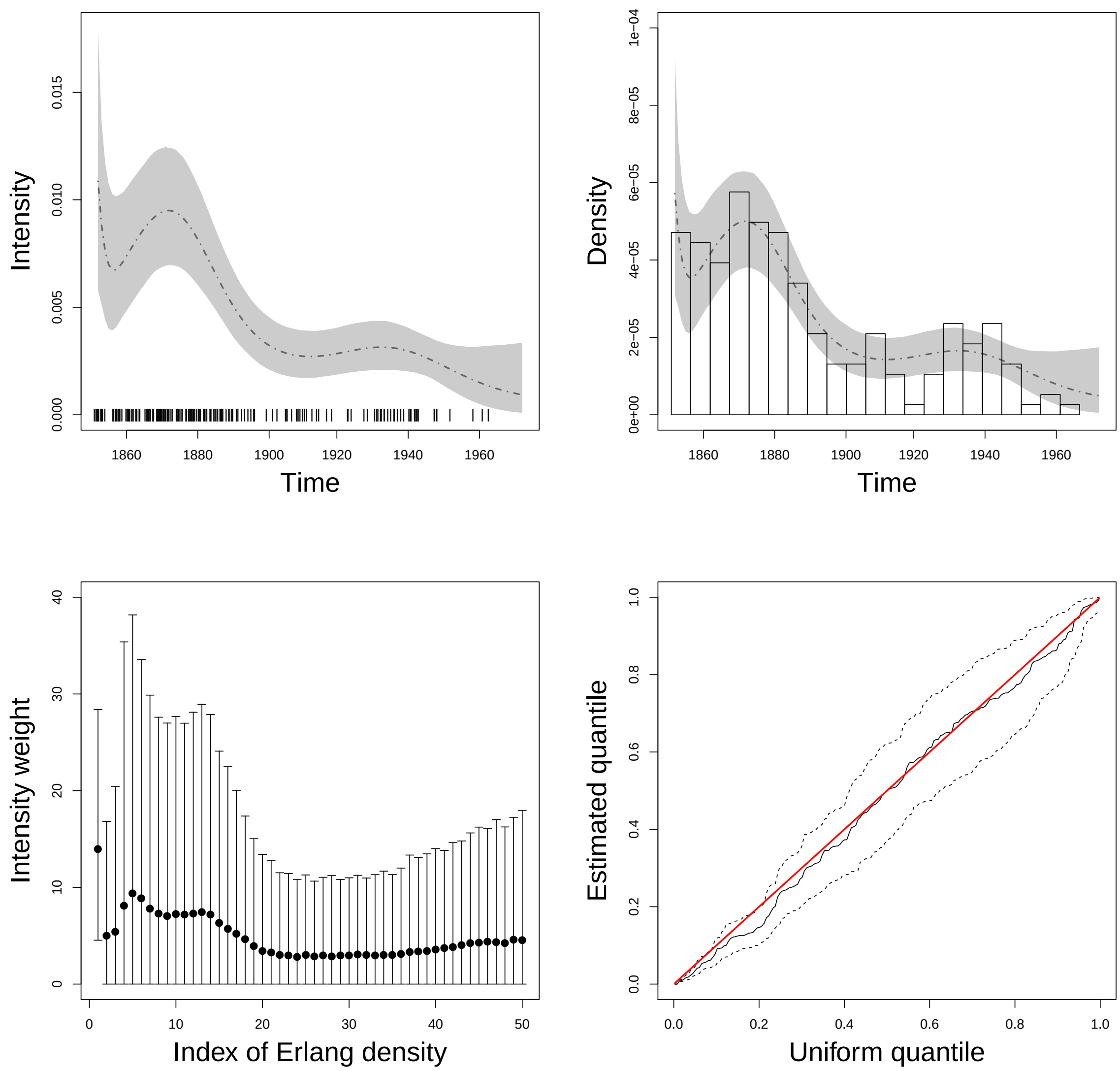


Figure 7: Coal-mining disasters data. The top left panel shows the posterior mean estimate (dashed-dotted line) and 95% interval bands (shaded area) for the intensity function. The bars at the bottom indicate the observed point pattern. The top right panel plots the posterior mean (dashed-dotted line) and 95% interval bands (shaded area) for the NHPP density, overlaid on the histogram of the accident times. The bottom left panel presents the posterior means (circles) and 90% interval estimates (bars) of the mixture weights. The bottom right panel plots the posterior mean and 95% interval bands for the time-rescaling model checking Q-Q plot.

based on quantile–quantile (Q-Q) plots to assess agreement of the estimated $U_i$ with the uniform distribution on the unit interval. Under the Bayesian inference framework, we can obtain a posterior sample for the $U_i$ for each posterior realization for the NHPP intensity, and we can thus plot posterior point and interval estimates for the Q-Q graph. These estimates suggest that the NHPP model with the Erlang mixture intensity provides a good fit for the coal-mining disasters data.

# 4 Modeling for spatial Poisson process intensities

In Section 4.1, we extend the modeling framework to spatial NHPPs with intensities defined on $\mathbb{R}^+ \times \mathbb{R}^+$. The resulting inference method is illustrated with synthetic and real data examples in Section 4.2 and 4.3, respectively.

## 4.1 The Erlang mixture model for spatial NHPPs

A spatial NHPP is again characterized by its intensity function, $\lambda(\boldsymbol{s})$, for $\boldsymbol{s} = (s_1, s_2) \in \mathbb{R}^+ \times \mathbb{R}^+$. The NHPP intensity is a non-negative and locally integrable function such that: (a) for any bounded $B \subset \mathbb{R}^+ \times \mathbb{R}^+$, the number of points in $B$, $N(B)$, follows a Poisson distribution with mean $\int_B \lambda(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$; and (b) given $N(B) = n$, the random locations $\boldsymbol{s}_i = (s_{i1}, s_{i2})$, for $i = 1, ..., n$, that form the spatial point pattern in $B$ are i.i.d. with density $\lambda(\boldsymbol{s})/\{\int_B \lambda(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\}$. Therefore, the structure of the likelihood for the intensity function is similar to the temporal NHPP case. In particular, for spatial point pattern, $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$, observed in bounded region $D \subset \mathbb{R}^+ \times \mathbb{R}^+$, the likelihood is proportional to $\exp\{-\int_D \lambda(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\} \prod_{i=1}^n \lambda(\boldsymbol{s}_i)$. As is typically the case in standard applications involving spatial NHPPs, we consider a regular, rectangular domain for the observation region $D$, which can therefore be taken without loss of generality to be the unit square.

Extending the Erlang mixture model in (1), we build the basis representation for the spatial NHPP intensity from products of Erlang densities, $\{\mathrm{ga}(s_1 \mid j_1, \theta_1^{-1})\,\mathrm{ga}(s_2 \mid j_2, \theta_2^{-1}) : j_1, j_2 = 1, ..., J\}$. Mixing is again with respect to the shape parameters $(j_1, j_2)$, and the basis densities share a pair of scale parameters $(\theta_1, \theta_2)$. Therefore, the model can be expressed as

$$\lambda(s_1, s_2) = \sum_{j_1=1}^{J} \sum_{j_2=1}^{J} \omega_{j_1 j_2} \, \text{ga}(s_1 \mid j_1, \theta_1^{-1}) \text{ga}(s_2 \mid j_2, \theta_2^{-1}), \quad (s_1, s_2) \in \mathbb{R}^+ \times \mathbb{R}^+.$$

Again, a key model feature is the prior for the mixture weights. Here, the basis density indexed by $(j_1, j_2)$ is associated with rectangle $A_{j_1 j_2} = [(j_1 - 1)\theta_1, j_1\theta_1) \times [(j_2 - 1)\theta_2, j_2\theta_2)$. The corresponding weight is defined through a random measure $H$ supported on $\mathbb{R}^+ \times \mathbb{R}^+$, such that $\omega_{j_1 j_2} = H(A_{j_1 j_2})$. This construction extends the one for the weights of the temporal NHPP model. We again place a gamma process prior, $\mathcal{G}(H_0, c_0)$, on $H$, where $c_0$ is the precision parameter and $H_0$ is the centering measure on $\mathbb{R}^+ \times \mathbb{R}^+$. As a natural extension of the exponential cumulative hazard used in Section 2.1 for the gamma process prior mean, we specify $H_0$ to be proportional to area. In particular, $H_0(A_{j_1 j_2}) = |A_{j_1 j_2}|/b = \theta_1 \theta_2 / b$, where $b > 0$. Using the independent increments property of the gamma process, and under the specific choice of $H_0$, the prior for the mixture weights is given by

$$\omega_{j_1 j_2} \mid \theta_1, \theta_2, c_0, b \overset{i.i.d.}{\sim} \text{gamma}(c_0 \, \theta_1 \, \theta_2 \, b^{-1}, c_0), \quad j_1, j_2 = 1, ..., J$$

which, as before, is a practically important feature of the model construction as it pertains to MCMC posterior simulation.

To complete the full Bayesian model, we place priors on the common scale parameters for the basis densities, $(\theta_1, \theta_2)$, and on the gamma process prior hyperparameters $c_0$ and $b$. The role played by these model parameters is directly analogous to the one of the corresponding parameters for the temporal NHPP model, as detailed in Section 2.1. Therefore, we apply similar arguments to the ones in Section 2.2 to specify the model hyperpriors. More specifically, we work with (independent) Lomax prior distributions for scale parameters $\theta_1$ and $\theta_2$, where the shape parameter of the Lomax prior is set equal to 2 and the scale parameter is specified such that $\Pr(0 < \theta_1 < 1)\Pr(0 < \theta_2 < 1) \approx 0.999$. Recall that the observation region is taken to be the unit square; in general, for a square observation region, this approach implies the same Lomax prior for $\theta_1$ and $\theta_2$. The gamma process precision parameter $c_0$ is assigned an exponential prior with mean 10. The result of Section 2.1 for the prior mean of the NHPP intensity can be extended to show that $\mathrm{E}(\lambda(s_1, s_2) \mid b, \theta_1, \theta_2)$ converges to $b^{-1}$, as $J \to \infty$, for any $(s_1, s_2) \in \mathbb{R}^+ \times \mathbb{R}^+$, and for any $(\theta_1, \theta_2)$ (and $c_0$). The prior mean for the spatial NHPP intensity is practically constant at $b^{-1}$ within its effective support given roughly by $(0, J\theta_1) \times (0, J\theta_2)$. Hence, taking the size of the observed spatial point pattern as a proxy for the total intensity, $b$ is assigned an exponential prior distribution with mean $1/n$. Finally, the choice of the value for $J$ can be guided from the approximate effective support for the intensity, which is controlled by $J$ along with $\theta_1$ and $\theta_2$. Analogously to the approach discussed in Section 2.2, the value of $J$ (or perhaps a lower bound for $J$) can be specified through the integer part of $1/\theta^*$, where $\theta^*$ is the median of the common Lomax prior for $\theta_1$ and $\theta_2$.

The posterior simulation method for the spatial NHPP model is developed through a straightforward extension of the approach detailed in Section 2.3. We work again with the augmented model that involves latent variables $\{\boldsymbol{\gamma}_i : i = 1, \dots, n\}$, where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2})$ identifies the basis density to which observed point location $(s_{i1}, s_{i2})$ is assigned. We omit the Gibbs sampler details, but note that the spatial NHPP model retains the practically relevant feature of efficient updates for the mixture weights, which, given the other model parameters and the data, have independent gamma posterior full conditional distributions.

## 4.2 Synthetic data example

Here, we illustrate the spatial NHPP model using synthetic data based on a bimodal intensity function built from a two-component mixture of bivariate logit-normal densities. Denote by $\mathrm{BLN}(\boldsymbol{\mu}, \Sigma)$ the bivariate logit-normal density arising from the logistic transformation of a bivariate normal with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$. A spatial point pattern of size 528 was generated over the unit square from a NHPP with intensity $\lambda(s_1, s_2) = 150\,\mathrm{BLN}((s_1, s_2) \mid \boldsymbol{\mu}_1, \Sigma) + 350\,\mathrm{BLN}((s_1, s_2) \mid \boldsymbol{\mu}_2, \Sigma)$, where $\boldsymbol{\mu}_1 = (-1, 1)$, $\boldsymbol{\mu}_2 = (1, -1)$, and $\Sigma = (\sigma_{11}, \sigma_{12}, \sigma_{21}, \sigma_{22}) = (0.3, 0.1, 0.1, 0.3)$. The intensity function and the generated spatial point pattern are shown in the top left panel of Figure 8.

The Erlang mixture model was applied setting $J = 70$ and using the hyperpriors for $\theta_1$, $\theta_2$, $c_0$ and $b$ discussed in Section 4.1. Figure 8 reports inference results. The posterior mean intensity estimate successfully captures the shape of the underlying intensity function. The structure of the Erlang mixture model enables ready inference for the marginal NHPP intensities associated with the two-dimensional NHPP. Although such inference is generally not of direct interest for spatial NHPPs, in the context of a synthetic data example it provides an additional means to check the model fit. The marginal intensity estimates effectively retrieve the bimodality of the true marginal intensity functions; the slight discrepancy at the second mode can be explained by inspection of the generated data for which the second mode clusters are located slightly to the left of the theoretical mode. Finally, we note the substantial prior-to-posterior learning for all model hyperparameters.

## 4.3 Real data illustration

Our final data example involves a spatial point pattern that has been previously used to illustrate NHPP intensity estimation methods (e.g., Diggle, 2014; Kottas and Sansó, 2007). The data set involves the locations of 514 maple trees in a 19.6 acre square plot in

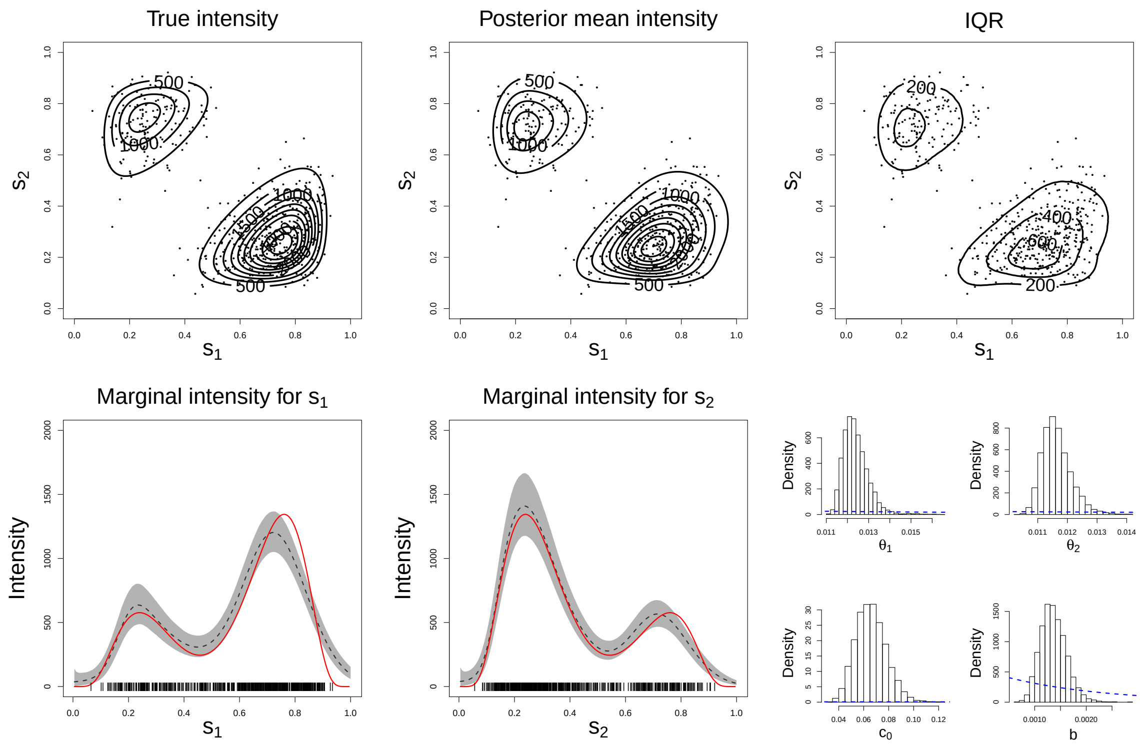


Figure 8: Synthetic data example from spatial NHPP. The top row panels show contour plots of the true intensity, and of the posterior mean and interquartile range estimates. The points in each panel indicate the observed point pattern. The first two panels at the bottom row show the marginal intensity estimates – posterior mean (dashed line) and 95% uncertainty bands (shaded area) – along with the true function (red solid line) and corresponding point pattern (bars at the bottom of each panel). The bottom right panel displays histograms of posterior samples for the model hyperparameters along with the corresponding prior densities (dashed lines).

Lansing Woods, Clinton County, Michigan, USA; the maple trees point pattern is included in the left column panels of Figure 9.

To apply the spatial Erlang mixture model, we specified the hyperpriors for $\theta_1$, $\theta_2$, $c_0$ and $b$ following the approach discussed in Section 4.1, and set $J = 70$. As with the synthetic data example, the posterior distributions for model hyperparameters are substantially concentrated relative to their priors; see the bottom right panel of Figure 9. The estimates for the spatial intensity of maple tree locations reported in Figure 9 demonstrate the model's capacity to uncover non-standard, multimodal intensity surfaces.

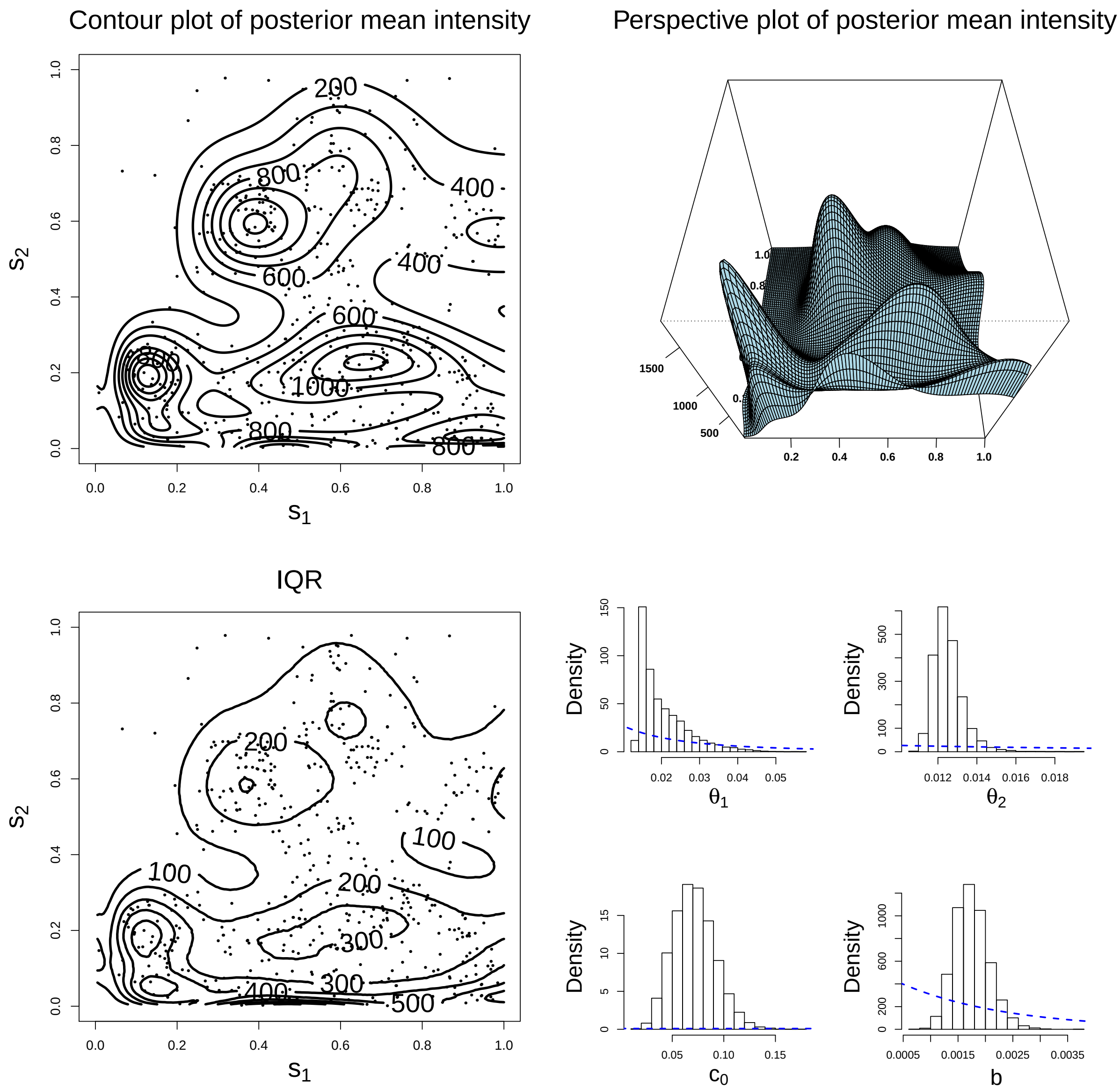


Figure 9: Maple trees data. The top row panels show the posterior mean estimate for the intensity function in the form of contour and perspective plots. The bottom left panel displays the corresponding posterior interquartile range contour plot. The bottom right panel plots histograms of posterior samples for the model hyperparameters along with the corresponding prior densities (dashed lines). The points in the left column plots indicate the locations of the 514 maple trees.

## 5 Summary

We have proposed a Bayesian nonparametric modeling approach for Poisson processes over time or space. The approach is based on a mixture representation of the point process intensity through Erlang basis densities, which are fully specified save for a scale parameter shared by all of them. The weights assigned to the Erlang densities are defined through increments of a random measure (a random cumulative intensity function in the temporal NHPP case) which is modeled with a gamma process prior. A key feature of the methodology, and the main motivation for its development, is that it offers a good balance between model flexibility and computational efficiency in implementation of posterior inference. Such inference has been illustrated with synthetic and real data for both temporal and spatial Poisson process intensities.

## References


Adams, R. P., Murray, I., and MacKay, D. J. C. (2009). Tractable nonparametric bayesian inference in poisson processes with gaussian process intensities. In *Proceedings of the 26th International Conference on Machine Learning*, Montreal, Canada.

Andrews, D. F. and Herzberg, A. M. (1985). *Data, A Collection of Problems from Many Fields for the Student and Research Worker*. Springer-Verlag.

Butzer, P. L. (1954). On the extensions of Bernstein polynomials to the infinite interval. *Proceedings of the American Mathematical Society*, 5:547–553.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes(2nd edition)*. Springer.

Diggle, P. J. (2014). *Statistical Analysis of Spatial and Spatio-Temporal Point Patert-erns(3rd edition)*. CRC press.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press.

Heikkinen, J. and Arjas, E. (1998). Non-parametric Bayesian estimation of a spatial Poisson intensity. *Scandinavian Journal of Statistics*, 25:435–450.

Heikkinen, J. and Arjas, E. (1999). Modeling a Poisson forest in variable elevations: A nonparametric Bayesian approach. *Biometrics*, 55:738–745.

Ishwaran, H. and James, L. F. (2004). Computational methods for multiplicative intensity models using weighted gamma processes: Proportional hazards, marked point processes, and panel count data. *Journal of the American Statistical Association*, 99:175–190.

Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society. Series B*, 40:214–221.

Kang, J., Nichols, T. E., Wager, T. D., and Johnson, T. D. (2014). A Bayesian hierarchical spatial point process model for multi-type neuroimaging meta-analysis. *The Annals of Applied Statistics*, 8:1800–1824.

Kingman, J. F. C. (1993). *Poisson Processes*. Clarendon Press.

Kottas, A. and Sansó, B. (2007). Bayesian mixture modeling for spatial poisson process intensities, with applications to extreme value analysis. *Journal of Statistical Planning and Inference*, 137:3151–3163.

Lee, S. C. K. and Lin, X. S. (2010). Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North American Actuarial Journal*, 14:107–130.

Lo, A. Y. (1982). Bayesian nonparametric statistical inference for Poisson point processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 59:55–66.

Lo, A. Y. (1992). Bayesian inference for Poisson process models with censored data. *Journal of Nonparametric Statistics*, 2:71–80.

Lo, A. Y. and Weng, C. S. (1989). On a class of Bayesian nonparametric estimates: ii. hazard rate estimates. *Annals of the Institute for Statistical Mathematics*, 41:227–245.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25:451–482.

Petrone, S. (1999a). Bayesian density estimation using Bernstein polynomials. *Canadian Journal of Statistics*, 27:105–126.

Petrone, S. (1999b). Random Bernstein polynomials. *Scandinavian Journal of Statistics*, 26:373–393.

Rodrigues, A. and Diggle, P. J. (2012). Bayesian estimation and prediction for inhomogeneous spatiotemporal log-Gaussian Cox processes using low-rank models, with application to criminal surveillance. *Journal of the American Statistical Association*, 107:93–101.

Rodriguez, A., Wang, Z., and Kottas, A. (2017). Assessing systematic risk in the S&P500 index between 2000 and 2011: A Bayesian nonparametric approach. *The Annals of Applied Statistics*, 11:527–552.

Taddy, M. (2010). Autoregressive mixture models for dynamic spatial Poisson processes: Application to tracking the intensity of violent crime. *Journal of the American Statistical Association*, 105:1403–1417.

Taddy, M. A. and Kottas, A. (2012). Mixture modeling for marked Poisson processes. *Bayesian Analysis*, 7:335–362.

Venturini, S., Dominici, F., and Parmigiani, G. (2008). Gamma shape mixtures for heavy-tailed distributions. *The Annals of Applied Statistics*, 2:756–776.

Wolpert, R. L. and Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika*, 85:251–267.

Xiao, S., Kottas, A., and Sansó, B. (2015). Modeling for seasonal marked point processes: An analysis of evolving hurricane occurrences. *The Annals of Applied Statistics*, 9:353–382.

Xiao, S., Kottas, A., Sansó, B., and Kim, H. (2020). Nonparametric Bayesian modeling and estimation for renewal processes. *Technometrics*, To appear.