Spatiotemporal point processes: regression, model specifications and future directions

Dani Gamerman

UFRJ

Abstract. Point processes are one of the most commonly encountered observation processes in Spatial Statistics. Model-based inference for them depends on the likelihood function. In the most standard setting of Poisson processes, the likelihood depends on the intensity function, and can not be computed analytically. A number of approximating techniques have been proposed to handle this difficulty. In this paper, we review recent work on exact solutions that solve this problem without resorting to approximations. The presentation concentrates more heavily on discrete time but also considers continuous time. The solutions are based on model specifications that impose smoothness constraints on the intensity function. We also review approaches to include a regression component and different ways to accommodate it while accounting for additional heterogeneity. Applications are provided to illustrate the results. Finally, we discuss possible extensions to account for discontinuities and/or jumps in the intensity function.

1 Introduction

Point pattern data or point process is an observation framework where events occur at random locations in a given region of interest S. This region is usually associated with space. The response of such process is denoted by Y, that can be also be identified with $\{s_1, \ldots, s_N\}$, the locations of the observed events. Examples include location of occurrences of a disease in a city, of a plant in a forest or of fires in a state or province. Figure 1 illustrates one of these situations.

One of the most common models for point pattern is the Poisson process. The model is specified by an intensity function λ. The likelihood is given by

$$l(\lambda; y) = \prod_{i=1}^{N} \lambda(s_i) \exp\left\{-\int_S \lambda(s)\, ds\right\}.$$

Note that the likelihood function for λ depends on the entire function and its evaluation depends on the ability to integrate λ. When this function is entirely unknown, the integration can not be performed. Solutions involve numerical approximation of the integral (Liang et al., 2008) or to approximate the intensity function by some representation (Dias et al., 2008).
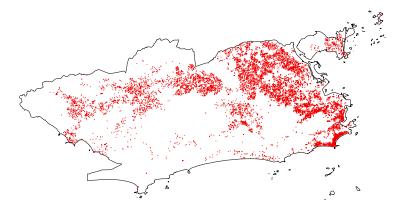
Figure 1 *Deaths from cerebrovascular disease in the city of Rio de Janeiro*, 2002–2007. *Each red dot is the location of the residence of a deceased* (*Pinto Jr.*, 2014).

In the case of a spatiotemporal point pattern, continuous time may be identified with the spatial case with an added dimension. The response Y for discrete times $\{t_1, \ldots, t_T\} = \mathcal{T}$ may be denoted $Y = (Y_1, Y_2, \ldots, Y_T)$, and each Y_t can be identified with $\{s_{t,1}, \ldots, s_{t,N_t}\}$ and has intensity λ_t, for $t = 1, \ldots, T$. The likelihood function is given by

$$l(\lambda; y) = \prod_{t=1}^{T} \prod_{i=1}^{N_t} \lambda_t(s_{i,t}) \exp\left\{ -\sum_{t=1}^{T} \int_S \lambda_t(s) \, ds \right\}.$$

Poisson processes are very useful models but rely on independence across space (and time). Thus, that may not be the most useful representation of situations where aversion or attraction of events take place. Other components must be added to the model to mitigate this deficiency. One such framework is provided by Cox processes (Cox, 1955). Cox processes assume that the intensity function is random and its distribution might induce dependence features in the marginal specification of the process.

A very frequently used Cox process assumes an isotropic Gaussian process (GP) for the log intensity. It was proposed by Moller et al. (1998) and is called log Gaussian Cox Process (LGCP). Letting $\eta = \log \lambda$, the LGCP specifies that $\eta \sim \text{GP}(\mu, \sigma, \rho_\theta)$, which means that for any $r > 1$, and any set of locations $\{s_1, \ldots, s_r\} \in S$, the joint distribution of $(\eta(s_1), \ldots, \eta(s_r))$ is multivariate normal with mean $\mu 1_r$ and covariance matrix R with elements $\sigma \rho_\theta(|s_i - s_j|)$, for all (i, j). The LGCP is implemented in a few statistical packages (see, for example, Brown (2015) and Taylor et al. (2013)). Isotropy is not required for model specification but is frequently imposed for computational tractability.

The logarithmic transformation of the intensity function is basically used for convenience but any other monotonic transformation could be equally used. These ideas have been explored in recent years and will be extensively used in the sequel.

The remainder of this paper is organized as follows. Section 2 addresses the incorporation of regression components. Section 3 provides the model definition with regression and revisits the inference difficulties above in this context. Section 4 presents one solution based on discretization. Section 5 presents another solution based on data augmentation. The paper is concluded in Section 6 with directions for future work. The structure of the paper follows closely my presentation at the Opening Address of the 14th Brazilian Meeting on Bayesian Statistics (EBEB XIV).

2 Regression

Point processes may be subject to the effect of explanatory variables like in any other area of Statistics. The standard approach in these settings is to add a linear predictor $W'\beta$ to the existing model. In the more general case of a spatiotemporal point processes this would imply

$$\lambda_t(s) = g[\beta_{0,t}(s) + W'\beta], \quad \text{for } (s,t) \in S \times \mathcal{T}.$$

where W is a p-dimensional vector of explanatory variables and g is some link function. In the case of LGCP, $g = \exp$.

Some applications show excess spatiotemporal heterogeneity making the separation between the intercept and the regression component not reasonable. One simple and flexible strategy to accommodate added variation is to allow for space/space-time varying regression coefficients. This idea was applied with success to other branches of Spatial Statistics. See, for example, the work of Gelfand, Banerjee and Gamerman (2005) on continuous spatial data or the work of Gamerman, Moreira and Rue (2003) on areal data.

This would translate in point pattern data as $\lambda_t(s) = g[\beta_{0,t}(s) + W'\beta_t^*(s)]$, for $(s,t) \in S \times \mathcal{T}$. The model becomes assumedly more complex but note that the extra complexity is mostly due to the spatiotemporal variability of β_0, now extended to all the other regression coefficients. Given the similarity between the nature of $\beta_{0,t}$ and β_t^*, they will be hereafter merged into $\beta_t = (\beta_{0,t}, \beta_t^*)$ and an additional component 1 is concatenated into W so that the model becomes

$$\lambda_t(s) = g[W'\beta_t(s)], \quad \text{for } (s,t) \in S \times \mathcal{T}.$$

The above model accommodates for local variation of effects that is sometimes encountered in such settings. This variation may provide insight into the nature of phenomenon under study. Note that it includes standard, fixed effects as the special case when there is no spatiotemporal variation of the effects. Other special cases are models where effects vary only in space but not in time and models where effects vary only in time but not in space.

There is an important distinction to be clarified among regressors. In the case these vary only over space/time (Benes et al., 2005), the above extension is

straightforward. Care must be exercised to avoid identification issues associated with the concurrent spatiotemporal variation of both regressors and their coefficients as these only appear in the intensity function via their product.

In the case they vary also over external configurations (Liang et al., 2008; Diggle et al., 2010), then (conditionally independent) point processes Y_v must be considered, each with its own intensity λ_v, for all possible values of the external configuration v. This is exactly what is done in any other regression context. Thus, W becomes W_v and $\lambda_t(s)$ becomes $\lambda_{t,v}(s)$, for all $(s, t) \in S \times \mathcal{T}$. This setting is particularly useful in epidemiological applications where individual characteristics such as age and gender are likely to affect the intensity function.

In the space-only context, the model for the random component process β is a multivariate GP, in a direct extension of the model without covariates. All that remains for model completion is the specification of the temporal variation of process β_t. A simple solution if a product of independent GP's. This route is not recommended for a number of reasons. Just like one expects spatial similarity between its values, one would expect temporal similarity between its values. This implies some form of relation between values of β_t. This relation would also help in borrowing information across time, making the inference more reliable.

One possible solution is provided by dynamic GP reviewed in Gamerman (2010). The link across time is provided by

$$\beta_{t'} = G_{t,t'}\beta_t + w_t, \quad \text{where } w_t \sim \text{GP}$$

$$\text{with } \beta_1 \sim \text{GP},$$

where $G_{t,t'}$ is the transition matrix describing the deterministic part of the evolution across time and the process disturbances w_t accommodate for possible deviations from this specification. This process is denoted by $\beta \sim \text{DGP}$. Note that if the process disturbances w_t vanish and $G = I$ then $\beta_t = \beta_1$, for all t, then the static process is recovered.

DGP are quite flexible and accommodate for stochastic space-varying trends and seasonality. They also induce sparcity over temporal components due to the use of first order Markovian evolution, thus helping to improve computations is such richly parametrized setting.

Continuous-time version of this model was proposed in Diggle and Brix (2001). More convolved forms of temporal evolution were proposed in Wikle and Cressie (1999). Both extensions require approximations to work. We would like solutions to be as exact as possible and thus retain our DGP specification in the sequel.

3 Model and inference

3.1 Model

The observation process in the spatiotemporal setting is given by $Y = (Y_1, Y_2, \ldots, Y_T)$ with $Y_t \sim \text{PP}(\lambda_t)$, for $t \in \mathcal{T}$, where PP stands for Poisson process. The link

with the regressors is established via the transformation $\lambda_t(s) = g[W'\beta_t(s)]$, for some g. If data varies over individual configurations v then $Y_t = \{Y_{t,v}\}$ with $Y_{t,v} \sim \mathrm{PP}(\lambda_{t,v})$ and $\lambda_{t,v}(s) = g[W_v'\beta_t(s)]$, for $v \in \mathcal{V}$, the space of all possible configurations. The model is completed with $\beta \sim \mathrm{DGP}$ and a prior distribution for θ, the collection of all other unknowns.

The above formulation includes many other models previously considered. Examples are:

- $\lambda(s) = g[\beta_0(s) + W'\beta]$ (Benes et al., 2005);
- $\lambda_v(s) = g[\beta_0(s) + W_v'\beta]$ (Liang et al., 2008; Diggle et al., 2010);
- $\lambda_t(s) = g[\beta_{0,t}(s)]$ (Reis et al., 2013).

In the current setting, θ is given by the DGP hyperparameters but more general model forms may require additional hyperparameters. One example is provided by evolutions with temporal transition $G_{t,t'}$ depending on unknown constants (Gamerman, 2010). In any case, inference for these higher levels parameters is difficult due to scarcity of information unless substantial amount of data is available. In these scenarios, ad-hoc procedures are frequently applied.

Frequentist inference is also possible for these models. Diggle and Brix (2001) suggest the use of empirical Bayes procedures to estimate θ. They used plug-in estimation for θ based on the method of moments. This approach has the advantage of by-passing the difficulties associated with lack of information in likelihood-based methods.

3.2 Inference (problem)

There are a few important issues to solve in order to perform inference for these models. The first issue is that there is no explicit form for the (joint) density of GPs due to its infinite dimensionality. However, it is useful to know that finite dimension versions do exist. This point will be explored in our approaches. The most serious issue however is the fact that the integral in the likelihood prevents its exact calculation. Thus posterior distributions can not be computed as well.

Some solutions to the above issues involve:

- parametric forms for the IF and/or β;
- approximation of the integral in the likelihood.

Both solutions above impose approximations and we would like our approach to be as exact as possible. Also, our desiderata list includes:

- having a fully model-based approach so that our assumptions can be readily checked and changed if necessary;
- retain the intensity function exactly as specified, without any representation.

The next two sections present different solutions that respect the above list, while presenting inference exactly for the model considered, instead of a representation and/or an approximation of it.

4 Discretization

Even tough the model is specified at the point-wise level, some situations require regional rather than a point-wise approach. These include the cases where all relevant inference to be drawn from the model is made at the regional level. In these scenarios, little is lost by assuming piece-wise constancy of the intensity function or the components that make it up. Once again, important and heavily used examples come from epidemiological studies where health decisions are not made at the point level but at the regional level where countries, states and provinces are organized at.

Therefore, this section follows Pinto Jr. et al. (2015) by assuming that $\lambda_t(s) = \lambda_{i,t}$, for $s \in R_i \subset S$ and for all i, t. The sets $\{R_i\}$ form a partition of S, with $a_i = \text{vol}(R_i)$, for all i. Shape, size and cardinality of $\{R_i\}$ depends on the phenomenon under study.

The likelihood function then becomes $l(\lambda; y) = \prod_t \prod_i \lambda_{i,t}^{N_{i,t}} \exp\{-\sum_t \sum_i a_i \lambda_{i,t}\}$. This is basically equivalent to observing data $N_{i,t} \sim \text{Poisson}(a_i \lambda_{i,t})$, $\forall(i,t)$. The only difference being the inclusion of the multiplicative term $[\prod_t \prod_i a_i^{N_{i,t}}]/\prod_t \prod_i N_{i,t}!$ to the likelihood from observing the $\{N_{i,t}\}$. This extra term is irrelevant for making inference about model parameters but is required for predictive quantities, used in some model comparison tools.

Since our framework was specified from a regression perspective with regression coefficients β, these must also be simplified to $\beta_t(s) = \beta_{i,t}$, for $s \in R_i \subset S$. the link between the discretized intensity and regression coefficients is provided by $\lambda_{i,t} = g(W'\beta_{i,t})$, $\forall(i,t)$. The most common link is the exponential but other monotonically increasing links could be used just as well.

The extension towards individual configurations is easily accommodated by noting the likelihood is $l(\lambda; y) = \prod_v l(\lambda_v; y)$, which is basically equivalent to observing data $N_{i,t,v} \sim \text{Poisson}(a_i \lambda_{i,t,v})$, $\forall(i, t, v)$ and the observation part of the model is completed with the link $\lambda_{i,t,v} = g(W'_v \beta_{i,t})$, $\forall(i, t, v)$, with the regression process β.

The latter process becomes finite-dimensional with values $\beta = (\beta_1, \beta_2, \ldots, \beta_T)$ where $\beta_t = \{\beta_{i,t}, \forall i\}$. The joint specification of this finite process becomes $\beta_{t'} = G_{t,t'}\beta_t + w_t$, the finitely-dimensional disturbances w_t and the initial process state β_1 are given multivariate normal distributions. These distributions may be induced by the underlying GP with covariance specification between pairs of units based on the some location of the units. The most common location choice to represent unit i is the centroid of R_i but other summarizers could be used.

The framework becomes similar to the Poisson version of the dynamic areal models of Vivar and Ferreira (2009). Inference for these models follow closely the time series literature. Examples of MCMC methods for these models include Gamerman (1998) and Fruhwirth-Schnatter and Wagner (2006). Standard, general-

purpose software for Bayesian inference can also be used here. They may require substantial computational overhead for large T since most of them do not make use of blocks and/or reparametrization. As a result, the typically strong temporal dependence present in the β process slows MCMC convergence down. In any case, the computational cost depend on the cardinality of the partition $\{R_i\}$).

4.1 Application

The Superintendência de Seguros Privados (SUSEP, in short) is a government entity concerned with regulation of insurance in Brazil. They contain a database of many sources of information. This illustration is concentrated on a specific dataset about vehicle policies in the city of Rio de Janeiro and the time span consist of the years 2009, 2010 and 2011. The dataset contains information on all insurance policies contracted in this period along with details about the insured vehicle. Our focus is on the assessment of the spatiotemporal variation of loss (theft and robbery) occurrence rate and of the possible effect that vehicle characteristics may have over this rate.

The nature of the problem is imposed by the need of insurance companies to have workable tables to formulate their policies and premiums. Thus, the study was discretized by regions $\{R_i\}$ defined according to the first 3 digits of the postal code. Figure 2 shows this result of this operation to the area of interest.

In this application, time is treated as discrete with time units given by semester, thus rendering a total of 6 time units. This unit is chosen because the database recorded the total exposure for each region with a half-yearly resolution, even though the losses are temporally registered in the database with monthly resolution. The explanatory variables chosen are x_1, the manufacturing year of the
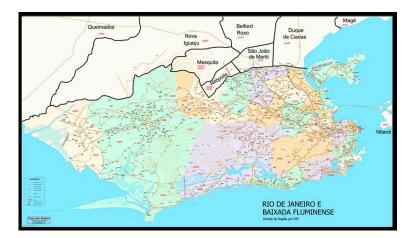

Figure 2 *City of Rio de Janero, colored according to discretization provided by the first 3 digits of the postal code of the location of the loss (Pinto Jr., 2014).*

vehicle (in years), and x_2, an indicator of the nature of the vehicle usage (personal/commercial). The total exposure $r_{v,i,s}$ per covariate configuration v, region R_i and semester t is used as an offset. It plays the crucial role of standardization thus allowing for valid comparisons across space-time-covariate configuration. The model for the point processes of occurrences $\{Y_v, \forall v\}$ can be written as

$$Y_v \sim \mathrm{PP}\big(\Lambda_v(s,t)\big), \quad \forall v$$

$$\Lambda_v(s,t) = r_{v,i,t}\lambda_{v,i,t}, \quad \text{for } s \in R_i, \forall(v,i,t)$$

$$\log \lambda_{v,i,t} = \alpha_{0,i,t} + \alpha_{1,i,t}x_{1,v} + \alpha_{2,i,t}x_{2,v}, \quad \forall(v,i,t)$$

$$\alpha_{l,t} = \alpha_{l,t-1} + w_{l,t}, \quad \text{where } w_{l,t} \sim N(0, R_l), \text{ for } l = 0,1,2 \text{ and } \forall t > 1,$$

where $\alpha_{l,t} = \{\alpha_{l,i,t}, \forall i\}$ is the set of regression coefficients at time t and the disturbances $w_{l,t}$ have their normal distribution drawn from the GP's that drive the evolution of DGP, for $l = 0, 1, 2$.

Figure 3 presents the temporal evolution of a subset of regions for the intercept α_0. The figure seem to indicate a general decreasing trend in losses over the period
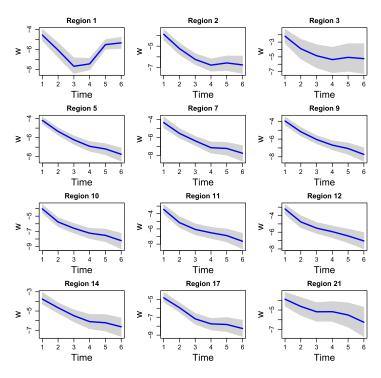

Figure 3 *Estimation summary of the temporal evolution of the intercept α_0 for some selected regions: median (full line) and 95% credibility intervals (shaded area). Extracted from Pinto Jr. (2014).*
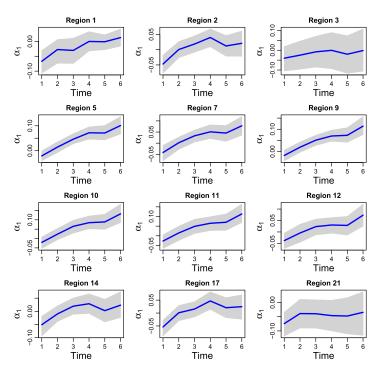
Figure 4   *Estimation summary of the temporal evolution of the coefficient α_1 of the vehicle manufacturing year for some selected regions*: *median* (*full line*) *and* 95% *credibility intervals* (*shaded area*). *Extracted from Pinto Jr.* (2014).

studied. This decrease is significant for most regions despite the small temporal length of the study window. This pattern exhibit substantial variation across space, with some regions presenting steeper declines and a few regions not providing evidence of any decrease. This spatial variation of the intercept only emphasize the importance of allowing them to vary over space, as expected.

Figures 4 and 5 present the temporal evolution of a subset of regions for the regression coefficients α_1 and α_2. They also show spatial variation for both regression coefficients for many regions. This variation is not contemplated in most regression models and therefore is typically not expected in such applications. They also show significant temporal variation for a number of regions. The effect of the manufacturing year is mostly positive indicating a higher loss intensity for newer vehicles. But the most striking feature of Figure 4 is the upward trend exhibited by many regions, indicating a temporal increase in the loss intensity for newer cars. Once again, the pattern is far from similar across regions, indicating the need for spatial variation of effects. The temporal variation of the effect of the nature of the vehicle usage seems to have a U-shaped form over the time span of the study, but the uncertainty of the estimates prevents further elaboration. It seems safer to assume a constant temporal trend of this effect overall.
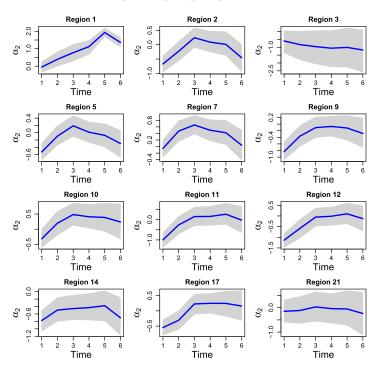
Figure 5 *Estimation summary of the temporal evolution of the coefficient α_2 of the nature of the vehicle usage for some selected regions*: *median* (*full line*) *and* 95% *credibility intervals* (*shaded area*). *Extracted from Pinto Jr.* (2014).

Nevertheless, many regions exhibit significantly positive and negative effects at some time periods. As before, the plots of Figure 5 exhibits a variety of shapes of temporal variation and uncertainty bounds highlighting the need for spatial variation.

5 Data augmentation

The material of the previous section is useful when one is prepared to assume local constancy of the intensity. More general scenarios require estimation of an intensity function varying over a continuous domain in space. However this scenario leads to an intractable situation due to the evaluation of the integral in the likelihood.

The solution comes from an augmentation approach, whose origin can be traced back to the Poisson thinning algorithm by Lewis and Shedler (1979), described in Algorithm 1.

This algorithm provides a computational description of an augmentation process X to the underlying process Y. The complication brought by the non-homogeneity

Algorithm 1 Simulation from $Y \mid \lambda \sim \text{PP}(\lambda)$ (space only)

(1) set $\lambda^* = \sup_s \lambda(s)$;
(2) sample $X \sim \text{PP}(\lambda^*)$, where X is an augmented data process. This is performed in two steps:
    (2a) sample $K \sim \text{Poisson}(\lambda^* \text{vol}(S))$;
    (2b) distribute K points $\{s_k\}_{k=1}^K$ uniformly over S;
(3) retain each point s_k with probability $\lambda(s_k)/\lambda^*$, $\forall k$;
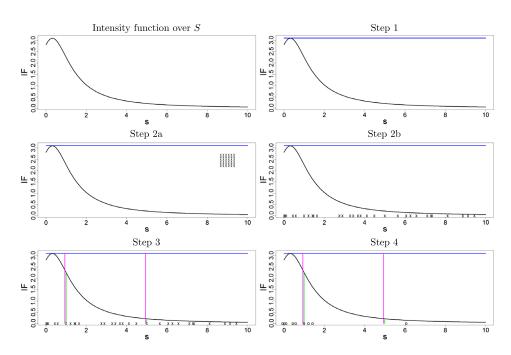(4) $\{s_i\}_{i=1}^N$ retained points are a sample from Y.

Figure 6 *Graphical representation of Algorithm* 1 *for a unidimensional spatial domain S. The K sampled points are presented at the top left corner in Step* 2a *and distributed to locations over the spatial domain in Step* 2b. *Two locations have their intensities values presented in Step* 3 *and compared to the intensity of the augmented process. The location to the left (left) (right) has a much higher (lower) acceptance rate and ends up accepted (rejected), as shown in Step* 4.

of the intensity function λ of Y is basically solved by the homogeneity of the intensity function of X. It augments the data process Y because its intensity is uniformly larger than (and blankets) the intensity of Y. The points generated from X are selected in a rejection procedure. This procedure is facilitated by the independence of Poisson processes and can thus be performed point by point independently. The steps of the algorithm are illustrated in Figure 6 for a unidimensional spatial domain, merely for ease of exposition.

The analytic version of Algorithm 1 is described by

$$\pi(K, x, y \mid \lambda) = \pi(K \mid \lambda^*) \times \pi(x \mid K) \times \pi(y \mid x, \lambda)$$

$$= e^{-\lambda^* \operatorname{vol}(S)} \frac{[\operatorname{vol}(S)\lambda^*]^K}{K!} \times \left[\frac{1}{\operatorname{vol}(S)} \right]^K$$

$$\times \prod_{i=1}^{N} \frac{\lambda(s_i)}{\lambda^*} \times \prod_{j=1}^{K-N} \left[1 - \frac{\lambda(s_j)}{\lambda^*} \right]. \tag{1}$$

The rejected points $\{s_i\}_{i=N+1}^K$ are sample from $\tilde{Y}$, the complement of Y with respect to X. It is useful to note that $\tilde{Y} = X \backslash Y \sim \operatorname{PP}(\lambda^* - \lambda)$ and that Y and $\tilde{Y}$ are independent conditionally on the intensity λ. Thus, the augmented data X is obtained as a result of the augmentation of real data Y by latent data $\tilde{Y}$.

The extension to the general space-time case is straightforward. The algorithm to simulate from $Y = (Y_1, \ldots, Y_T) \mid \lambda \sim \operatorname{PP}(\lambda)$ can be obtained after recalling that $Y_t \mid \lambda_t \sim \operatorname{PP}(\lambda_t), \forall t$, are conditionally independent point processes given λ. So, a sample from Y is obtained by repeating Algorithm 1 for all t, with augmented process $X = (X_1, \ldots, X_T)$, with intensities $\lambda_t^* = \sup_s \lambda_t(s), \forall t$. The sample from Y is given by $\{\{s_{i,t}\}_{i=1}^{N_t}, \forall t\}$. $\{\{s_{i,t}\}_{i=N_t+1}^{K_t}, \forall t\}$ is the set of rejected points and consist on a sample from $\tilde{Y} = (\tilde{Y}_1, \tilde{Y}_2, \ldots)$, where $X_t \backslash Y_t = \tilde{Y}_t \sim \operatorname{PP}(\lambda_t^* - \lambda_t), \forall t$. The analytic version of the algorithm now is simply a product of the version above for the space-only case.

In either case, the integral disappeared from the augmented likelihood and it depends on a finite subset of λ. These features will be crucial for exact evaluation of the likelihood and for easy sampling of model parameter processes, respectively.

The requirement of a bounding constant λ^* forces changes in the model specification. In order to benefit from the advantages of Algorithm 1 the link function g relating the covariates to the intensity function must be bounded. This seems like a small price to pay in many applications as typically there is no physical need for unlimited values of the intensity.

Thus, the intensity function is now given by $\lambda(s) = \lambda^* g[W'\beta(s)]$, where the link function g must be a monotonically increasing, taking values in the unit interval. So, any distribution function can be used. Adams et al. (2009) proposed the use of Algorithm 1 for inference of Poisson process intensities, in conjunction with a logistic or sigmoidal link. Gonçalves and Gamerman (2018) chose the probit link where $g = \Phi$, the distribution function of the standard normal distribution. It is basically undistinguishable from the sigmoidal link but presents useful computational advantages.

The model is completed with independent prior distributions for the regression coefficient processes β and the maximum intensity λ^*. The former is given by a multivariate GP, possibly depending on hyperparameters θ, and the latter can be given any distribution over the positive semi-line. An obvious choice is the Gamma

distribution, that is conditionally conjugate. Details about these specifications are provided in Gonçalves and Gamerman (2018).

In the sequel, we will use the notation β_N and W_K to represent the values of β at the N accepted locations and W at all K locations. Similar reasoning is valid to define β_K, β_{K-N}, W_N and β_{K-N}.

5.1 Inference

The presentation here will be restricted to the space-only case with only spatial covariates. The other cases will be briefly addressed in the sequel. Inference must be performed on all unknown quantities. These are the model parameters β and λ^*, the hyperparameters θ and the latent data $\tilde{y}$.

The Bayesian inference for these unknowns is based on their posterior distribution. Bayes theorem informs that the posterior distribution is based on the product of prior distribution and the (augmented) likelihood. Thus,

$$\pi(\tilde{y}, \beta, \lambda^*, \theta \mid y) \propto l(\tilde{y}, \beta_K, \lambda^*; y) \times \pi(\beta_K, \beta_{-K}, \lambda^*, \theta),$$

where $\beta = (\beta_K, \beta_{-K})$, the (augmented) likelihood is given by

$$l(\tilde{y}, \beta, \lambda^*; y) = \pi(y, \tilde{y} | \beta_K, \lambda^*)$$
$$= e^{-\lambda^* \operatorname{vol}(S)} \frac{[\lambda^*]^K}{K!} \times \prod_{i=1}^{N} \Phi[W_i' \beta(s_i)] \prod_{j=1}^{K-N} \Phi[-W_j' \beta(s_j)],$$

and the prior distribution is represented by

$$\pi(\beta_K, \beta_{-K}, \lambda^*, \theta) = \pi_{\mathrm{GP}}(\beta_{-K} \mid \beta_K, \theta)\pi_{\mathrm{GP}}(\beta_K \mid \theta)\pi_\lambda(\lambda^*)\pi(\theta).$$

The above results show that the likelihood depends on β only through β_K. Thus, inference about β can be split into two steps: inference for β_K based on the data, followed by inference on β_{-K} conditional only on β_K and θ. The inference procedure for β is basically reduced to finitely many values of β, like in the discretization approach of the previous section.

The complicated nature of the posterior distribution precludes exact posterior calculations. One of the attractive option for inference is MCMC (Gamerman and Lopes, 2006). This can be implemented to the current setting by block sampling from full conditional distribution of components of the array of unknowns as follows:

$\tilde{y}$: sample from Poisson($\lambda^* - \lambda$), using Algorithm 1;

β_K: sample directly from full conditional distribution that is multivariate skew normal;

λ^*: sample directly from full conditional distribution that is Gamma, if prior for λ^* is Gamma;

θ: sample with a Metropolis–Hastings proposal;

β_{-K}: sample required finite components of β_{-K} via kriging using the GP prior.

Sampling $\tilde{y}$ requires retrospective sampling of components of β_{-K} at the new chosen locations. The multivariate skew normal distribution of β_K stems from the combination of the multivariate normal prior distribution with the product of probit functions from the augmented likelihood. Details of the above sampling operations are provided in Gonçalves and Gamerman (2018).

Extension of these results to the dynamic setting is not difficult but depend on specification of dynamic extensions of λ^* and β. Appropriate temporal evolutions for these quantities must be specified. DGPs seem a natural choice for β. A computationally convenient evolution for the dynamic λ_t^*'s was provided by Gamerman et al. (2013).

These results allow for inference for any computable function of parameters. Among these, stand out the integrated intensity $\int_C \lambda(s)\,ds$, for any subset C of the region of interest. It is the mean of the Poisson distribution of the number of occurrences in C and is a useful tool for the assessment of model fit and also for prediction.

5.2 Comments

This line of work is an extension of the work done by Adams et al. (2009). There are a few important differences however. Adams et al. (2009) did not consider the temporal extension, even tough it can be trivially adapted in the continuous time setting. They did not consider covariates and therefore they obviously did not consider space(-time) variation of regression coefficient. Also, their choice of the sigmoidal link makes it difficult to provide easy sampling of β. They resorted to Hamiltonian MCMC which rendered a more convoluted coding and more costly mixing in terms of time and convergence speed. Figure 7 illustrates the relative mixing merits of both proposals.
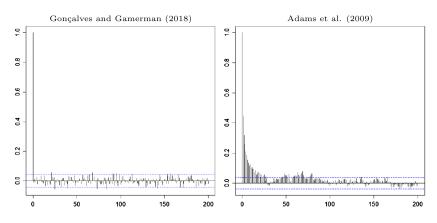

Figure 7 *Autocorrelation function of* $-2\log$ *posterior density for the two approaches in a simulated example (from Gonçalves and Gamerman (2018)).*

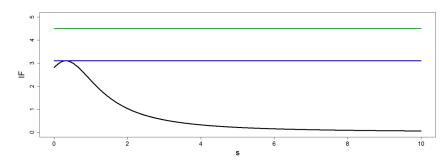**Figure 8** *Introduction of an upper bound $\lambda^\dagger$ (in red) larger than the maximum λ^* (in blue) to the intensity.*

Extra care must be exercised when setting priors for the maximum intensity λ^*. Any $\lambda^\dagger > \lambda^*$ is a valid constant for the intensity of the augmented data X of Algorithm 1. The larger the value of $\lambda^\dagger$, the less efficient the computation becomes. Figure 8 illustrates this issue as the expected value of the number K of drawn locations is proportional to the maximum intensity used and the value of K governs the computational cost.

The augmentation procedure may also be inefficient if the intensity function exhibit substantial variation in magnitude even if λ^* is used. In the example of Figure 8, only around 20% of the K drawn points using λ^* are accepted. The acceptance rate would reduce to around 14% if point were drawn using $\lambda^\dagger$.

5.3 Application

Fires in forest land are a major source of concern for the society. As an example recent fires throughout the globe caused a sizeable number for fatal casualties among neighboring citizens and visitors to these areas. The study of their occurrence may indicate spatial and/or temporal trends and thus provide directions for the authorities to help prevent future events.

This application comes from a dataset available in `spatstat` (Baddeley and Turner, 2005). It concerns location of fires in the province of New Brunswick, Canada, from 1987 to 2003, with the year 1988 missing, comprising a total of 16 years of data. The analysis below uses the same dataset from Gonçalves and Gamerman (2018), shown in Figure 9, over a large rectangular area contained in the province. The model follows the description of Section 5 with a single latent process β_t undergoing a random walk $\beta_t = \beta_{t-1} + w_t$, with $w_t \sim GP$, $\forall t$ and $\beta_1 \sim GP$, with a slightly different configuration to that used in Gonçalves and Gamerman (2018).

Inference is performed via MCMC, sampling each β_t one at a time. This strategy is known to be less efficient than sampling β in a single block (Gamerman, 1998). But this strategy is far simpler to implement. Block sampling of β would require
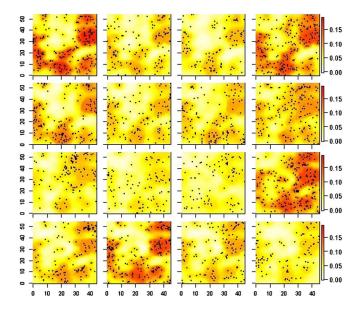
Figure 9 *Maps of the data* (*dots*) *and IF posterior mean in the New Brunswick fires example*: *years are ordered from the top row to the bottom row, and then from left to right in each row in the sequence* 1987, 1989, 1990, ..., 2003.
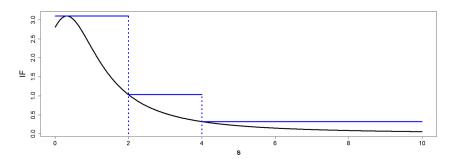

Figure 10 *Implementation of regional maxima* (*in blue*) *over the intensity of Figure* 2.

FFBS-like strategy or alternatively handling a joint multivariate skew normal distribution of much larger dimension. Also, the relative gains of block sampling are modest for such small time spans (16 time units in this case). Improvement over MCMC sampling in such dynamic settings is still subject of current research and report on other attractive alternatives is hoped to be released in the near future.

Results of the inference for the intensity function are also presented in Figure 10. It shows a nice balance between the information of the point pattern observed at each time with the estimated intensity function at neighboring times.

6 Discussion

Two approaches to handle inference for point processes were presented and discussed. Discretization is useful when regional effects are more useful than pointwise effects. The computational cost depends on number of regions specified. Augmentation also leads to discretization but the computational cost in this case depends on number of locations both real and augmented.

Either way, the computation depends crucially on GP's. These processes make use of multivariate normal distributions, that require matrix inversion. These operations may be extremely costly, specially if dimensions are large. This is a well-known problem in Geostatistics and a variety of approximation solutions has been proposed. Shirota and Banerjee (2018) applied some of these solutions to the models presented in this paper.

The need for an additional parameter λ^* in the augmentation approach opens up an additional possibility for introducing the effect of non-spatial covariates. Their effect was incorporated into the probit link but it may well affect the maximum. The latter option is not unreasonable. Intensities are affected by values of the explanatory variables and therefore it is expected that non-spatial covariates affect the maximum intensity. In this case, λ^* may depend on the individual configuration v, for example, via $\lambda^*(v) = \lambda_0^* h(v)$, for some positive transformation h.

The last section discussed the inefficiency of the augmentation route due to the use of the unique maximum intensity over the entire spatial domain. This approach could be made more efficient by varying λ^* over regions in space. Figure 10 illustrates the substantial gain of efficiency that could be obtained. The regional variation of λ^* in the example improved the acceptance of points from 20% to around 60%.

This variation of λ^* will induce jumps in the intensity function unless the GP assumption for β is replaced by another specification that would compensate the discontinuities. Such specification is far from trivial and it will likely introduce additional computational burden even if it were found.

The above point brings in a discussion of the very nature of the model. The GP assumption was a useful prior representation for smooth intensity functions. This smoothness is expected in many applications but may also be unreasonable for other situations. Pockets of violence in the illustration of Section 4.1 may cause an abrupt, substantial increase in the loss intensity. These hot spots provide a vivid example of discontinuities frequently encountered in applications. Also, the degree of smoothness may experience substantial variation across space. In these cases, changes in space of the GP characteristics could be beneficial.

This remark calls for a modeling variation of a single GP set-up into more general frameworks. There is a long list of alternatives available in the Geostatistics literature. Many of them can be written as some form of mixtures of GPs but very few of them are dedicated to specific case of point patterns. Liang et al. (2014) is a notable exception.

A simple, promising generalization to nonstationarity is provided by partition into local GPs (Kim et al., 2005; Gramacy and Lee, 2008). The partition approach is particularly attractive from a computational perspective. It may be applied to the regression coefficient process β from both discretization and augmentation routes. In the latter route, these ideas can also be applied to λ^*. This remark connects with the comments above about varying λ^*, that were seemingly concerned only with computation. This approach induces loss of information by not borrowing information across the entire spatial domain but Kim et al. (2005) provided some empirical evidence that this loss may not greatly affect inference.

The above discussion is just a sample of the wealth of possibilities that are open for use in the context of point pattern analysis. They are the subject of current research and we hope to report on them shortly. We hope that our paper will foster further development to the growing literature in the area and will increase our knowledge about the possible reach of this type of analysis.

Acknowledgments

References

Adams, R. P., Murray, I. and Mackay, D. J. C. (2009). Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proc. 26th Int. Conf. Machine Learning* (L. Bottou and M. Littman, eds.) Montreal: Omnipress.

Baddeley, A. and Turner, R. (2005). Spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software* **12**, 1–42.

Benes, V., Boldak, K., Møller, J. and Waagepetersen, R. (2005). A case study on point process modelling in disease mapping. *Image Analysis and Stereology* **24**, 159–168.

Brown, P. E. (2015). Model-based Geostatistics the easy way. *Journal of Statistical Software* **63**.

Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society, Series B* **17**, 129–164. MR0092301

Dias, R., Ferreira, C. S. and Garcia, N. L. (2008). Penalized maximum likelihood estimation for a function of the intensity of a Poisson point process. *Statistical Inference for Stochastic Processes* **11**, 11–34. MR2357551

Diggle, P. J. and Brix, A. (2001). Spatio-temporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society, Series B* **63**, 823–841. MR1872069

Diggle, P. J., Guan, Y., Hart, A. C., Paize, F. and Stanton, M. (2010). Estimating individual-level risk in spatial epidemiology using spatially aggregated information on the population at risk. *Journal of the American Statistical Association* **105**, 1394–1402. MR2796558

Frühwirth-Schnatter, S. and Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika* **93**, 827–841. MR2285074

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika* **85**, 215–227. MR1627273

Gamerman, D. (2010). Dynamic spatial models including spatial time series. In *Handbook of Spatial Statistics* (E. A. E. Gelfand, P. Diggle, P. Guttorp and M. Fuentes, eds.) 437–448. Londres: CRC/Chapman & Hall. MR2730959

Gamerman, D. and Lopes, H. F. (2006). Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. 2a edição. Londres: Chapman and Hall. MR2260716

Gamerman, D., Moreira, A. R. B. and Rue, H. (2003). Space-varying regression models: Specifications and simulation. *Computational Statistics & Data Analysis* **42**, 513–533. MR2005406

Gamerman, D., Santos, T. R. and Franco, G. C. (2013). A non-Gaussian family of state-space models with exact marginal likelihood. *Journal of Time Series Analysis* **34**, 625–645. MR3127211

Gelfand, A. E., Banerjee, S. and Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *EnvironMetrics* **16**, 465–479. MR2147537

Gonçalves, F. B. and Gamerman, D. (2018). Exact Bayesian inference in spatio-temporal Cox processes driven by multivariate Gaussian processes. *Journal of the Royal Statistical Society, Series B* **80**, 157–175. MR3744716

Gramacy, R. and Lee, H. K. H. (2008). Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association* **103**, 1119–1130. MR2528830

Kim, H.-M., Mallick, B. K. and Holmes, C. C. (2005). Analyzing nonstationary spatial data using piecewise Gaussian processes. *Journal of the American Statistical Association* **100**, 653–668. MR2160567

Lewis, P. A. W. and Shedler, G. S. (1979). Simulation of a nonhomogeneous Poisson process by thinning. *Naval Research Logistics Quarterly* **26**, 403–413. MR0546120

Liang, S., Carlin, B. and Gelfand, A. (2008). Analysis of Minnesota colon and rectum cancer point patterns with spatial and nonspatial covariate information. *Annals of Applied Statistics* **3**, 943–962. MR2750381

Liang, W. W. J., Colvin, J. B., Sansó, B. and Lee, H. K. H. (2014). Modeling and anomalous cluster detection for point processes using process convolutions. *Journal of Computational and Graphical Statistics* **23**, 129–150. MR3173764

Moller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics* **25**, 451–482. MR1650019

Pinto, J. A. Jr. (2014). Point processes with spatio-temporal heterogeneity. Unpublished Ph.D. thesis, Universidade Federal do Rio de Janeiro. Brazil. (in Portuguese).

Pinto, J. A. Jr., Gamerman, D., Paez, M. S. and Alves, R. H. F. (2015). Point pattern analysis with spatially varying covariate effects, applied to the study of cerebrovascular deaths. *Statistics in Medicine* **34**, 1214–1226. MR3322745

Reis, E. A., Gamerman, D., Paez, M. S. and Martins, T. G. (2013). Bayesian dynamic models for space–time point processes. *Computational Statistics & Data Analysis* **60**, 146–156. MR3007026

Shirota, S. and Banerjee, S. (2018). Scalable inference for space-time Gaussian Cox processes. preprint, ArXiV.

Taylor, B., Davies, T. M., Rowlingson, B. S. and Diggle, P. J. (2013). lgcp: An R Package for Inference with Spatial and Spatio-Temporal Log-Gaussian Cox Processes. *Journal of Statistical Software* **52**. MR3161587

Vivar, J. C. and Ferreira, M. A. R. (2009). Spatio-temporal models for Gaussian areal data. *Journal of Computational and Graphical Statistics* **18**, 658–674. MR2751645

Wikle, C. K. and Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika* **86**, 815–829. MR1741979

Instituto de Matemática
UFRJ
21941-909, Rio de Janeiro, RJ
Brazil
E-mail: dani@im.ufrj.br