Supplementary material for "Regularised semi-parametric composite likelihood intensity modelling of a Swedish spatial ambulance call point pattern"

S1. Statistical methods

S1.1. Parametric spatial intensity function modelling

If X is an inhomogeneous Poisson point process, then the associated log-likelihood function is given by

$$\log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}) = \sum_{i=1}^{n} \log \rho_{\boldsymbol{\theta}}(x_i) - \int_W \rho_{\boldsymbol{\theta}}(x)\, dx. \tag{S1}$$

Using any quadrature rule, the integral in (S1) can be approximated by a finite sum

$$\int_W \rho_{\boldsymbol{\theta}}(x)\, dx \approx \sum_{j=1}^{m} \rho_{\boldsymbol{\theta}}(s_j) w_j, \tag{S2}$$

where the positive numbers w_j, $j = 1, 2, \ldots, m$, are quadrature weights summing to the area $|W|$ and $s_j \in W$, $j = 1, 2, \ldots, m$, are quadrature points. Following this approximation, the approximated log-likelihood function may be expressed as

$$\log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}) \approx \sum_{i=1}^{n} \log \rho_{\boldsymbol{\theta}}(x_i) - \sum_{j=1}^{m} \rho_{\boldsymbol{\theta}}(s_j) w_j, \tag{S3}$$

where $\mathbf{s} = \{s_1, \ldots, s_m\} \subseteq W$ represents the union of the observed spatial locations of events $\mathbf{x} = \{x_1, \ldots, x_n\}$ and the set of dummy points $\mathbf{s} \setminus \mathbf{x} = \{v_1, \ldots, v_q\}$, $q = m - n$. Here we assume that m is much larger than n for a better approximation of the log-likelihood function $\log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x})$. The

approximated log-likelihood function in expression (S3) can then be rewritten as

$$\log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}) \approx \sum_{j=1}^{m} \left(y_j \log \rho_{\boldsymbol{\theta}}(s_j) - \rho_{\boldsymbol{\theta}}(s_j) \right) w_j, \tag{S4}$$

$$y_j = \frac{a_j}{w_j} \quad \text{and} \quad a_j = \mathbf{1}\{s_j \in \mathbf{x}\}. \tag{S5}$$

Note that $a_j = 0$ means that s_j is a dummy point. Exploiting the approximation in equation (S2), a large number of dummy points are required to obtain accurate parameter estimation using equation (S4). Waagepetersen (2008) proposed two ways of obtaining dummy points and quadrature weights. With regard to the quadrature weights, the first method is a grid approach in which the observation window W is partitioned into a collection of rectangular tiles. The quadrature weight for a quadrature point $s \in \mathbf{s}$ falling in a tile R is the area of R divided by the total number of quadrature points falling in R. This approach is advantageous since the computation of the quadrature weights is easy. The second approach is the Dirichlet approach (Okabe et al., 2009) in which the quadrature weights are the areas of the tiles of the Dirichlet/Voronoi tessellation generated by the quadrature points in $\mathbf{s}$. With regard to the dummy points, Waagepetersen (2008)'s first approach is to use stratified dummy points combined with grid-type weights while the second approach is to exploit binomial dummy points with the Dirichlet-type weights. According to Baddeley and Turner (2000) and Thurman et al. (2015), a computationally cheaper approach to generate dummy points and compute quadrature weights is to partition the study region W into tiles R of equal area. To generate dummy points, we place one dummy point exactly in each tile either systematically or randomly. It follows that the quadrature weights for quadrature points s_j can be set to $w_j = \Delta / E_j$, where Δ is the area of each tile and E_j is the number of events and dummy points in the same tile as point s_j.

Modelling spatial intensity functions parametrically, in particular modelling based on spatial covariates, we often assume that $\rho_{\boldsymbol{\theta}}$ has a log-linear form. More specifically, letting $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_K)$

and $\boldsymbol{\theta} = (\beta_0, \boldsymbol{\beta})$, we assume that

$$\rho_\theta(x) = \exp\{\beta_0 + \mathbf{z}(x)\boldsymbol{\beta}'\}, \tag{S6}$$

where $\mathbf{z}(x) = (z_1(x), z_2(x), \cdots, z_K(x))$ is a vector of spatial covariates at location $x \in W$. Combining equations (S4) and (S6), we will get the approximated log-likelihood function $\log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x})$ given in equation (1) in the article.

S2. Optimisation methods

To deal with the optimisation problem in (4), we carry out the optimisation using a cyclical coordinate descent method, which optimises a target function/optimisation problem with respect to a single parameter at a time and iteratively cycles through all parameters until a convergence criterion is reached. Here, we present the coordinate descent algorithm for solving the regularized log-likelihood function with the elastic-net penalty.

Let $f(\boldsymbol{\theta})$ be the approximated log-likelihood function (1), i.e.,

$$f(\boldsymbol{\theta}) = f(\beta_0, \boldsymbol{\beta}) = \sum_{j=1}^{m} w_j \left(y_j \left(\beta_0 + \mathbf{z}(s_j)\boldsymbol{\beta}' \right) - \exp\left\{ \beta_0 + \mathbf{z}(s_j)\boldsymbol{\beta}' \right\} \right).$$

Let r denote the step number in the optimisation algorithm, and $\boldsymbol{\theta}^{(r-1)} = (\beta_0^{(r-1)}, \boldsymbol{\beta}^{(r-1)})$ represent the current estimates of the parameters. A quadratic approximation of $f(\boldsymbol{\theta})$ at the point $\boldsymbol{\theta}^{(r-1)}$ is given by

$$f(\boldsymbol{\theta}) \approx f\left(\boldsymbol{\theta}^{(r-1)}\right) + \frac{df\left(\boldsymbol{\theta}^{(r-1)}\right)}{d\boldsymbol{\theta}} \left(\boldsymbol{\theta} - \boldsymbol{\theta}^{(r-1)}\right)' + \frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}^{(r-1)}\right) \frac{d^2 f\left(\boldsymbol{\theta}^{(r-1)}\right)}{d\boldsymbol{\theta} d\boldsymbol{\theta}'} \left(\boldsymbol{\theta} - \boldsymbol{\theta}^{(r-1)}\right)', \tag{S7}$$

where the first and the second-order derivatives of the function f with respect to $\boldsymbol{\theta}$ are given by

$$\frac{df(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = \sum_{j=1}^{m} w_j \left(y_j \bar{\mathbf{z}}(s_j) - \exp\left\{ \beta_0 + \mathbf{z}(s_j)\boldsymbol{\beta}' \right\} \bar{\mathbf{z}}(s_j) \right), \tag{S8}$$

$$\frac{d^2 f(\boldsymbol{\theta})}{d\boldsymbol{\theta} d\boldsymbol{\theta}'} = -\sum_{j=1}^{m} w_j \exp\left\{ \beta_0 + \mathbf{z}(s_j)\boldsymbol{\beta}' \right\} \bar{\mathbf{z}}(s_j)' \bar{\mathbf{z}}(s_j), \tag{S9}$$

and $\bar{\mathbf{z}}(s_j) = (1, \mathbf{z}(s_j))$. Hence, a quadratic approximation of the approximated log-likelihood function in (1) can be obtained through equations (S8), (S9), and (S7), i.e.,

$$\mathcal{L}_q(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{j=1}^{m} u_j \left(y_j^* - \beta_0 - \mathbf{z}(s_j)\boldsymbol{\beta}' \right)^2 + C\left(\boldsymbol{\theta}^{(r-1)} \right), \tag{S10}$$

where $C(\boldsymbol{\theta}^{(r-1)})$ is a constant function of $\boldsymbol{\theta}$ and the remaining variables in equation (S10) are given by

$$y_j^* = \beta_0^{(r-1)} + \mathbf{z}(s_j)\boldsymbol{\beta}^{(r-1)\prime} + \frac{y_j}{\exp\left\{ \beta_0^{(r-1)} + \mathbf{z}(s_j)\boldsymbol{\beta}^{(r-1)\prime} \right\}} - 1 \quad \text{and} \quad u_j = w_j \exp\left\{ \beta_0^{(r-1)} + \mathbf{z}(s_j)\boldsymbol{\beta}^{(r-1)\prime} \right\} \tag{S11}$$

As can be seen from equations (S10) and (S11), the variable y_j^* is the working response variable while u_j is the updated weight. Replacing the log-likelihood function $\log \mathcal{L}(\boldsymbol{\theta})$ in equation (4) by the quadratic approximation $\mathcal{L}_q(\boldsymbol{\theta})$, the optimisation problem of the regularised quadratic approximation of the log-likelihood function becomes

$$\underset{\boldsymbol{\theta} \in \mathbb{R}^{K+1}}{\operatorname{argmin}} \mathcal{L}_{qp}(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta} \in \mathbb{R}^{K+1}}{\operatorname{argmin}} \left\{ -\frac{1}{m} \mathcal{L}_q(\boldsymbol{\theta}) + \lambda \sum_{k=1}^{K} \left\{ \frac{1}{2}(1-\alpha)\beta_k^2 + \alpha|\beta_k| \right\} \right\}. \tag{S12}$$

The optimisation problem in equation (S12) can be solved by the coordinate descent algorithm. More specifically, for any pre-specified value of the tuning parameter λ, in iteration $r = 1, 2, \ldots$, the coordinate descent algorithm partially optimises the optimisation problem with respect to β_k, given the estimates $\beta_0^{(r-1)}$ and $\beta_h^{(r-1)}$, $h \in \{1, \ldots, K\} \setminus \{k\}$. Explicitly, the optimisation can be described by

$$\underset{\boldsymbol{\theta} \in \mathbb{R}^{K+1}}{\operatorname{argmin}} \mathcal{L}_{qp}(\boldsymbol{\theta}) \approx \underset{\beta_k \in \mathbb{R}}{\operatorname{argmin}} \mathcal{L}_{qp}\left(\beta_0^{(r-1)}, \beta_1^{(r-1)}, \cdots, \beta_{k-1}^{(r-1)}, \beta_k, \beta_{k+1}^{(r-1)}, \cdots, \beta_K^{(r-1)} \right). \tag{S13}$$

According to Friedman et al. (2007), there are closed form coordinate-wise updates to estimate the parameters of the optimisation problem. Letting $\beta_k \geq 0$, the first-order derivative of $\mathcal{L}_{qp}(\boldsymbol{\theta})$ in equation (S13) with respect to β_k is given by

$$\frac{d\mathcal{L}_{qp}(\boldsymbol{\theta})}{d\beta_k} = -\frac{1}{m}\sum_{j=1}^{m} u_j z_k(s_j)\left(y_j^* - \tilde{y}_j^{(k)}\right) + \frac{1}{m}\sum_{j=1}^{m} u_j z_k^2(s_j)\beta_k + \lambda(1-\alpha)\beta_k + \lambda\alpha, \tag{S14}$$

where $\tilde{y}_j^{(k)} = \beta_0^{(r-1)} + \sum_{h\neq k}^{K} \beta_h^{(r-1)} z_h(s_j)$ is the fitted value excluding the covariate $z_k(s_j)$. Similarly, the first-order derivative of $\mathcal{L}_{qp}(\boldsymbol{\theta})$ for the case $\beta_k < 0$ can easily be obtained. It follows that the coordinate-wise updates for parameter estimation in the elastic-net penalisation can be obtained by

$$\beta_k^{(r)} = \frac{S\left(\frac{1}{m}\sum_{j=1}^{m} u_j z_k(s_j)\left(y_j^* - \tilde{y}_j^{(k)}\right), \lambda\alpha\right)}{\frac{1}{m}\sum_{j=1}^{m} u_j z_k^2(s_j) + \lambda(1-\alpha)}, \quad r = 1, 2, \ldots, \quad k = 1, 2, \ldots, K, \tag{S15}$$

where $S(z, \vartheta)$ is the soft-thresholding operator given by

$$S(z,\vartheta) = \text{sign}(z)(|z| - \vartheta)_+ = \begin{cases} z - \vartheta, & \text{if } z > 0 \text{ and } \vartheta < |z|, \\ 0, & \text{if } \vartheta \geq |z|, \\ z + \vartheta, & \text{if } z < 0 \text{ and } \vartheta < |z|. \end{cases}$$

The intercept parameter need not be penalised as it has no role in the variable selection. The estimate of the intercept term can be obtained by

$$\beta_0^{(r)} = \frac{1}{\sum_{j=1}^{m} u_j}\sum_{j=1}^{m} u_j\left(y_j^* - \mathbf{z}(s_j)\boldsymbol{\beta}'^{(r-1)}\right), \quad r = 1, 2, \ldots.$$

The parameter estimates are updated until the algorithm converges. With regard to the tuning parameter $\lambda \in [\lambda_{max}, \lambda_{min}]$, we start with the smallest value λ_{max} of the tuning parameter for which the entire vector is zero. That is, we begin with a λ_{max} for which $\widehat{\boldsymbol{\beta}} = 0$ to obtain solutions for a

decreasing sequence of λ values. Using a prediction performance measure, e.g. cross-validation, for the estimated model, the user can select a particular value of λ from the candidate sequence of λ values. Since the parameter estimation updating equation (S15) is obtained for elastic-net penalisation, we may set $\alpha = 0$ to implement ridge regression and $\alpha = 1$ to use the lasso approach; other kinds of elastic-net regularisation can be implemented by picking $\alpha \in (0, 1)$. Recall that elastic-net is particularly useful when there are many correlated covariates in the statistical model and the data are high-dimensional, i.e., the data have the property that $K \gg m$. Cyclical coordinate descent methods are a natural approach to solving convex problems and each coordinate-descent step of the algorithm is fast with an explicit formula for each coordinate-wise optimisation. It also exploits the sparsity of the model and it has better computational speed both for high dimensional data and big data (Friedman et al., 2010a).

Algorithm 1. Algorithm for regularized Poisson process log-likelihood estimation.

1: Identify the spatial domain W,
2: Generate a set of dummy points $\mathbf{v} = \{v_1, v_2, \ldots, v_q\}$ in W,
3: Combine the dummy points $\mathbf{v} = \{v_1, v_2, \ldots, v_q\}$ with the data points $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ to form a set of quadrature points $\mathbf{s} = \{s_1, s_2, \ldots, s_m\}$,
4: Compute the quadrature weights w_j, $j = 1, \ldots, m$,
5: Following equation (S5), determine the indicator a_j and compute the variable $y_j = a_j/w_j$, $j = 1, \ldots, m$,
6: Obtain the vector of spatial covariates $\mathbf{z}(s_j) = (z_1(s_j), \ldots, z_K(s_j))$ at each quadrature point s_j, $j = 1, \ldots, m$,
7: Use existing model-fitting software such as `glmnet` (Friedman et al., 2010b), specifying that the model is a log-linear Poisson regression model, $\log \rho_{\boldsymbol{\theta}}(s_j) = \beta_0 + \mathbf{z}(s_j)\boldsymbol{\beta}'$, in order to fit the responses y_j and vector of covariate values $\mathbf{z}(s_j)$ with weights w_j, $j = 1, \ldots, m$,
8: The coefficient estimates returned by the software give the approximate maximum log-likelihood estimate of $\boldsymbol{\theta}$,

The optimisation problem in equation (S13) can be implemented using Algorithm 1. The approximated log-likelihood function in equation (1) and the log-likelihood function of the weighted generalised linear model (Poisson distribution) have the same deviance function $\mathcal{D}(\boldsymbol{\theta}) = 2\{\log \mathcal{L}(\mathbf{y} \mid \mathbf{y}) - \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})\}$. Hereby, the deviance $\mathcal{D}$ of the regularised weighted generalised linear model (Poisson distribution) obtained by the model-fitting software, e.g. `glmnet`, can be exploited to select an optimal tuning parameter in the optimisation of the regularised quadratic approximation of

the log-likelihood function of the inhomogeneous Poisson process in equation (S12). Choosing an optimal tuning parameter value can be done using K-fold cross-validation where, in short, we set a sequence of tuning parameter candidate values $\lambda_1 > \lambda_2 > \cdots > \lambda_T$ and split the data into K-folds. Then, for each λ-candidate value, we leave out a data fold/piece and perform parameter estimation on all the remaining $K-1$ data folds, thus obtaining a deviance for the left-out data fold. We repeat the parameter estimation and deviance computation for the remaining $K-1$ possible folds to be left out. This means that we obtain K out-of-sample deviances for each λ value. Among the sequenced λ values, the one giving the smallest mean deviance can be picked as an optimal estimate of the tuning parameter λ of the regularised quadratic approximation of the log-likelihood function. We then use the selected optimal λ to again carry out the regularized fitting, this time using the full dataset, in order to obtain a final estimate of the model parameter $\boldsymbol{\theta}$. Finally, the stopping criterion for the cyclic coordinate descent algorithm is generally based on the change of the fitted quadratic approximation of the log-likelihood function value.

S3. Semi-parametric intensity function modelling

The chosen modelling approach is suitable for the setting where i) the main goal is a prediction/predictive model, i.e., one wants to predict a collection of further/unseen events as precisely as possible, ii) one believes that the observed covariates can only describe a part of the spatial intensity variation, and iii) added spatial flexibility is warranted in the modelling. In the case of our ambulance data, as our end goal is to build optimal dispatching strategies, we mainly want a predictive model.

We want to emphasise that the demographic covariates we have at hand only reflect where different demographic groups live but not how they move about (recall Section 3.6). Specifically, we do not have access to explicit mobility covariates such as aggregated movement patterns in the population, and, as one may guess, people do not only need access to ambulances when they are at home. Our solution is quite pragmatic and simple. We simply add a further spatial covariate to the existing collection of covariates, which is given by a non-parametric spatial intensity estimate $\widetilde{\rho}(x)$, $x \in W$,

and will be referred to as the *benchmark spatial intensity*. As an alternative one could instead use $\log\widetilde{\rho}(x)$, $x \in W$, as a further spatial covariate, resulting in $\widetilde{\rho}$ being treated as an offset in (S6); note that this would require that $\widetilde{\rho}$ is strictly positive. Hence, we include the spatial intensity estimate $\widetilde{\rho}$ as a covariate in the approximated log-likelihood expression in (1) and the inclusion has the effect that the modelling steps away from a purely parametric setting to a semi-parametric approach. This added covariate should pick up on regions where there is an increased intensity due to human mobility, which cannot be explained by the existing list of covariates. Note that this is similar in nature to the semi-parametric (spatio-temporal) log-Gaussian Cox process modelling approach advocated for in e.g. Diggle (2013). To be able to discern whether this added covariate is in fact necessary/useful in the presence of the other covariates, we carry out elastic-net regularisation-based variable selection (see Section 3.5) to indicate whether the benchmark spatial intensity has any added value in terms of describing the true intensity function.

A natural question here is what kind of non-parametric intensity estimator one should use to generate the benchmark spatial intensity. There are different candidates for this, and the main distinction one usually makes is between global and adaptive/local smoothers. Adaptive smoothing techniques include adaptive kernel intensity estimation (Davies et al., 2018) and (resample-smoothed) Voronoi intensity estimation (Moradi et al., 2019; Ogata, 2004; Ogata et al., 2003). These have some clear benefits (in particular the latter, Moradi et al. (2019)), but here we do not want to put too much weight on the local features since we may run the risk of overfitting. Instead, we here consider (global) kernel intensity estimation (Baddeley et al., 2015; Diggle, 1985), which is arguably the most prominent approach to global smoothing and is defined as

$$\widetilde{\rho}(x) = \sum_{y \in \mathbf{x}} \kappa_h(x - y)/w_h(x, y), \quad x \in W,$$

where $\kappa_h(\cdot) = h^{-1}\kappa(\cdot/h)$, κ is a symmetric density function and the smoothing parameter $h > 0$ is the bandwidth. The function $w_h(x, y)$ is a suitable edge correction factor which adjusts for the effect of unobserved events outside W on the intensity of the observed events (Baddeley et al., 2015); we

here use the local corrector $w_h(x,y) = \int_W \kappa_h(u-y)du$ which ensures mass preservation, i.e. that $\int_W \widetilde{\rho}(x)dx = n$, the number of observed events. Practically, to carry out kernel intensity estimation we make use of the function `density.ppp` in the R package spatstat (Baddeley et al., 2015).

Although the choice of kernel may play a certain role, the choice of bandwidth is absolutely the main determinant for the quality of the intensity estimate; recall that the bandwidth governs how much we smooth the data. However, optimal bandwidth selection is a well-studied and challenging problem. Concerning the state of the art, the bandwidth selection criterion of Cronie and Van Lieshout (2018) is generally the most stable with respect to accounting for spatial interaction; observed clusters of points in a point pattern may be the effect of aggregation/clustering (dependence) or intensity peaks, or a combination of the two. However, there is one scenario where it tends to not perform too well, and that is when there are large regions in W where there are no points present, which is the case for our ambulance dataset. Other standard methods for bandwidth selection include the state estimation approach of Diggle (1985) (called `bw.diggle` in spatstat), the Poisson process likelihood leave-one-out cross-validation approach in Baddeley et al. (2015) and Loader (1999) (called `bw.ppl` in spatstat), the rule of thumb of Scott (2015) (called `bw.scott` in spatstat), and the recent machine learning-based approach of Bayisa et al. (2020) (see Algorithm 2).

S3.1. New heuristic algorithm for bandwidth selection

In K-means clustering, the dataset is partitioned into a number of clusters, and each cluster consists of data points whose intra-point distances, i.e., distances between points within a cluster, are smaller than their inter-point distances, i.e., their distances to points outside of the cluster. In a recent study, Bayisa et al. (2020) proposed a K-means clustering-based bandwidth selection approach for kernel intensity estimation, where the average of the standard deviations of the clusters is used as an optimally selected bandwidth. Although the approach performed well in terms of non-parametrically describing the current ambulance call dataset, it has some limitations/issues. Firstly, the number of clusters used in the K-means algorithm has been selected visually, and thereby subjectively. Secondly, clusters with high point densities and clusters with widely dispersed data points

tend to uniformly determine the resulting bandwidth. Evidently, a cluster with widely dispersed data points has a larger standard deviation, which can be an outlier, and hence, it distances the estimated bandwidth away from the standard deviations of clusters with highly clustered data points. As a result, the selected bandwidth leads to oversmoothing of the spatial intensity.

Algorithm 3. Subalgorithm of the main algorithm

1: **while** $\Delta > \varepsilon$ and $v \leq max.iter$ **do**

2: Obtain optimal classes k_i and 1-of-P class indicator variables ι_{iq} for data points x_i:

$$k_i^{(v)} = \underset{q\in\{1,2,\ldots,P\}}{\arg\min} \left\|x_i - \varpi_q^{(v)}\right\|^2, \quad \iota_{iq}^{(v)} = \mathbf{1}_{\left\{q=k_i^{(v)}\right\}}, \quad q = 1, 2, \ldots, P; i = 1, 2, \ldots, n,$$

3: Update the centroids of the clusters (the mean locations of the clusters) ϖ_q:

$$\varpi_q^{(v+1)} = \left\{\sum_{i=1}^{n} \iota_{iq}^{(v)} x_i\right\} \Big/ \sum_{i=1}^{n} \iota_{iq}^{(v)}, \quad q = 1, 2, \ldots, P,$$

4: $\Delta = \sum_{q=1}^{P} \left\|\varpi_q^{(v+1)} - \varpi_q^{(v)}\right\|^2$.

5: **end while**

To overcome these limitations, we propose a new heuristic algorithm, which is outlined in Algorithm 2, to establish the ideal number of clusters and thereby to obtain an optimal estimate of the bandwidth. The main algorithm, which is Algorithm 2, consists of two crucial steps. It continually invokes Algorithm 3, which is a K-means algorithm, to establish the optimal number of clusters using the KL index of Krzanowski and Lai (1988). Once an optimal number of clusters has been obtained, the main algorithm determines an optimal bandwidth by calling the K-means algorithm and using a weighted mean of dispersion measures for the clusters, where the weight for each cluster is given by the inverse of the average of the squares of the distances from the centroid of the cluster to each observation. When establishing the bandwidth, the weights help in balancing the contributions of clusters with closely spaced spatial points and clusters with widely spaced spatial points. In Algorithm 3, the spatial data points are re-assigned to clusters (see step 2 in Algorithm 3), the cluster means are re-computed (see step 3 in Algorithm 3), and these steps are repeated until the sum composed of the squared Euclidean distances between all successive centroids is smaller than

Algorithm 2. K–means clustering-based heuristic algorithm for bandwidth selection

1: Consider the observed spatial location data: $\mathbf{x} = \{x_1, x_2, \cdots, x_n\} \subseteq W \subset \mathbb{R}^2$,
2: Candidates for the number of clusters: $K = 2, 3, \cdots, K_{max}$,
3: Let $\mathrm{tr}(\cdot)$ denote the trace of a square matrix,
4: Let d denote the number of variables,
5: Set while condition determining parameter: $\Delta = 1$,
6: Set the maximum number of iterations for the while loop: $max.iter$,
7: Let ϖ_q, $q = 1, 2, \cdots, P$ represent the centroid of a cluster q,
8: Let C_q denote the collection of observations in cluster q,
9: Let n_q is the number of data points in C_q,
10: Set a parameter determining the convergence of the algorithm: ε,
11: Let $D_P = \sum_{q=1}^{P} \sum_{x \in C_q} \left(x - \varpi_q\right)' \left(x - \varpi_q\right)$ denote within-cluster dispersion matrix for P clusters,
12: Let $KL_P = \left| \frac{(P-1)^{2/d} \mathrm{tr}(D_{P-1}) - P^{2/d} \mathrm{tr}(D_P)}{P^{2/d} \mathrm{tr}(D_P) - (P+1)^{2/d} \mathrm{tr}(D_{P+1})} \right|$ represent KL index,
13: **for** $K \leftarrow 2, 3, \cdots, K_{max}$ **do**
14: **for** $P \leftarrow K + 1$ **do**
Initialize the centroids: $\varpi_q^{(v)}$, $v = 0$ and $q = 1, 2, \ldots, P$,
Call algorithm 3,
15: **end for**
16: **if** $K = 2$, **then**
17: **for** $P \leftarrow K$ **do**
Initialize the centroids: $\varpi_q^{(v)}$, $v = 0$ and $q = 1, 2, \ldots, P$,
Call algorithm 3,
18: **end for**
19: **for** $P \leftarrow K - 1 = 1$ **do**
$\varpi_1 = \left(\sum_{i=1}^{n} x_i \Big/ n \right)$,
20: **end for**
21: **end if**
22: **if** $K > 2$, **then**
23: **for** $P \leftarrow K - 1, K$ **do**
Initialize the centroids: $\varpi_q^{(v)}$, $v = 0$ and $q = 1, 2, \ldots, P$,
Call algorithm 3,
24: **end for**
25: **end if**
26: Obtain the optimal centroids $\left\{\varpi_q\right\}_{q=1}^{P}$ for $P = K - 1$, K, and $K + 1$ from step 14 to 25,
27: Using the result from step 26, compute the expression D_P in step 11 for $P = K - 1$, K, and $K + 1$,
28: Based on the result from step 27, compute KL_P in step 12 for $P = K$.
29: **end for**

30: From step 13 to 29, obtain the optimal number of clusters: $P_0 = \underset{P\in\{2,3,\ldots,K_{max}\}}{\arg\max} \left\{KL_P\right\}$,

31: Initialize the centroids for the optimal number of clusters P_0: $\varpi_q^{(v)}$, $v = 0$ and $q = 1, 2, \ldots, P = P_0$,

32: Based on step 31, call algorithm 3 and obtain the optimal centroids $\left\{\varpi_q\right\}_{q=1}^{P}$ and classes of data points $\{k_i\}_{i=1}^{n}$,

33: Compute cluster dispersion measure: $\sigma_q^2 = \dfrac{0.5}{n_q - 1} \sum_{x_i \in C_q} \left\|x_i - \varpi_q\right\|^2$, $q = 1, 2, \ldots, P$,

34: Compute the weight: $w_q = \dfrac{1}{g_q\left(x, \varpi_q\right)}$, $q = 1, 2, \ldots, P$ and $g_q\left(x, \varpi_q\right) = \dfrac{1}{n} \sum_{i=1}^{n} \left\|x_i - \varpi_q\right\|^2$,

35: Obtain an optimal bandwidth estimate: $h = \sqrt{\dfrac{\sum_{q=1}^{P} w_q \sigma_q^2}{\sum_{q=1}^{P} w_q}}$.

a user-specified value (see step 4 in Algorithm 3). Alternatively, one may control the convergence of Algorithm 3 by repeating steps 2 and 3 until there is either no further change in the assignments of data points to clusters or until some maximum number of iterations has been reached. In our case, we have used $\varepsilon = 10^{-5}$ and $max.iter = 100$.

S4. Spatial covariates

This section presents a short description of the spatial covariates (some are graphically illustrated in Figure S1):

- **Population density:** Per DeSO, this gives us the ratio of the total population size of the DeSO to the area of the DeSO.

- **Population by age (counts):** The number of individuals of a given age category living in a given DeSO zone. There are a total of 17 such covariates, reflecting the following age categories (ages in years): 0-5, 6-9, 10-15, 16-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79 and 80+.

- **Population by sex (counts):** The number of individuals of a given sex classification living in a given DeSO zone. There are 2 sex related covariates: women and men.

- **Population by Swedish/foreign background (counts):** The number of individuals of a given Swedish/foreign classification living in a given DeSO zone. There are 2 such covariates: Swedish background and foreign background (the latter includes people born outside Sweden as well as people born in Sweden with both parents born outside Sweden).

- **Population 20-64 years of age by occupation (counts):** The number of individuals of a given occupational status living in a given DeSO zone. There are a total of 2 such covariates: (gainfully) employed and unemployed.

- **Population 25-64 years of age by education level (counts):** The number of individuals of a given maximal education level category living in a given DeSO zone. There are a total of 4 such covariates: no secondary education, secondary education, post-secondary education of at most three years, and post-secondary education of more than three years (including doctoral degrees).

- **Population 20+ years of age by accumulated income (counts):** The number of individuals of a given income level category living in a given DeSO zone. There are a total of 2 such covariates: below the national median income and above the national median income.

- **Population 20+ years of age by economical standard (counts):** The number of individuals of a given economical standard category living in a given DeSO zone (this is a measure based on each household's total income and composition). There are a total of 2 such covariates: below the national median economical standard and above the national median economical standard.

S5. Created covariates

Recall the notions of 'original' and 'created' covariates from Section 2. We illustrate some of the selected 'original' and 'created' covariates considered in this study in Figure S1. Below, we provide

a description of the construction of all created covariates.

S5.1. Benchmark intensity covariate for the ambulance data

Recall that our semi-parametric approach is based on the idea of including a non-parametric intensity estimate, the so-called benchmark intensity, as a covariate in our log-linear intensity form. To generate the benchmark intensity for the ambulance data, we employ our new algorithm to obtain the intensity estimate found in panel (e) of Figure S2. We further compare the result with the aforementioned approaches (recall Section S3), in particular the approach of Bayisa et al. (2020), which is based on 5 clusters (this number has been obtained through visual inspection). The resulting intensity estimates for the ambulance data can be found in panels (b)-(g) in Figure S2. Note that throughout we have used a Gaussian kernel in combination with the aforementioned local edge correction factor. We argue that the state estimation approach and the Poisson process likelihood leave-one-out cross-validation approach tend to under-smooth the data and thereby do not reflect the general overall variations of the data, whereas the approach of Cronie and Van Lieshout (2018) and the K-means clustering based bandwidth selection of Bayisa et al. (2020) instead tend to over-smooth the ambulance data. Recall that the number of clusters, which is a necessary input in the K-means clustering based bandwidth selection of Bayisa et al. (2020), has to be selected through visual inspection.

Looking at panel (g) of Figure S2, we see that by employing our proposed KL index to automatically select the number of clusters, we obtain a total of 16 clusters. We argue that, compared to the different panels in Figure S2, the new heuristic algorithm performs the best in terms of balancing over- and under-smoothing of the ambulance call events. Interestingly, we see that this bandwidth is very similar to the one obtained with the rule of thumb of Scott (2015); see panel (e).

S5.2. Creation of (road) network-related covariates

The role of the road network-related covariates is to control that the fitted model generates events close to the underlying road network. But, as previously indicated, the original road network-related covariates are not of the form $z_i(s)$, $s \in W$. We here propose to treat the road networks under

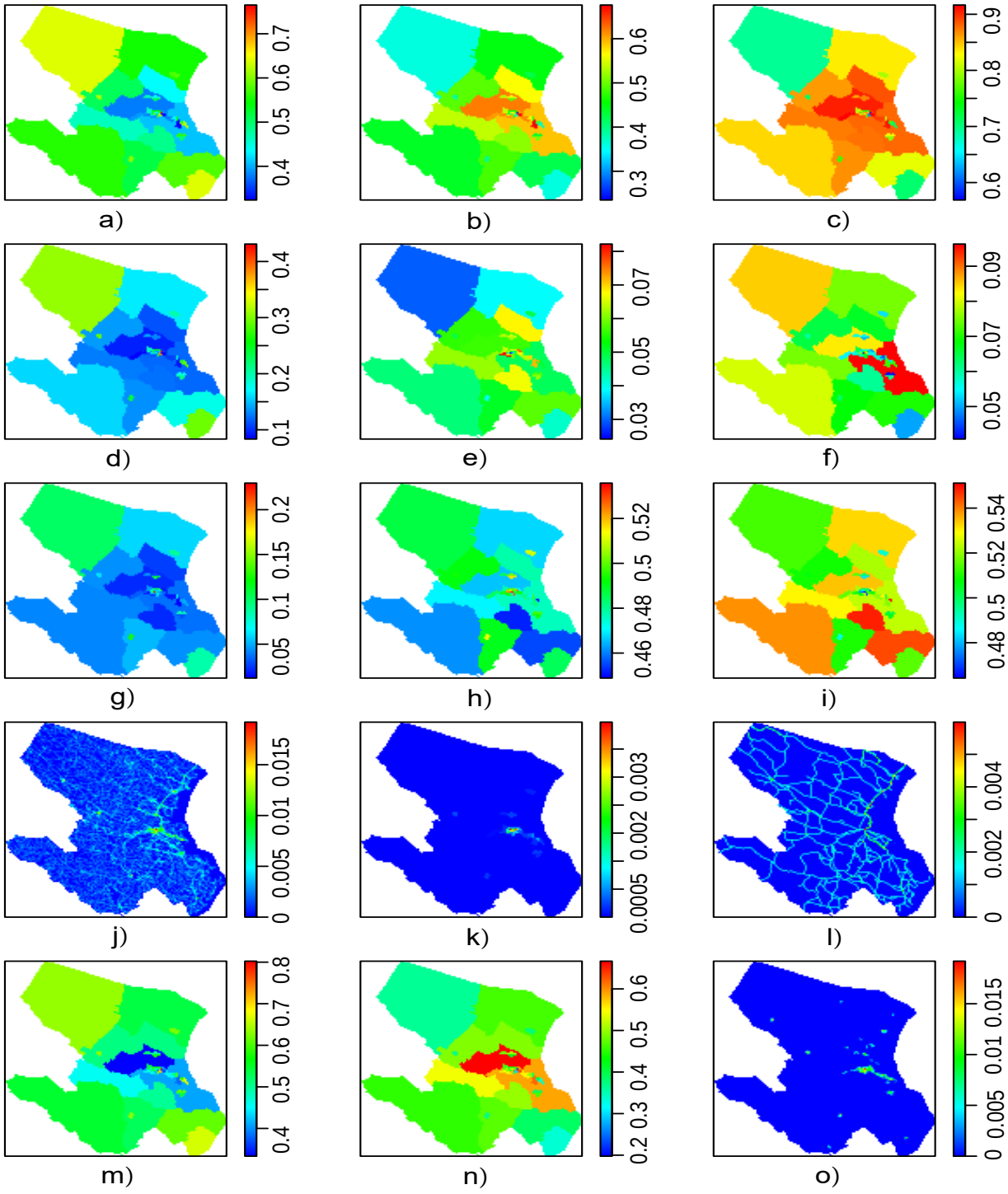

Fig. S1. Spatial demographic and road network-related covariates. Population 20+ years old income status: a) below and b) above median income. Population 20-64 years old employment status: c) employed and d) unemployed. Population by age: e) 35-39 years old, f) 55-59 years old, and g) 80+ years old. Population by sex: h) male and i) female. j) complete road network line density. k) main road network line density. l) population density. Households 20+ years old economic status: m) below and n) above median income. o) Densely populated line density. All covariates, except for j), k), l) and o), are representing proportions.

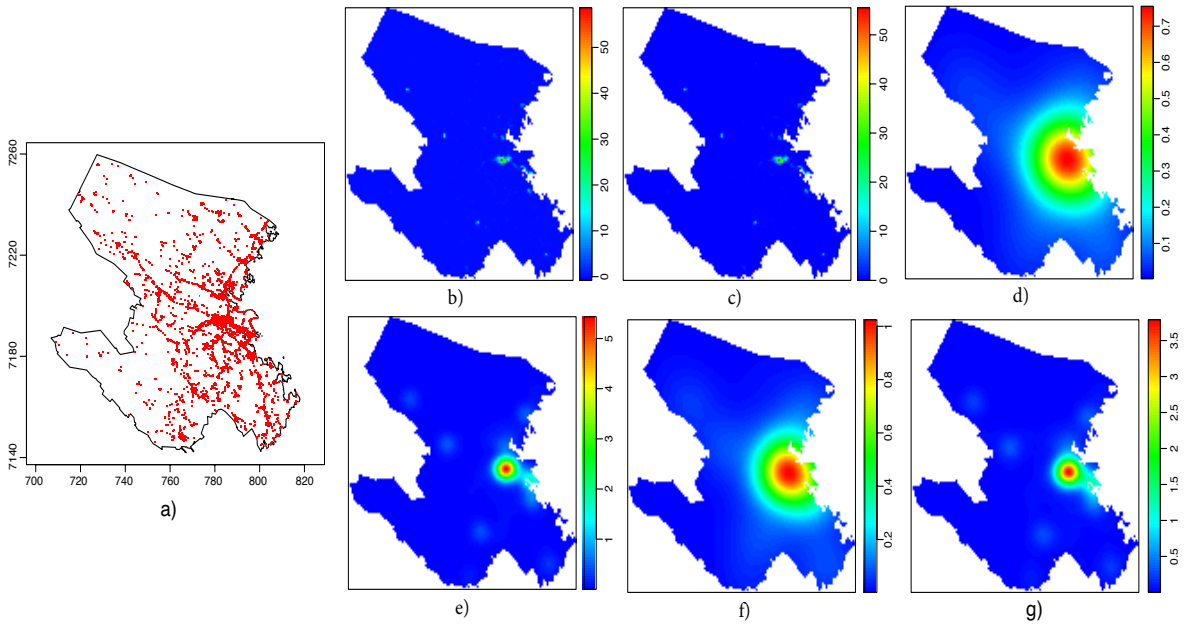

Fig. S2. The roles of different bandwidth selection methods using Gaussian kernel density estimation for the unmarked ambulance call data in panel (a). b) State estimation (Diggle, 1985). c) Poisson likelihood cross-validation (Loader, 1999). d) Cambell's formula-based bandwidth selection (Cronie and Van Lieshout, 2018), e) Scott's rule (Scott, 2015), f) K-means clustering (Bayisa et al., 2020). g) The new heuristic algorithm developed in this work. Note that what interests us here are the relative scales rather than the raw scales; the values in the plots above have been multiplied by 1000 for ease of visualisation.

consideration as line segment patterns (Baddeley et al., 2015), which essentially means that each road network considered is a realisation of a point process in the space of line segments in $\mathbb{R}^2$. The spatial covariate corresponding to a given line segment pattern is then given by the estimated line segment intensity, which is obtained as the convolution of an isotropic Gaussian kernel with the line segments of the pattern in question. Practically, such an estimate may be obtained through the function `density.psp` in the R package spatstat, and the standard deviation of the Gaussian kernel, the bandwidth, determines the degree of smoothing. The default bandwidth choice in `density.psp` is given by the diameter of the observation window multiplied by 0.1. As an alternative, we propose to use our new heuristic algorithm for bandwidth selection, which is achieved by letting $W = [0, \infty)$, and letting the observations $\mathbf{x} = \{x_1, x_2, \cdots, x_n\} \subseteq W$ considered in Algorithm 2 represent the lengths of the line segments. We here consider the following network-related covariates:

- Complete road networks line density/intensity. It is a spatial pattern of line segments, which is converted to a pixel image. The value of each pixel in the image is measured as the total

length of intersection between the pixel and the line segments [1].

- Main road networks line density/intensity. It is a spatial pattern of line segments, which is converted to a pixel image. The value of each pixel in the image is measured as the total length of intersection between the pixel and the line segments [1].

- Densely populated line density/intensity. It is a spatial pattern of line segments, which is converted to a pixel image. The value of each pixel in the image is measured as the total length of intersection between the pixel and the line segments [1].

- Bus stops intensity. It is a spatial point pattern, which is converted to a pixel image. The value of each pixel is an intensity, which is measured as "points per unit area" [1].

- Shortest distance to bus stops. It is the shortest distance in meters from the ambulance location data to the bus stops [1].

- Shortest distance to densely populated areas. It is the shortest distance in meters from the ambulance location data to the densely populated areas [1].

- Shortest distance to main road networks. It is the shortest distance in meters from the ambulance location data to the main road networks [1].

- Shortest distance to complete road networks. It is the shortest distance in meters from the ambulance location data to the complete road networks [1].

Figure S3 compares the default bandwidth choice of `density.psp` to our heuristic algorithm, and it clearly suggests that the line segment intensities generated using the heuristic algorithm bandwidth selection have captured the spatial pattern of the line segments in the observed data better than the default bandwidth choice. Note that what interests us here are the relative scales rather than the raw scales; the values in the plots have been multiplied by 1000 for ease of visualisation.

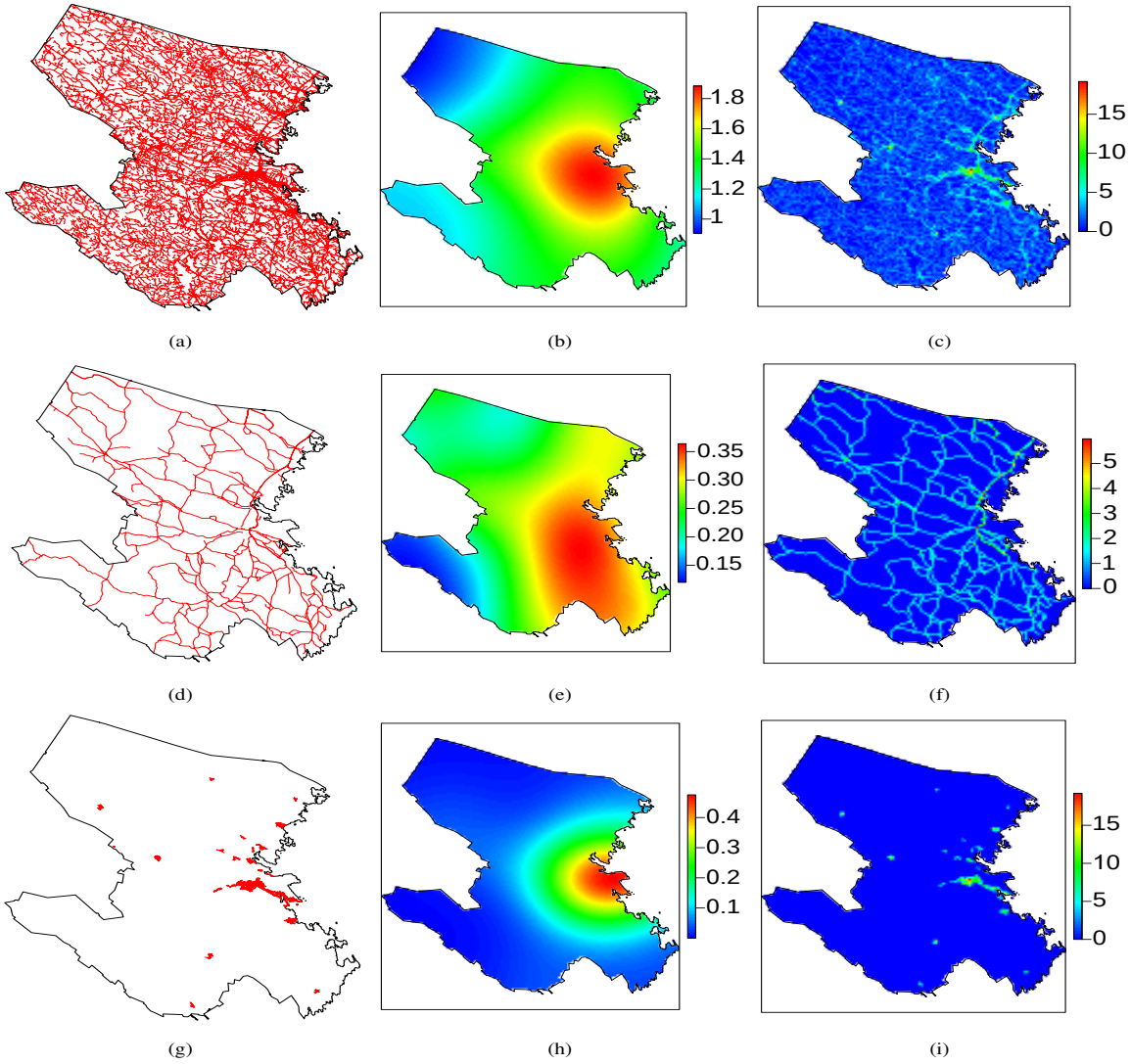

Fig. S3. Creation of covariates from line segment patterns. The first column represents a) the complete road network, d) the main road network and g) the densely populated area road network, respectively. Created road network covariates using the spatstat function `density.psp`: (b), (e) and (h) have been obtained using the default bandwidth of `density.psp`, whereas (c), (f) and (i) have been obtained using the new heuristic bandwidth selection algorithm. The values in the plots have been multiplied by 1000 for ease of visualisation.

S6. Data analysis results

S6.1. Modelling the ambulance call intensity function

Figure S4 shows the estimated coefficients of the spatial covariates, using both lasso regularisation and lasso-like elastic-net regularisation, for the spatial point pattern constituting the events with priority level 1. Note that the numbers at the top of each panel in Figure S4 indicate the number of

spatial covariates with non-zero coefficients, i.e. the number of covariates that are associated with the fitted spatial intensity functions for the indicated value of λ. We clearly see how a large number of covariates are quickly excluded as we increase $\log \lambda$.

Using the aforementioned covariates and their first-order interaction terms, the trace plots of the estimated parameters for lasso regularization and lasso-like elastic-net regularisation with $\alpha = 0.95$ are shown in Figure S4. At their corresponding optimal tuning parameters, which are shown by blue-coloured vertical lines in Figure S4, lasso and lasso-like elastic-net regularisations have selected 185 (18.69%) and 207 (20.91%) of the covariates that are associated with the intensity of the events. As it is expected, even though the lasso-like elastic-net regularisation selects more variables than the lasso regularization, it forces the highly correlated covariates to have similar coefficients (Friedman et al., 2010a; Yue and Loh, 2015). As a result, the correlated covariates have similar roles for the intensity of the events, and their presence in the fitted model can be advantageous for interpretation. Therefore, hereafter, we use lasso-like elastic-net regularisation with $\alpha = 0.95$.

A grid of λ values has been exploited to train the proposed model, and among the candidate λ values, the one which gives the smaller deviance, D, has been selected as an optimal estimate of λ. The optimal elastic-net regularisation parameter λ has been selected using ten-fold cross-validation as shown in Figure S5.

The two vertical lines in the figure have been drawn to show the location of the logarithm of the optimal estimate of the tuning parameter (blue) and the location of the logarithm of the estimate of the tuning parameter that is one standard error away from the optimal estimate (red); the one-standard-error-rule (Hastie et al., 2017) says that one should go with the simplest model, which is no more than one standard error worse than the best model.

For the priority level 1 and 2 events, about 20.91% and 36.06% of the spatial covariates have been associated with/included in the final intensity function estimates, respectively. The lasso-like elastic-net has discerned about 17.68% and 35.76% of the spatial covariates which determine the spatial intensities of the point patterns corresponding to male and female, respectively.

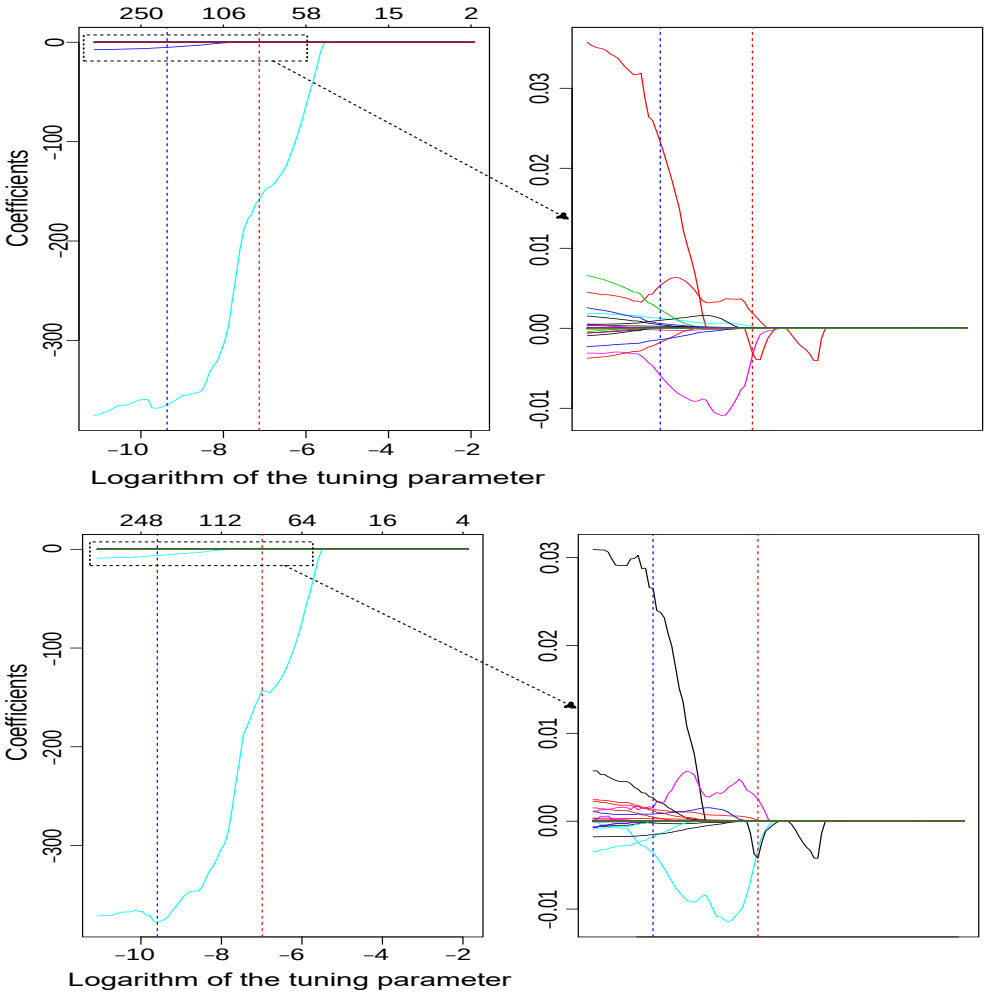

Fig. S4. A trace plot of the estimated coefficients for lasso (the first row) and lasso-like elastic-net (the second row) regularisations. The numbers at the top of each panel indicate the number of spatial covariates with non-zero estimated coefficients. The plots in the right panel of the left plots represent the zoomed in portion of the region marked by the rectangular region. For the purpose of visualization, the estimated coefficients are scaled (divided) by 10^{10}.

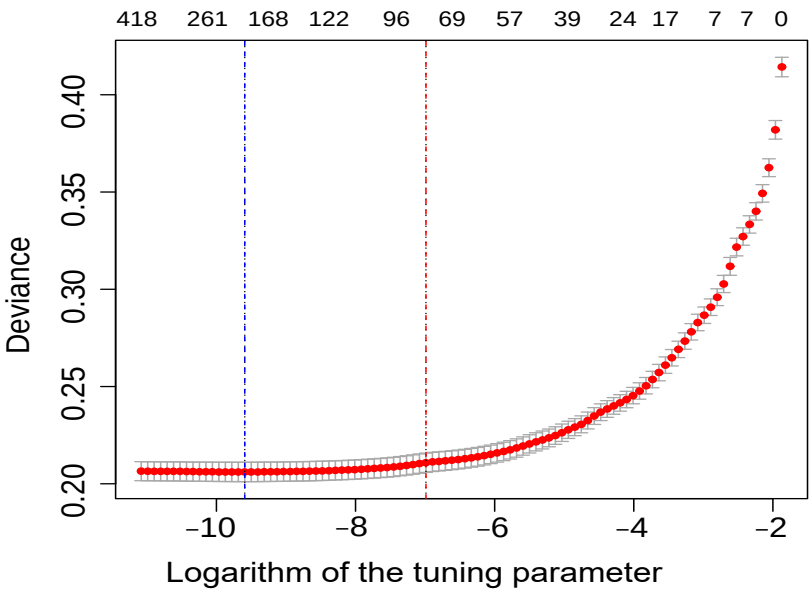

Fig. S5. An optimal tuning parameter selection via ten-fold cross-validation: location of the logarithm of the optimal estimate of the tuning parameter (blue) and the location of the logarithm of the estimate of the tuning parameter that is one standard error away from the optimal estimate (red).

Let P_1, P_2, M, and F denote emergency alarm call events with marks priority level 1, priority level 2, male and female, respectively, whereas Unm denotes unmarked emergency alarm call events.

?tablename? S1. The estimated dense models for the marginals and unmarked emergency alarm call events/patterns. The dots in the table represent small coefficients of covariates that are shrunk to zero.

No.	Covariates	P_1	P_2	M	F	Unm
	Spatial locations:					
1[1]	x	0.823	0.826	0.951	0.522	0.725
2[2]	y	-0.144	-0.225	-0.175	-0.463	-0.243
	Shortest distance to:					
3[2]	Bus stops	-42.810	-52.120	-53.820	-49.810	-44.776
4[2]	Densely populated area	-3.685	-4.380	-2.344	-4.246	-3.674
5[2]	Main road networks	-32.030	-16.740	-16.890	-20.800	-24.907
6[2]	Complete road networks	-206.300	-238.600	-203.100	-259.200	-229.287
7	Population density	18.991	-141.524	68.728	·	-31.185
8	Benchmark intensity	949.209	3802.066	-15119.642	·	276.511
	Line densities:					
9	Complete road networks	243.727	280.890	308.540	304.280	275.264
10	Main road networks	42.073	-54.408	-78.472	-57.196	25.836
11	Densely populated	-158.792	-155.344	-175.181	-157.708	-161.800
12[2]	Bus stops density	5.481	-1.633	6.046	0.819	4.941
	Population by:					
	a) **employment status:**					
13	Employed	·	·	·	·	·

Continued on next page

[1]The parameter estimates are divided by 10^{-5} for better visualisation of the estimates
[2]The parameter estimates are divided by 10^{5} for better visualisation of the estimates

?tablename? S1 – continued from previous page

No.	Covariates	P_1	P_2	M	F	Unm
14	Unemployed	·	·	·	·	·
	b) **income status:**					
15	Below national median income	4.747	0.387	0.365	0.866	4.446
16	Above national median income	-0.116	-0.001	-0.001	-0.010	-0.079
	c) **educational level**:					
17	Pre-high school	-0.355	-2.112	-1.996	-2.260	-2.150
18	High school	1.808	0.962	2.001	1.365	1.190
19	Post-secondary – less than 3 years	3.105	1.196	2.534	·	1.034
20	Post-secondary – 3 years or longer	·	·	·	·	·
	d) **age:**					
21	0 – 5	7.999	5.926	4.683	6.189	6.000
22	6 – 9	-1.115	-12.052	-3.768	-12.189	-4.575
23	10 – 15	4.505	1.310	5.901	3.828	7.542
24	16 – 19	·	4.433	2.148	0.247	-3.813
25	20 – 24	·	-3.429	·	·	0.063
26	25 – 29	-0.179	-0.764	-1.902	-4.533	-0.163
27	30 – 34	-3.306	-3.821	·	·	-0.599
28	35 – 39	1.486	12.824	4.757	15.668	5.583
29	40 – 44	-3.386	0.112	-8.503	-3.040	0.737
30	45 – 49	-1.064	-3.946	-4.763	-4.098	-4.794
31	50 – 54	-10.938	-8.441	-7.539	-5.996	-10.981
32	55 – 59	3.340	4.501	6.950	7.621	6.450
33	60 – 64	-1.003	0	1.370	-0.171	-0.422
34	65 – 69	-13.011	-5.694	-6.573	-8.789	-7.743

Continued on next page

?tablename? S1 – continued from previous page

No.	Covariates	P_1	P_2	M	F	Unm
35	70 – 74	0.131	0.025	·	0.098	-4.211
36	75 – 79	3.020	·	·	3.359	2.524
37	80 +	·	0.294	·	·	2.507
	e) **gender:**					
38	Female	-0.816	2.586	1.431	2.145	-1.302
39	Male	0.058	-0.017	-0.004	-0.043	0.106
	f) **background:**					
40	Swedish background	-0.177	-1.460	-1.715	·	-0.671
41	Foreign background	0.046	0.028	0.062	·	0.044
	g) **household economic standard:**					
42	Below the national median	-3.741	·	·	·	-3.195
43	Above the national median	0.131	·	·	·	0.079
	Intercept parameter estimate:	2.410	8.409	3.646	25.929	10.886

?tablename? S2. The estimated parsimonious/sparse models for the marginal and unmarked emergency alarm call events/patterns. The dots in the table represent small coefficients of covariates that are shrunk to zero.

No.	Covariates	P_1	P_2	M	F	Unm
	Spatial locations :					
1[1]	x	0.098	0.324	0.390	0.119	0.128
2[2]	y	-0.077	·	-0.002	-0.015	-0.133
	Shortest distance to:					
3[2]	Bus stops	-38.910	-45.250	-48.560	-44.100	-41.486
4[2]	Densely populated area	-2.633	-2.882	-1.567	-2.899	-2.898
5[2]	Main road networks	-22.740	-8.121	-8.359	-11.310	-19.045
6[2]	Complete road networks	-65.710	-54.770	-57.790	-64.720	-101.270
7	Population density	·	·	·	·	·
8	Benchmark intensity	·	·	·	·	·
	Line densities:					
9	Complete road networks	273.826	315.707	319.707	335.114	309.844
10	Main road networks	74.692	·	·	·	34.869
11	Densely populated	-141.438	-129.585	-152.325	-141.335	-159.128
12[2]	Bus stops density	6.014	0.526	0.766	·	6.046
	Population by:					
	a) **employment status:**					
13	Employed	·	·	·	·	·
14	Unemployed	·	·	·	·	·
	b) **income status:**					

Continued on next page

[1]The parameter estimates are divided by 10^{-5} for better visualisation of the estimates
[2]The parameter estimates are divided by 10^{5} for better visualisation of the estimates

?tablename? S2 – continued from previous page

No.	Covariates	P_1	P_2	M	F	Unm
15	Below national median income	·	·	·	·	·
16	Above national median income	·	·	·	·	·
	c) **educational level:**					
17	Pre-high school	·	·	·	·	·
18	High school	·	·	·	·	·
19	Post-secondary – less than 3 years	·	·	·	·	·
20	Post-secondary – 3 years or longer	·	·	·	·	·
	d) **age:**					
21	0 – 5	1.917	2.120	2.780	3.782	3.786
22	6 – 9	0.045	·	1.197	·	·
23	10 – 15	2.123	2.552	3.126	0.937	0.924
24	16 – 19	·	·	·	·	·
25	20 – 24	·	·	·	·	·
26	25 – 29	·	·	·	·	·
27	30 – 34	·	·	·	·	·
28	35 – 39	·	4.098	·	9.585	1.289
29	40 – 44	·	·	·	·	·
30	45 – 49	·	·	·	·	·
31	50 – 54	·	·	·	·	-1.541
32	55 – 59	·	-0.901	·	·	·
33	60 – 64	-6.267	-2.451	-3.532	·	-3.204
34	65 – 69	-0.466	·	·	·	-0.144
35	70 – 74	·	·	·	·	·
36	75 – 79	·	·	·	·	·

Continued on next page

?tablename? S2 – continued from previous page

No.	Covariates	P_1	P_2	M	F	Unm
37	80 +	·	·	·	·	·
	e) **gender:**					
38	Female	·	·	·	·	·
39	Male	·	·	·	·	·
	f) **background**:					
40	Swedish background	·	·	·	·	·
41	Foreign background	·	·	·	·	·
	g) **household economic standard**:					
42	Below the national median	·	·	·	·	·
43	Above the national median	·	·	·	·	·
	Intercept parameter estimate:	3.255	-4.821	-4.872	-2.757	6.651

S7. Further analysis on emergency alarm call events with priority 2 and unmarked events

In a broad sense, the three approaches, i.e., i) the estimated spatial intensity based on the original spatial covariates plus the first-order interaction terms, and the estimated spatial intensities that are obtained using ii) the estimated dense model and iii) the parsimonious/sparse model based on only the original spatial covariates, have similar performance for the marginal point patterns corresponding to the marks priority level 1, male, and female. Here we present the results of the three approaches, i.e., the estimated spatial intensity using the estimated model based on the original spatial covariates and the first-order interaction terms; the estimated spatial intensities using the estimated dense and the parsimonious/sparse models based on only the original spatial covariates. Analogously to Figure 2, Figure S6 illustrates the fitted spatial intensities using the three approaches for the marginal point pattern with priority level 2.

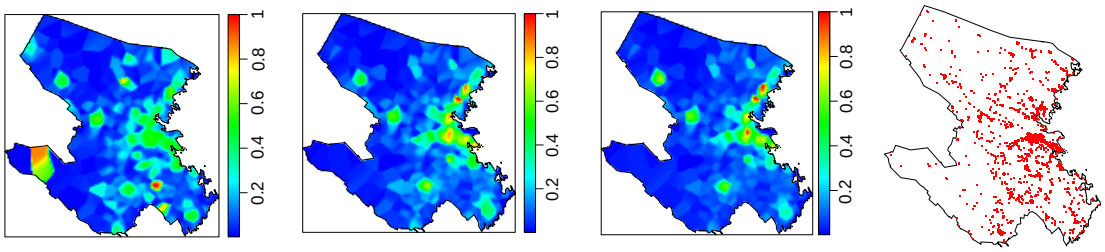

Fig. S6. The estimated spatial intensities of the emergency alarm call events with priority level 2 using the estimated model based on the original spatial covariates and the first-order interaction terms, the estimated dense and sparse models, which are based on only the original covariates, respectively. The performance of the estimated models can be evaluated based on the pattern of the events in the observed marginal spatial point pattern. Note that we have scaled the intensity estimates to range between 0 and 1 so that we may compare them more easily.

The figure clearly demonstrates that the dense and sparse models capture the spatial distribution of the events more accurately than the estimated model based on the original spatial covariates and the first-order interaction terms. Those spatial covariates in the dense model with non-zero estimated coefficients that continue to exist in the sparse model, i.e., those spatial covariates with non-zero estimated coefficients, have a strong association with the spatial intensity of the events. As the figure shows, the estimated dense and sparse models perform well in capturing the spatial variation of the events, and, relatively speaking, we can loosely and strongly interpret the results from the dense and sparse models, respectively.

Next, we examine the modelling of the spatial intensity function for unmarked ambulance call data, i.e., we ignore the marks of the ambulance call data, using only the original spatial covariates. Here, we are interested in recognising how the marks influence the inclusion of different covariates. As in the modelling of each marginal spatial point pattern, the spatial intensity function modelling of the unmarked spatial point pattern based solely on the original spatial covariates is expected to better capture the spatial distribution of the emergency alarm call events than the corresponding model setting including both the original spatial covariates and the first-order interaction terms. Then, using only the original spatial covariates, we focus on modelling the spatial intensity function of the unmarked ambulance call data. The last column in Table S1 presents the estimated dense model for the unmarked spatial point pattern, while the last column in Table S2 displays the corresponding sparse model for the unmarked spatial point pattern that is obtained by utilising the one-standard-

error rule. Figure S7 shows the estimated spatial intensities of the unmarked ambulance call data using the estimated dense and sparse models.

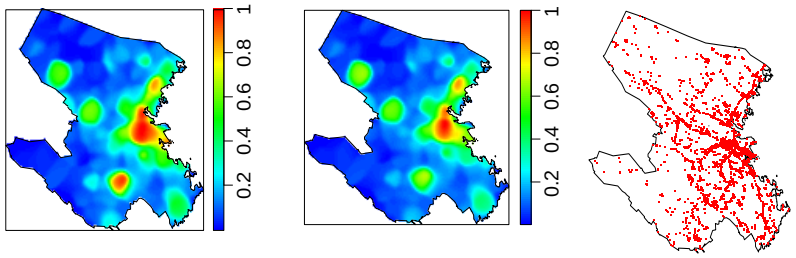

Fig. S7. The estimated spatial intensities of the unmarked ambulance call data using the estimated dense model (left) and sparse model (middle). The pattern of the events in the observed unmarked spatial point pattern (right) can be used to assess the performance of the estimated models. To make it easier to compare the intensity estimates, we scaled them to have a range between 0 and 1.

In the presence of the other covariates, from the coefficients for the spatial locations in Table S1, we see that there is a slight positive trend to the east/the coast and to the south. This is not surprising as this is where the majority of the population lives; most likely, we here catch some of the residual effects which the demographic covariates do not manage to catch. Moreover, regarding the distance-based covariates, which are all on the same scale, the biggest effect, in general, seems to come from the distance to the road network, which is not surprising, followed by the distance to bus stops. When looking simultaneously at the coefficients for the population density and the benchmark intensity, which is intended to reflect mobility, for priority level 1 calls both tend to increase the risk; they thus enhance each other which we interpret as there being a mobility structure which we actually catch here. In the case of the unmarked events and the priority level 2 events, on the other hand, these two go in opposite directions, which could indicate that they cancel each other out, to some degree, as well as that there may be places with a low population density where there, relatively speaking, is a higher frequency of events. Interestingly, for females these two covariates do not seem to have an effect on the outcome whereas for men we see the reversed of what we saw in the case of priority level 2. We further note that the bus stop density, which is intended to reflect an additional layer of mobility, has positive coefficients for all marks except priority level 2; we interpret this as more severe cases happening in places where there is a larger amount of human mobility. Turning to the line density covariates, not surprisingly, we have that the main road

network impacts the risk positively. Turning to the socioeconomic covariates, we find that living in an impoverished area, versus in an above-national median income area, markedly increases the risk, irrespectively of the mark/group considered. Quite interestingly, areas with a higher proportion of educated persons seem to have a higher event frequency, which may be a reflection of lack of education making people adverse to calling an ambulance, in favour of seeking help in other ways. Employment rate does not seem to play a role. We further see that the most priority level 1 calls occur in areas inhabited by larger proportions of certain young and certain old age categories. With a few exceptions and a bit of variation, the picture is more or less the same for the other marks/groups. In fact, having a high proportion of young living in an area seems to be one of the main sociodemographic drivers here. Despite the effect being weak, we see that having a higher proportion of males in an area decreases number of priority level 2 calls but increases the number of priority level 1 calls. Moreover, having a (non-)Swedish background does not seem to have a big impact, which could indicate that it is primarily the poverty of one's area (impoverished areas are to a larger degree inhabited by people of non-Swedish background) that has an impact on the number events taking place there. By considering the intercept terms as indicators for how much of the spatial variation is caught by the included covariates, we see that the mark category female has a markedly higher intercept, which could indicate that additional covariates could have increased the fit of the model. We would like to point out that the above interpretations are contingent on the presence of other covariates in the model with non-zero coefficients.

Based on the results in Table S2, we can make a strong interpretation in comparison to the results in Table S1. Here we see that there is a slight positive trend to the east/the coast and to the south and that, in the case of the distance-based covariates, the biggest effect seems to come from the distance to the road network, followed by the distance to bus stops. The actual bus stop density in an area also seems to increase the risk. Moreover, here, none of the sociodemographic covariates, except for the age proportion covariates, seem to have an impact on the outcome. In particular, for each marks/groups we see that having a higher proportion of children in an area has a clear impact on the

outcome. Here, having a larger proportion elderly does not seem have an effect. In fact, when the number of retirees increases, the risk seems to decrease. Here, throughout, the orders of magnitude of the intercept terms is quite small, which could be an indication of a good fit.

?refname?

Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial point patterns: methodology and applications with R*. Chapman and Hall/CRC.

Baddeley, A. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns: (with discussion). *Australian & New Zealand Journal of Statistics*, 42(3):283–322.

Bayisa, F. L., Ådahl, M., Rydén, P., and Cronie, O. (2020). Large-scale modelling and forecasting of ambulance calls in northern sweden using spatio-temporal log-gaussian cox processes. *Spatial Statistics*, page 100471.

Cronie, O. and Van Lieshout, M. N. M. (2018). A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika*, 105(2):455–462.

Davies, T. M., Marshall, J. C., and Hazelton, M. L. (2018). Tutorial on kernel estimation of continuous spatial and spatiotemporal relative risk. *Statistics in medicine*, 37(7):1191–1221.

Diggle, P. (1985). A kernel method for smoothing point process data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 34(2):138–147.

Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. CRC press.

Friedman, J., Hastie, T., Höfling, H., Tibshirani, R., et al. (2007). Pathwise coordinate optimization. *The annals of applied statistics*, 1(2):302–332.

Friedman, J., Hastie, T., and Tibshirani, R. (2010a). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1.

Friedman, J., Hastie, T., and Tibshirani, R. (2010b). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, Articles*, 33(1):1–22.

Hastie, T., Tibshirani, R., and Friedman, J. (2017). The elements of statistical learning, 12th printing, ser.

Krzanowski, W. J. and Lai, Y. (1988). A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics*, pages 23–34.

Loader, C. (1999). *Local regression and likelihood*. Springer, New York.

Moradi, M. M., Cronie, O., Rubak, E., Lachieze-Rey, R., Mateu, J., and Baddeley, A. (2019). Resample-smoothing of Voronoi intensity estimators. *Statistics and Computing*, pages 1–16.

Ogata, Y. (2004). Space-time model for regional seismicity and detection of crustal stress changes. *Journal of Geophysical Research: Solid Earth*, 109(B3).

Ogata, Y., Katsura, K., and Tanemura, M. (2003). Modelling heterogeneous space–time occurrences of earthquakes and its residual analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4):499–509.

Okabe, A., Boots, B., Sugihara, K., and Chiu, S. N. (2009). *Spatial tessellations: concepts and applications of Voronoi diagrams*, volume 501. John Wiley & Sons.

Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

Thurman, A. L., Fu, R., Guan, Y., and Zhu, J. (2015). Regularized estimating equations for model selection of clustered spatial point processes. *Statistica Sinica*, pages 173–188.

Waagepetersen, R. (2008). Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika*, 95(2):351–363.

Yue, Y. and Loh, J. M. (2015). Variable selection for inhomogeneous spatial point process models. *Canadian Journal of Statistics*, 43(2):288–305.