

Multiplicative component models for replicated
point processes

Daniel Gervini
Department of Mathematical Sciences
University of Wisconsin–Milwaukee

May 26, 2017

Abstract

We propose a multiplicative semiparametric model for the intensity function of replicated point processes. Two examples of applications are given: a temporal one, about the dynamics of Internet auctions, and a spatial one, about the spatial distribution of street robberies in Chicago.

Key words: Doubly-stochastic process; functional data analysis; latent-variable model; Poisson process; spline smoothing.

1 Introduction

Point processes in time and space have a broad range of applications, in diverse areas such as neuroscience, ecology, finance, astronomy, seismology, and many others. Examples are given in classic textbooks like Cox and Isham (1980), Diggle (2013), Møller and Waagepetersen (2004), Streit (2010), and Snyder and Miller (1991), and in the papers cited below. However, the point-process literature has mostly focused on single-realization cases, such as the distribution of trees in a single forest (Jalilian et al., 2013) or the distribution of cells in a single tissue sample (Diggle et al., 2006). Situations where several replications of a process are available are increasingly common, but this area is still relatively unexplored in the literature. We can cite Diggle et al. (1991), Baddeley et al. (1993), Diggle et al. (2000), Bell and Grunwald (2004), Landau et al. (2004), Wager et al. (2004), and Pawlas (2011). However, these papers propose estimators for summary statistics of the processes rather than the intensity functions, which would be more informative.

When several replications of a process are available, it is possible to estimate the intensity functions by “borrowing strength” across replications. Along these lines Wu et al. (2013) propose estimators for the mean and principal components of independent and identically distributed realizations of a temporal doubly stochastic process based on kernel estimators of covariance functions. Gervini (2016) proposes an additive independent component model that has the advantages, over Wu et al., of treating the temporal and spatial cases in a unified way and of being easy to extend beyond the i.i.d. case, for instance, to regression and multivariate settings. In fact, Gervini and Baur (2017) is an extension of this method to marked point processes.

In this paper we propose an alternative to the additive model of Gervini (2016), namely an additive model for the log-intensity functions. This simplifies the numerical and theoretical aspects of the procedure by eliminating the nonnegativity constraints, but the interpretability is somewhat hampered by the fact that the additive model for the log-intensities translates into a multiplicative model for the intensities. At the end of this brief paper we present two examples of application, one temporal and one spatial, to illustrate these issues.

2 The model

A point process X is a random countable set in a space \mathcal{S} , where \mathcal{S} is usually \mathbb{R} for temporal processes and \mathbb{R}^2 or \mathbb{R}^3 for spatial processes (Møller and Waagepetersen, 2004, ch. 2; Streit, 2010, ch. 2). A process is locally finite if $\#(X \cap B) < \infty$ with probability one for any bounded $B \subseteq \mathcal{S}$. In that case we can define the count function $N(B) = \#(X \cap B)$ for any bounded $B \subseteq \mathcal{S}$, which essentially characterizes the process and is equivalent to X in this case.

Let X be locally finite and define $X_B = X \cap B$. Given a locally integrable function $\lambda : \mathcal{S} \rightarrow [0, \infty)$, i.e. a function λ such that $\int_B \lambda < \infty$ for any bounded $B \subseteq \mathcal{S}$, we say that X is a Poisson process with intensity function λ , denoted by $X \sim \mathcal{P}(\lambda)$, if (i) $N(B)$ follows a Poisson distribution with rate $\int_B \lambda$ and (ii) conditionally on $N(B) = m$, the m points in X_B are independent and identically distributed with density $\tilde{\lambda} = \lambda / \int_B \lambda$.

For $X \sim \mathcal{P}(\lambda)$, then, the density function of X_B at $x_B = \{t_1, \dots, t_m\}$ is

$$\begin{aligned} f(x_B) &= f(m)f(t_1, \dots, t_m|m) & (1) \\ &= \exp \left\{ - \int_B \lambda(t) dt \right\} \frac{\left\{ \int_B \lambda(t) dt \right\}^m}{m!} \times \prod_{j=1}^m \tilde{\lambda}(t_j) \\ &= \exp \left\{ - \int_B \lambda(t) dt \right\} \frac{1}{m!} \prod_{j=1}^m \lambda(t_j). \end{aligned}$$

What we mean by density of X_B , whose realizations are sets, not vectors, is the following: if \mathcal{N} is the family of locally finite subsets of \mathcal{S} , i.e. $\mathcal{N} = \{A \subseteq \mathcal{S} : \#(A \cap B) < \infty \text{ for all bounded } B \subseteq \mathcal{S}\}$, then for any $F \subseteq \mathcal{N}$,

$$\begin{aligned} P(X_B \in F) &= \sum_{m=0}^{\infty} \int_B \cdots \int_B \mathbb{I}(\{t_1, \dots, t_m\} \in F) f(\{t_1, \dots, t_m\}) dt_1 \cdots dt_m \\ &= \sum_{m=0}^{\infty} \frac{\exp \left\{ - \int_B \lambda(t) dt \right\}}{m!} \int_B \cdots \int_B \mathbb{I}(\{t_1, \dots, t_m\} \in F) \left\{ \prod_{j=1}^m \lambda(t_j) \right\} dt_1 \cdots dt_m, \end{aligned}$$

and, more generally, for any function $h : \mathcal{N} \rightarrow [0, \infty)$

$$E\{h(X_B)\} = \sum_{m=0}^{\infty} \int_B \cdots \int_B h(\{t_1, \dots, t_m\}) f(\{t_1, \dots, t_m\}) dt_1 \cdots dt_m. \quad (2)$$

A function h on \mathcal{N} is a function well defined on \mathcal{S}^m for any integer m and invariant under permutation of the coordinates; for example, $h(\{t_1, \dots, t_m\}) = \sum_{j=1}^m t_j/m$.

Single realizations of point processes are often modeled as Poisson processes with fixed λ s, but for replicated point processes a single intensity function λ rarely provides an adequate fit for all replications. It is more reasonable to assume that the λ s are subject-specific and treat them as latent random effects. Such processes are called doubly stochastic or Cox processes (Møller and Waagepetersen, 2004, ch. 5; Streit, 2010, ch. 8). A doubly stochastic process is a pair (X, Λ) where $X|\Lambda = \lambda \sim \mathcal{P}(\lambda)$ and Λ is a random function that takes values on the space \mathcal{F} of non-negative locally integrable functions on \mathcal{S} . The n replications of the process are then i.i.d. realizations $(X_1, \Lambda_1), \dots, (X_n, \Lambda_n)$ of (X, Λ) , where X is observable but Λ is not. In this paper we will assume that all X_i s are observed on a common region B of \mathcal{S} ; the method can be extended to X_i s observed on non-conformal regions B_i at the expense of higher computational complexity.

The latent intensity process Λ characterizes the distribution of X . Gervini (2016) proposes an additive model for Λ , but here we will explore the alternative approach of assuming an additive model for $\log \Lambda$, which is not constrained to be nonnegative. Let us assume, then, that

$$\log \Lambda(t) = \mu(t) + \sum_{k=1}^p U_k \phi_k(t) \quad (3)$$

where $\mu \in L^2(B)$ and ϕ_1, \dots, ϕ_p are orthonormal functions in $L^2(B)$. The U_k s are assumed independent $N(0, \sigma_k^2)$ random variables. Model (3), minus the Gaussianity assumption, is a truncated version of the Karhunen–Loève expansion (Ash and Gardner, 1975, ch. 1) that any process in $L^2(B)$ must follow, so it requires little justification. The Gaussianity assumption on the U_k s is added in order to derive maximum likelihood estimators; see next section. Model (3) translates into a multiplicative model for $\Lambda(t)$:

$$\Lambda(t) = \lambda_0(t) \prod_{k=1}^p \xi_k(t)^{U_k}, \quad (4)$$

where $\lambda_0 = \exp \mu$ is the baseline intensity function and $\xi_k = \exp \phi_k$ is a multiplicative component.

The mean and components of model (3) are functional parameters that need to

be estimated. We will follow a semiparametric approach, modeling μ and the ϕ_k s in terms basis functions β_1, \dots, β_q which can be, for example, B-splines for temporal processes or radial Gaussian kernels for spatial processes. Simplicial bases are another possibility for spatial processes, particularly if the domain B is irregular. In any case, we will have $\mu(t) = \mathbf{c}_0^T \boldsymbol{\beta}(t)$ and $\phi_k(t) = \mathbf{c}_k^T \boldsymbol{\beta}(t)$, where $\boldsymbol{\beta}$ is the vector of the β_k s. From (3) we can express

$$\log \Lambda(t) = (\mathbf{c}_0 + \mathbf{C}\mathbf{U})^T \boldsymbol{\beta}(t)$$

where $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_p]$ and $\mathbf{U} = (U_1, \dots, U_p)^T$. The parameters \mathbf{c}_0 and \mathbf{c}_k s, along with the variances σ_k^2 s of the U_k s, are estimated by penalized maximum likelihood, as explained next.

3 Estimation

Let us collect the parameters \mathbf{c}_0 , \mathbf{c}_k s and σ_k^2 s into a single vector $\boldsymbol{\theta}$. From now on we will omit the subindex B in x_B , since B is fixed. Then the marginal density of X_B at x is

$$\begin{aligned} f(x; \boldsymbol{\theta}) &= \int \int f(x, \mathbf{u}) d\mathbf{u} \\ &= \int \int f(x | \mathbf{u}) f(\mathbf{u}) d\mathbf{u} \end{aligned} \tag{5}$$

where, for $x = \{t_1, \dots, t_m\}$,

$$\begin{aligned} \log f(x | \mathbf{u}) &= - \int_B \lambda_{\mathbf{u}}(t) dt + \sum_{j=1}^m \log \lambda_{\mathbf{u}}(t_j) - \log m! \\ &= - \int_B \exp\{(\mathbf{c}_0 + \mathbf{C}\mathbf{u})^T \boldsymbol{\beta}(t)\} dt \\ &\quad + (\mathbf{c}_0 + \mathbf{C}\mathbf{u})^T \sum_{j=1}^m \boldsymbol{\beta}(t_j) - \log m! \end{aligned}$$

and

$$\log f(\mathbf{u}) = \sum_{k=1}^p \left(-\frac{1}{2} \log 2\pi\sigma_k^2 - \frac{u_k^2}{2\sigma_k^2} \right).$$

There is no closed form for $f(x; \boldsymbol{\theta})$ but it can be easily computed by Monte Carlo integration, as explained in the Technical Supplement.

The model parameters are estimated by penalized maximum likelihood. Since the dimension q of the functional basis $\boldsymbol{\beta}$ may be large, a roughness penalty is necessary to obtain smooth μ and ϕ_k s. We use penalties of the form $P(g) = \int_B \|\mathbf{H}g(t)\|_F^2 dt$, where \mathbf{H} denotes the Hessian and $\|\cdot\|_F$ the Frobenius matrix norm. Then for a temporal process $P(g) = \int (g'')^2$ and for a spatial process $P(g) = \int \{(\frac{\partial^2 g}{\partial t_1^2})^2 + 2(\frac{\partial^2 g}{\partial t_1 \partial t_2})^2 + (\frac{\partial^2 g}{\partial t_2^2})^2\}$, both of which are quadratic in the basis coefficients when evaluated at μ and the ϕ_k s.

Then the penalized maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ based on n independent realizations x_1, \dots, x_n is

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \rho_n(\boldsymbol{\theta})$$

where

$$\rho_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \log f(x_i; \boldsymbol{\theta}) - \nu_1 P(\mu) - \nu_2 \sum_{k=1}^p P(\phi_k)$$

and ν_1 and ν_2 are smoothing parameters. We use two different parameters for μ and the ϕ_k s because the latter have unit norm but μ does not, so it may be necessary to use ν_1 and ν_2 of different magnitudes to attain the same degree of smoothness. As mentioned before, $P(\mu) = \mathbf{c}_0^T \boldsymbol{\Omega} \mathbf{c}_0$ and $P(\phi_k) = \mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k$ for a matrix $\boldsymbol{\Omega}$ that depends on $\boldsymbol{\beta}$ and is derived in the Technical Supplement.

The smoothing parameters and the number of components p can be chosen by cross-validation, by maximizing

$$\operatorname{CV}(\nu_1, \nu_2, p) = \sum_{i=1}^n \log f(x_i; \hat{\boldsymbol{\theta}}_{(-i)}), \quad (6)$$

where $\hat{\boldsymbol{\theta}}_{(-i)}$ is the estimator for the reduced sample obtained after deleting x_i .

4 Applications

4.1 Internet auction data

In this section we analyze eBay auction data for Palm M515 Personal Digital Assistants (PDA) on week-long auctions that took place between March and May of

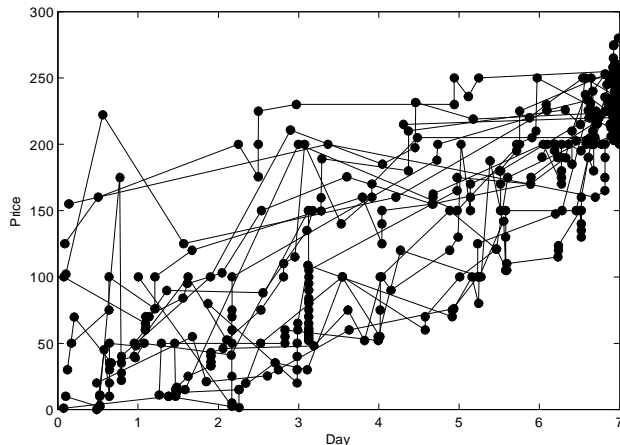


Figure 1: Internet Auction Data. Price trajectories of Palm Digital Assistants auctioned at eBay (first 20 trajectories in a sample of 194).

2003. The data was downloaded from the companion website of Jank and Shmueli (2010). There were 194 auctioned items in this sample; a subsample of 20 bid price trajectories are shown in Figure 1. The dots are the actual bids; the solid lines are for better visualization only. Individual trajectories are hard to follow in Figure 1, but some general trends are visible. For example, bidding activity seems to concentrate at the beginning and at the end of the auctions, in patterns that have been called “early bidding” and “bid sniping”, respectively. In this paper we are interested in the bidding times as a temporal point process, not on the bidding prices (the relationship between the two is explored in Gervini and Baur (2017) via additive models).

For these data we fitted a model (4) with $p = 2$ components, using cubic B-splines with 10 equally spaced knots as basis β . We found the smoothing parameters ν_1 and ν_2 by cross-validation, obtaining $\nu_1 = 10^{-4.5}$ and $\nu_2 = 10^{-2}$. We did not attempt to find an optimal p by cross-validation, since for illustrative purposes $p = 2$ suffices. The resulting baseline intensity function λ_0 and components ξ_1 and ξ_2 are shown in Figure 2. We see in Figure 2(a) that, as mentioned above, bidding generally intensifies towards the end of the auction period. The component ξ_1 , shown in Figure 2(b), is greater than one everywhere, so it is a size component: items with component scores $u_{i1} > 0$ will tend to have intensity functions λ_i that are overall larger than the baseline λ_0 , so they are items that attracted lots of bidders; whereas items with $u_{i1} < 0$ will tend to have λ_i s overall smaller than the baseline and therefore are

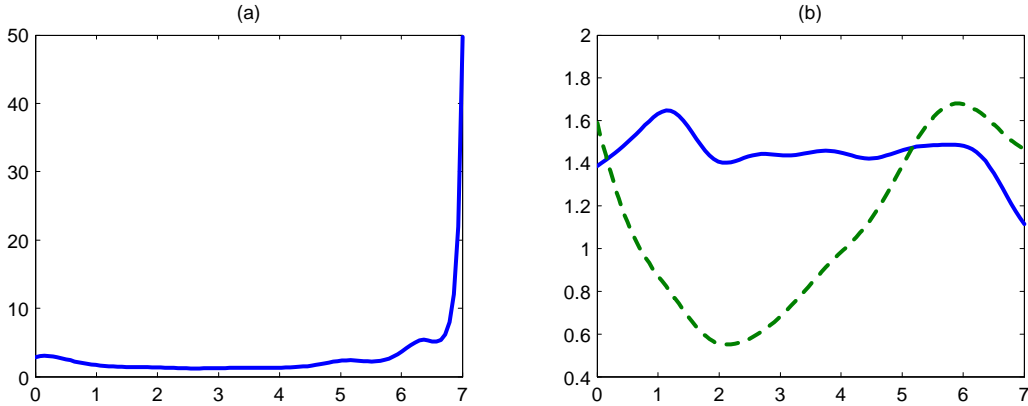


Figure 2: Internet Auction Data. (a) Baseline intensity function λ_0 . (b) Multiplicative components ξ_1 (solid line) and ξ_2 (dashed line).

items that attracted few bidders. This interpretation is in fact corroborated by the correlation between $\{u_{i1}\}$ and the number of bids per item, $\{m_i\}$, which is .88.

The second component, ξ_2 , is a contrast or shape component, because $\xi_2(t) > 1$ for $t < 1$ or $t > 4$, and $\xi_2(t) < 1$ for $1 < t < 4$, roughly. So, for an item i with $u_{i2} > 0$, the intensity λ_i will tend to be below the baseline for $t \in (1, 4)$ and above the baseline for $t \notin (1, 4)$. In particular, items subject to strong “bid snipping” will tend to have positive u_{i2} s while items that show more “early bidding” will tend to have negative u_{i2} s.

4.2 Street theft in Chicago

As a second example, this time of a spatial process, we analyzed the spatial distribution of street robberies in Chicago during the year 2014. The data was downloaded from the City of Chicago Data Portal, a very extensive data repository that provides, among other things, detailed information about every crime reported in the city. The information provided includes type, date, time, and coordinates (latitude and longitude) of the incident. Here we focus on crimes typified as of primary type “theft” and location “street”. There were 16,278 reported incidents of this type between January 1, 2014 and December 31, 2014. Their locations cover most of the city, as shown in Figure 3(a); a kernel-density estimator of these data is shown in Figure 3(b).

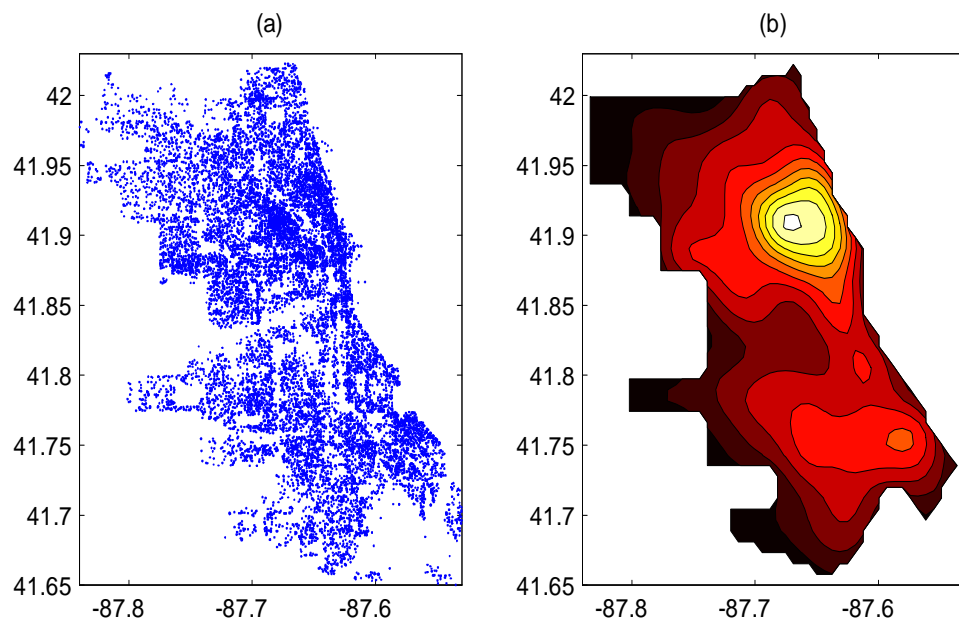


Figure 3: Chicago Street Theft. (a) Location of reported incidents in the year 2014. (b) Kernel density estimator of the data in (a).

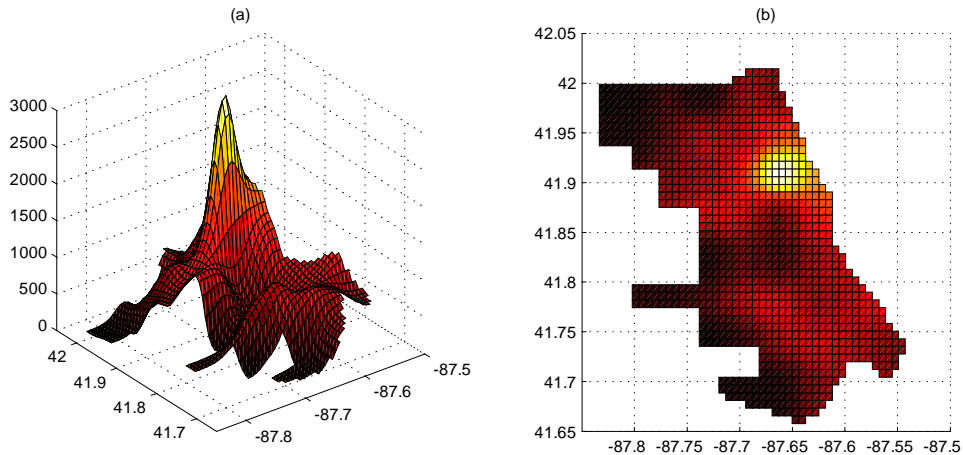


Figure 4: Chicago Street Theft. (a) Lateral view and (b) top view of baseline intensity function λ_0 .

We grouped up the data by day and considered them as $n = 365$ replications of a spatial point process, for which we fitted a multiplicative model (4). For illustrative purposes, we fitted a model with $p = 3$ components (we did not attempt to find an optimal p). As basis β we used renormalized Gaussian radial kernels $\beta_k(\mathbf{t}) = \exp\{-\|\mathbf{t} - \tau_k\|^2 / 2\delta_k^2\} / \sum_{j=1}^q \exp\{-\|\mathbf{t} - \tau_j\|^2 / 2\delta_j^2\}$, where the τ_k s were initially 100 uniformly spaced points in $[-87.84, -87.53] \times [41.65, 42.03]$, the smallest rectangle that includes the domain B (the city of Chicago), but those τ_k s outside B were eliminated, leaving $q = 40$ basis functions. The parameter δ_k was taken as half the distance between τ_k and the closest τ_j . The optimal smoothing parameters were obtained by cross-validation, $\nu_1 = 10^{-6.5}$ and $\nu_2 = 10^{-6}$.

The baseline intensity λ_0 is shown in Figure 4 and essentially coincides with the kernel smoother of the aggregated data (Figure 3(b)), as is to be expected. The mode of λ_0 occurs at Pulaski and Wicker Park, which are generally safe and affluent neighborhoods, but this is precisely what attracts street thieves; the poorer, crime-riddled neighborhoods of the West and South sides of the city are less populated and have less foot traffic, so street theft is actually rarer there.

The multiplicative components ξ_1 , ξ_2 and ξ_3 are shown in Figures 5, 6 and 7, respectively. The corresponding components of the log-intensity, ϕ_1 , ϕ_2 and ϕ_3 , are shown in Figure 8. The latter are sometimes easier to interpret due to their scale. For instance, we clearly see that ϕ_1 is nonnegative everywhere, whereas it is not

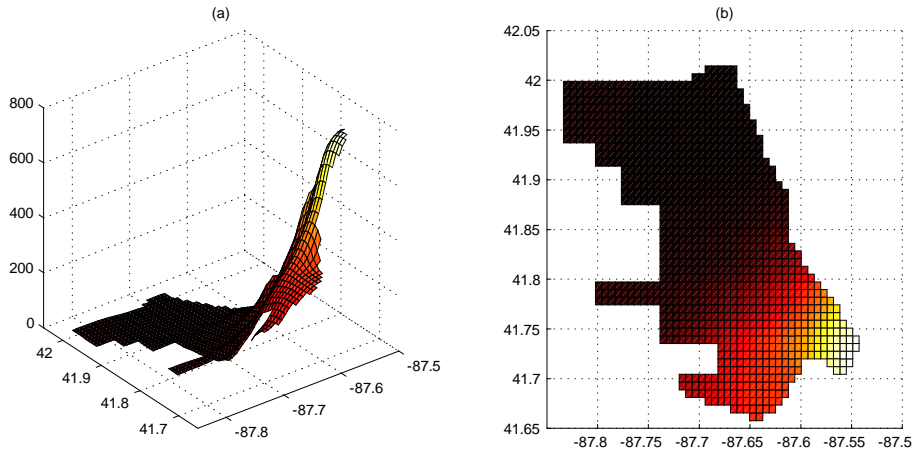


Figure 5: Chicago Street Theft. (a) Lateral view and (b) top view of first multiplicative component, ξ_1 .

easy to determine from Figure 5 if ξ_1 is greater than one everywhere or not. It also helps interpretation to plot the baseline intensity λ_0 versus $\lambda_+ = \exp(\mu + 2\sigma_k\phi_k)$ and $\lambda_- = \exp(\mu - 2\sigma_k\phi_k)$, since this shows the overall effect on λ of moving in the direction of the components. For the first component this is shown in Figure 9. This plot confirms that ξ_1 is a size component: λ will be greater than λ_0 everywhere for positive scores and smaller than λ_0 everywhere for negative scores, and the difference in amplitude will be more noticeable in the South-eastern part of the city, but not only in this part, as Figure 5 may seem to indicate. To further corroborate this interpretation, Figure 10 shows the incidents in the days with highest and lowest scores on the first component, which is in line with what has been said.

A similar analysis reveals that the second and third components are contrasts. For the second component, we see in Figure 11 that positive scores correspond to λ s that are above the baseline in the North-west part of the city and below the baseline in the South side, and the other way around for negative scores. The individual plots of the two extreme days (Figure 12) confirms this. For the third component, Figure 13 shows that positive scores correspond to λ s that are above the baseline in the narrow strip of affluent North-east neighborhoods by the lake and below the baseline everywhere else, and the other way around for negative scores. This is confirmed by the individual plots of the two extreme days (Figure 14).

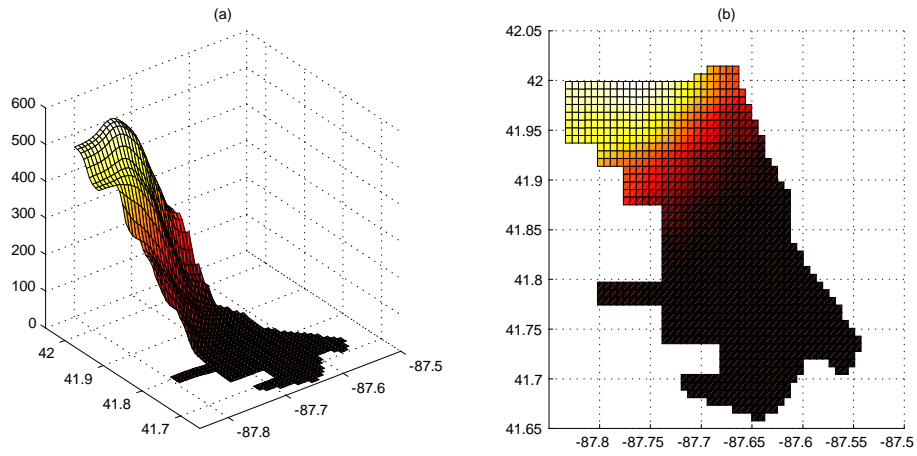


Figure 6: Chicago Street Theft. (a) Lateral view and (b) top view of second multiplicative component, ξ_2 .

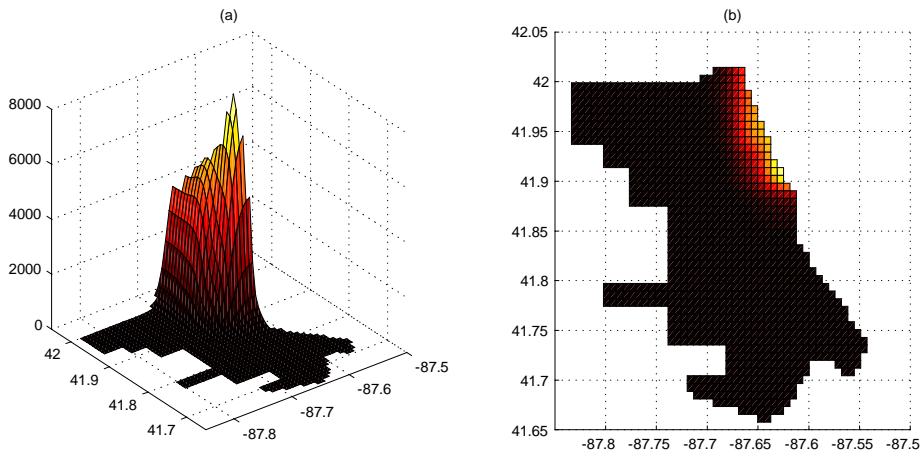


Figure 7: Chicago Street Theft. (a) Lateral view and (b) top view of third multiplicative component, ξ_3 .

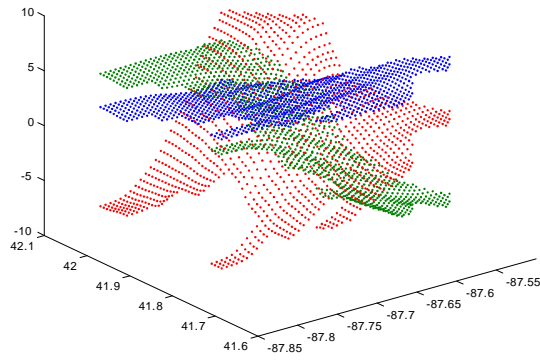


Figure 8: Chicago Street Theft. Log-intensity components ϕ_1 (blue), ϕ_2 (green) and ϕ_3 (red).

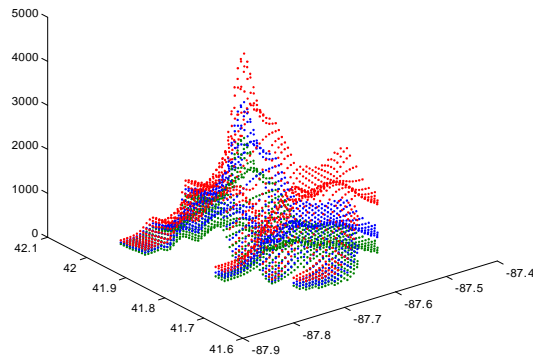


Figure 9: Chicago Street Theft. Baseline intensity function λ_0 (blue) versus λ_- (green) and λ_+ (red) for the first component.

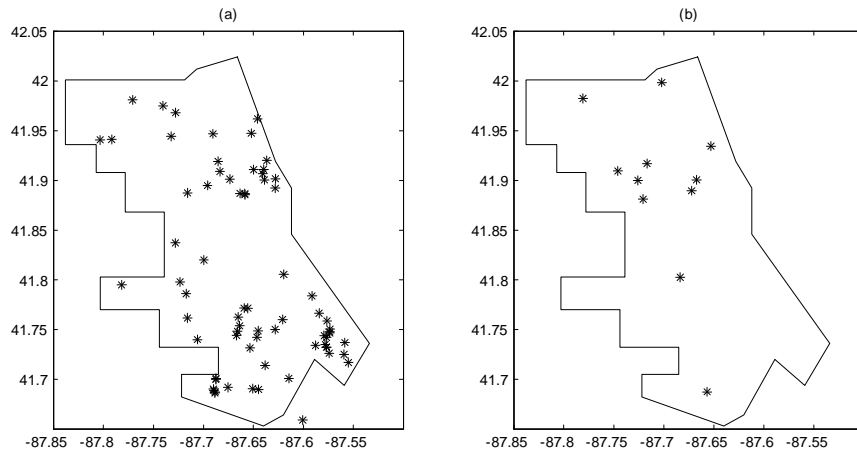


Figure 10: Chicago Street Theft. Days with highest [(a)] and lowest [(b)] scores on the first component.

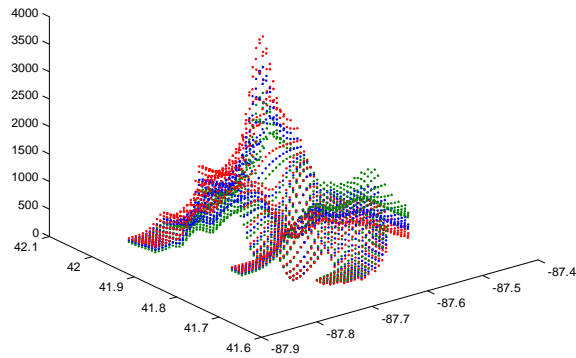


Figure 11: Chicago Street Theft. Baseline intensity function λ_0 (blue) versus λ_- (green) and λ_+ (red) for the second component.

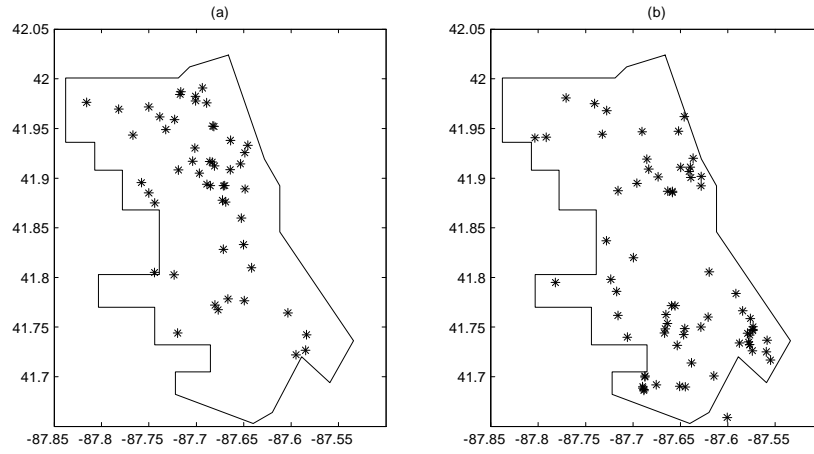


Figure 12: Chicago Street Theft. Days with highest [(a)] and lowest [(b)] scores on the second component.

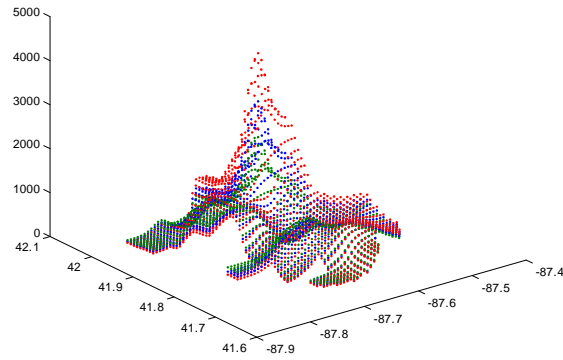


Figure 13: Chicago Street Theft. Baseline intensity function λ_0 (blue) versus λ_- (green) and λ_+ (red) for the third component.

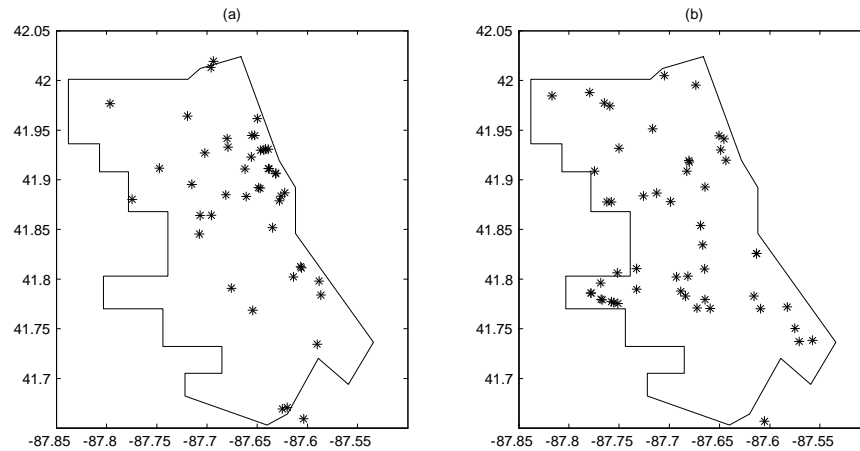


Figure 14: Chicago Street Theft. Days with highest [(a)] and lowest [(b)] scores on the third component.

Acknowledgement

This research was partly supported by US National Science Foundation grant DMS 1505780.

References

- Ash, R.B. and Gardner, M.F. (1975). *Topics in stochastic processes*. Academic Press, New York.
- Baddeley, A.J., Moyeed, R.A., Howard, C.V., and Boyde, A. (1993). Analysis of a three-dimensional point pattern with replication. *Applied Statistics* **42** 641–668.
- Bell, M.L., and Grunwald, G.K. (2004). Mixed models for the analysis of replicated spatial point patterns. *Biostatistics* **5** 633–648.
- Cox, D.R., and Isham, V. (1980). *Point Processes*. Chapman and Hall/CRC, Boca Raton.
- Diggle, P.J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition*. Chapman and Hall/CRC, Boca Raton.

- Diggle, P.J., Lange, N., and Beneš, F.M. (1991). Analysis of variance for replicated spatial point patterns in clinical neuroanatomy. *Journal of the American Statistical Association* **86** 618–625.
- Diggle, P.J., Mateau, J., and Clough, H.E. (2000). A comparison between parametric and nonparametric approaches to the analysis of replicated spatial point patterns. *Advances in Applied Probability* **32** 331–343.
- Diggle, P.J., Eglen, S.J., and Troy, J.B. (2006). Modeling the bivariate spatial distribution of amacrine cells. In *Case Studies in Spatial Point Process Modeling*, eds. A. Baddeley et al., New York: Springer, pp. 215–233.
- Gervini, D. (2016). Independent component models for replicated point processes. *Spatial Statistics* **18** 474–488.
- Gervini, D. and Baur, T.J. (2017). Regression models for replicated marked point processes. *ArXiv* 1705.06259.
- Jalilian, A., Guan, Y., and Waagepetersen, R. (2013). Decomposition of variance for spatial Cox processes. *Scandinavian Journal of Statistics* **40** 119–137.
- Jank, W., and Shmueli, G. (2010). *Modeling Online Auctions*. Wiley & Sons, New York.
- Landau, S., Rabe-Hesketh, S., and Everall, I.P. (2004). Nonparametric one-way analysis of variance of replicated bivariate spatial point patterns. *Biometrical Journal* **46** 19–34.
- Li, Y., and Guan, Y. (2014). Functional principal component analysis of spatiotemporal point processes with applications in disease surveillance. *Journal of the American Statistical Association* **109** 1205–1215.
- Møller, J., and Waagepetersen, R.P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.
- Pawlas, Z. (2011). Estimation of summary characteristics from replicated spatial point processes. *Kybernetika* **47** 880–892.
- Snyder, D.L., and Miller, M.I. (1991). *Random Point Processes in Time and Space*. Springer, New York.

- Streit, R.L. (2010). *Poisson Point Processes: Imaging, Tracking, and Sensing*. Springer, New York.
- Wager, C.G., Coull, B.A., and Lange, N. (2004). Modelling spatial intensity for replicated inhomogeneous point patterns in brain imaging. *Journal of the Royal Statistical Society Series B* **66** 429–446.
- Wu, S., Müller, H.-G., and Zhang, Z. (2013). Functional data analysis for point processes with rare events. *Statistica Sinica* **23** 1–23.