Large-scale modelling and forecasting of ambulance calls in northern Sweden using spatio-temporal log-Gaussian Cox processes

Fekadu L. Bayisa, Patrik Rydén, Ottmar Cronie*

Department of Mathematics and Mathematical Statistics, Umeå University, 901 87, Umeå, Sweden

Abstract

In order to optimally utilise the resources of a country’s prehospital care system, i.e. ambulance service(s), it is crucial that one is able to spatio-temporally forecast hot-spots, i.e. spatial regions and periods with an increased risk of seeing a call to the emergency number 112 which results in the dispatch of an ambulance. Such forecasts allow the dispatcher to make strategic decisions regarding e.g. the fleet size and where to direct unoccupied ambulances. In addition, simulations based on forecasts may serve as the starting point for different optimal routing strategies. Although the associated data typically comes in the form of spatio-temporal point patterns, point process based modelling attempts in the literature has been scarce. In this paper, we study a unique set of Swedish spatio-temporal ambulance call data, which consists of the spatial (gps) locations of the dispatch addresses and the associated days of occurrence of the calls. The spatial study region is given by the four northernmost regions of Sweden and the study period is January 1, 2014 to December 31, 2018. Motivated by the non-infectious disease nature of the data, we here employ log-Gaussian Cox processes (LGCPs) for the spatio-temporal modelling and forecasting of the calls. To this end, we propose a K-means based bandwidth selection method for the kernel estimation of the spatial component of the separable spatio-temporal intensity function. The temporal component of the intensity function is modelled by means of Poisson regression, using different calendar covariates, and the spatio-temporal random field component of the random intensity of the LGCP is fitted using simulation via the Metropolis-adjusted Langevin algorithm. A study of the spatio-temporal dynamics of the data shows that a hot-spot can be found in the south eastern part of the study region, where most people in the region live and our fitted model/forecasts manage to capture this behaviour quite well. The fitted temporal component of the intensity functions reveals that there is a significant association between the expected number of calls and the day of the week as well as the season of the year. In addition, non-parametric second-order spatio-temporal summary statistic estimates indicate that LGCPs seem to be reasonable models for the data. Finally, we find that the fitted forecasts generate simulated future spatial event patterns which quite well resemble the actual future data.

Keywords: Ambulance call data, Forecasting/prediction, K-means clustering based bandwidth selection, Metropolis-adjusted Langevin Markov chain Monte Carlo, Minimum contrast estimation, Spatio-temporal point process modelling

1. Introduction

To obtain a desired outcome of the prehospital care in a given country, which includes keeping the mortality low, the ambulance response times should be as small as possible (Blackwell and Kaufman, 2002; O’keefe et al., 2011; Pell et al., 2001). Ideally, any given country would have the financial means to let the number of active ambulances, at any given time of the day and in any given region, be so large that the higher quantiles of the empirical response time distribution would be very small. Unfortunately, this is not/rarely the case and many countries struggle to make the most of their existing resources. Even Sweden, for example, which is a prime example of a well-fare state with large resources at hand, struggles...
The data at hand, which consists of the emergency alarm calls in the four (northernmost) Swedish regions Västerbotten, Norrbotten, Västernorrland and Jämtland-Härjedalen, represents a typical case of a spatio-temporal point pattern \((\{(s_i, t_i)\}_{i=1}^n)\) where \(s_i\) is the spatial (gps) location of the \(i\)th call and \(t_i\) is the associated call time (here obtained as the date). Although there is a vast literature on prehospital care optimisation (see e.g. Aringhieri et al. (2017) and the references therein), surprisingly very little effort has been spent on actually modelling the calls by means of point processes, which are the models used for point patterns. Some attempts have been made (Zhou et al., 2015, 2016) but, based on our understanding, the proposed approaches did not justify their model choices on the basis of non-parametric point pattern analyses so it remains unclear how well the proposed models are generally suited for this type of data.

The primary goal of this work is to describe the spatio-temporal dynamics of the calls with the main focus of the study being to identify hotspot-regions within the spatial study region as well as generating a short-term forecast-model that allows us to simulate realistic future emergency alarm call locations, which e.g. may be used to optimally allocate prehospital resources and design prehospital strategies, such as ambulances dispatching strategies. Since the location and time of each emergency alarm call are known, a spatio-temporal point process provides the right framework to model the space-time dynamics in the emergency alarm call data. As will be verified by our non-parametric analysis, due to the data’s non-infectious epidemiological nature, inhomogeneous log-Gaussian Cox processes (LGCPs) (Møller et al., 1998) tend to be particularly suited to model the space-time dynamics of the emergency alarm calls – LGCPs take spatial and/or temporal clustering/aggregation into account. Moreover, through the LGCP framework we may do spatial forecasting for future time periods and, consequently, simulate “realistic” future emergency alarm scenarios. Following Diggle (2013), the space-time intensity function of the LGCP considered is expressed as a product of two deterministic baseline functions and one stochastic log-Gaussian random field component. The deterministic components represent purely spatial and purely temporal variations while the stochastic component deals with the unexplained space-time variation; the employed approach is semi-parametric in the sense that the spatial baseline function is taken to be a spatial density estimate whereas the temporal one is taken to be a temporal intensity estimate which is based on temporal covariates. The spatial density estimate, which is a normalised spatial intensity estimate (Cronie and Van Lieshout, 2018; Moradi et al., 2019; van Lieshout, 2012), is obtained through kernel intensity estimation and we here propose a new clustering approach to select the associated bandwidth. In the case of the temporal intensity estimation, we employ a Poisson regression model which is based on temporal covariates such as ”day of the week”. To fit the stochastic component of the model, minimum contrast estimation (Baddeley et al., 2015; Diggle, 2013; Møller and Waagepetersen, 2003) is the natural way to proceed. Finally, the forecasting is carried out by exploiting the Ornstein-Uhlenbeck approximation of Brix and Diggle (2001).

The structure of the article is as follows. Section 2 briefly presents the ambulance call data and Section 4 provides an overview of the statistical methods used in this work. We present the results of the study in section 5 and discuss the implication of the results as well as provide a precise summary of the work in section 6.
2. Data

The spatio-temporal point pattern \( \{(s_i, t_i)\}_{i=1}^{n} \) under study consists of \( n = 444283 \) events, i.e. ambulance/emergency alarm calls, and for each event we have recorded its spatial (gps) location \( s_i \in \mathbb{R} \subseteq \mathbb{R}^2 \) as well as its date of occurrence, \( t_i \in T \). The temporal domain, \( T \), of the spatio-temporal data ranges from January 1, 2014 to December 31, 2018.

The left-hand panel in Figure 1 displays the spatial locations of the calls while the right-hand panel presents the call count over time. From the left-hand panel of the figure, we can see that the calls are unevenly distributed over the four northern regions (Norrbotten, Västerbotten, Västernorrland, and Jämtland-Härjedalen) of Sweden. In addition, as expected, the calls tend to be located in populated areas, which in turn lie along the road network – this will make the semi-parametric statistical inference quite challenging. For data privacy issues, the axis labels of the left-hand panel in the figure have been obtained by dividing the xy-coordinates of the alarm call locations by 1000.

![Fig. 1. The spatial and temporal components of the spatio-temporal emergency alarm calls data. Left: Spatial locations of the calls. Right: The call count as a function of time (day-scale).](image)

Note that since we will use spatio-temporal point processes for the modelling, we will model daily accumulated data as if they were in fact observed on a temporal continuum, i.e. on the interval between midnight and midnight on these two dates, we will in effect treat the calls occurring during a given day as though they were uniformly distributed over that day.

3. Point process preliminaries

A (simple) spatio-temporal point process (STPP) (Diggle, 2013; Diggle and Gabriel, 2010; González et al., 2016) is essentially a random countable subset of \( \mathbb{R}^2 \times \mathbb{R} \) such that, with probability one, i) the number of points of \( \mathcal{Y} \) within any bounded (Borel) subset \( A \) of \( \mathbb{R}^2 \times \mathbb{R} \), denoted \( N(A) \), is finite, ii) the cardinality \( N((s, t)) \in \{0, 1\} \) for any \( (s, t) \in \mathbb{R}^2 \times \mathbb{R} \), and iii) we equip the temporal domain \( \mathbb{R} \) with the natural complete ordering of the real numbers. Moreover, throughout we will consider the STPP obtained as the restriction

\[
\mathcal{Y} = \{(s_i, t_i)\}_{i=1}^{N} \subseteq \mathcal{R} \times [0, T], \tag{1}
\]

of an STPP in \( \mathbb{R}^2 \times \mathbb{R} \), where \( \mathcal{R} \subseteq \mathbb{R}^2 \) is a bounded spatial domain and \( [0, T] \subseteq \mathbb{R} \) is a bounded temporal domain. We will interchangeably use the term event for both a point \((s, t)\) of \( \mathcal{Y} \) and the actual event that took place in the application to give rise to a space-time measurement \((s, t)\). Hence, in the case of our
ambulance call data, an event is (the location and the time of) a call made to the dispatcher/emergency centre, which resulted in the dispatch of an ambulance. The events are indexed in accordance with their time of occurrence.

3.1. Intensity functions

Provided that it exists, the intensity function of an STPP governs its univariate properties and is defined as

\[ \lambda(s, t) = \lim_{|\delta s \times \delta t| \to 0} \frac{E[N(\delta s \times \delta t)]}{|\delta s \times \delta t|}, \quad (s, t) \in \mathbb{R}^2 \times \mathbb{R}, \]

where \( s \in \delta s \subseteq \mathbb{R}^2 \) and \( t \in \delta t \subseteq \mathbb{R} \), and \(|\delta s \times \delta t|\) denotes the size of \( \delta s \times \delta t \). Heuristically, since our STPPs can have at most one event at any location, we interpret \( \lambda(s, t) \) at the density of the probability that the STPP has an event in an infinitesimal neighbourhood of \((s, t)\). Whenever \( \lambda \) is a non-constant function we say that the STPP is inhomogeneous, and otherwise we say that it is homogeneous.

Similarly, we may study joint occurrences of points of the STPP by considering the intensity function \( \lambda^2 \) of the point process formed by distinct pairs \((s_i, t_i), (s_j, t_j)\) \( \subseteq (\mathbb{R}^2 \times \mathbb{R})^2 \), \( i \neq j \), of points of the STPP:

\[ \lambda^2((s, t), (s', t')) = \lim_{|\delta s' \times \delta t'| \to 0} \frac{E[N(\delta s' \times \delta t')]}{|\delta s' \times \delta t'|}, \quad (s, t), (s', t') \in \mathbb{R}^2 \times \mathbb{R}, \]

which is referred to as the second-order intensity/product density function. Here \(|\delta s \times \delta t|\) and \(|\delta s' \times \delta t'|\) denote the volumes of neighbourhoods \( \delta s \times \delta t \) and \( \delta s' \times \delta t' \) of the spatio-temporal locations \((s, t)\) and \((s', t')\), respectively. Note that \( \lambda^2((s, t), (s, t)) = 0 \) for any \((s, t)\) and \( \lambda^2((s, t), (s', t')) \) is interpreted as the density of the joint probability that the STPP has events in infinitesimal neighbourhoods of \((s, t)\) and \((s', t')\).

3.2. Poisson processes

A Poisson process, often playing the role of a benchmark which represents the case of no spatio-temporal interaction, is such that the event counts in disjoint subsets of the spatio-temporal domain are independent and the random number of events found in any subset of the spatio-temporal domain is Poisson distributed. If \( \mathcal{N} \) is a Poisson process with intensity function \( \mathcal{N} \), then \( \lambda(A) \) follows a Poisson distribution with mean \( \int_A \lambda(A) \, ds \, dt \) for any subset \( A \) of the spatio-temporal domain. Due to the independence property of a Poisson process, we have that its second-order intensity function satisfies \( \lambda^2((s, t), (s', t')) = \lambda(s, t) \lambda(s', t') \).

3.3. Intensity reweighted second-order moments

One can easily make the mistake of thinking that \( \lambda^2((s, t), (s', t')) \) being large means that there is (strong) dependence between points of the underlying STPP located around \((s, t)\) and \((s', t')\). For instance, in the case of a Poisson process, if \( \lambda(s, t) \) and \( \lambda(s', t') \) are/is large enough then we also obtain that \( \lambda^2((s, t), (s', t')) \) is large, despite the fact that there is independence between all points. Hence, to study actual dependence, we have to scale away the individual intensity contributions and this results in the pair correlation function \( \gamma(s, t) = \frac{g((s, t), (s', t'))}{\lambda(s, t) \lambda(s', t')} \) (Diggle, 2013; Gabriel and Diggle, 2009; González et al., 2016), which is one of the most commonly used summary statistics for studying second-order dependence structures. It is defined as

\[ g((s, t), (s', t')) = \lambda^2((s, t), (s', t')) \frac{1}{\lambda(s, t) \lambda(s', t')}, \quad (2) \]

and we see that for a Poisson process with intensity function \( \lambda \) we have that \( g((s, t), (s', t')) = 1 \). Hence, if the pair correlation function of an arbitrary STPP satisfies \( g((s, t), (s', t')) > 1 \) then there is clustering/aggregation between points located around \((s, t)\) and \((s', t')\). Reversely, \( g((s, t), (s', t')) < 1 \) instead indicates inhibition. Note that the so-called covariance density \( \gamma(s, t) \) is given by \( c((s, t), (s', t')) = \lambda^2((s, t), (s', t')) - \lambda(s, t) \lambda(s', t') = \lambda(s, t) \lambda(s', t') (g((s, t), (s', t')) - 1) \), which is 0 for a Poisson process.

It is further common to consider cumulative versions of pair correlation functions, so-called \( K \)-functions. Originally defined by Ripley for the spatial homogeneous case, it has been extended in a variety of directions. Baddeley et al. (2000) considered the spatial inhomogeneous case and this was extended to the spatio-temporal case by Gabriel and Diggle (2009), which in turn was further studied
by Møller and Ghorbani (2012); Gabriel (2014). Versions for marked (spatio-temporal) point processes have been treated by Cronie and van Lieshout (2016); Ifitimi et al. (2019). Inhomogeneous K-functions, both spatial and spatio-temporal, have been used in several application areas, see e.g. Diggle et al. (2007); Diggle and Gabriel (2010); Law et al. (2009); Arbia et al. (2012); Gabriel and Diggle (2009). Inhomogeneous K-functions are defined under the assumption of second-order intensity reweighted stationarity (SOIRS), which may defined as

$$g((s, t), (s', t')) = g(s - s', t - t'),$$

i.e. the pair correlation function only depends on the separation vectors, in combination with the intensity function being positive. Note that SOIRS is only well-defined for STPPs on $\mathbb{R}^2 \times \mathbb{R}$, i.e. not for STPPs on bounded regions, since SOIRS is based on Euclidean shifts and this is the reason why we consider restrictions of the kind (1) rather than defining our STPPs directly on $R \times [0, T]$. Gabriel and Diggle (2009) further defined SOIRS (with isotropy) as (additionally) having $g((s, t), (s', t')) = g(\|s - s'\|, |t - t'|)$, where $\| \cdot \|$ denotes the Euclidean norm, and considered the spatio-temporal K-function

$$K(r, t) = 2\pi \int_0^r u \int_0^t g(u, v) dv du,$$

where $r \geq 0$ and $t \geq 0$ are the spatial and temporal lags, respectively. We may think of (3) as reflecting the expected number of further points within spatial distance $r$ and temporal distance $t$ of an arbitrary point, having scaled for the varying intensity (if the intensity is high in a region then we scale down the contribution from that region). For an inhomogeneous spatio-temporal Poisson process with intensity function $\lambda(s, t)$ we have that $K(r, t) = 2\pi r^2 t$ and we may use this as a benchmark: if $K(b, d) - 2\pi b^2 d > 0$ we have indications of clustering for spatial lags less than $r$ and temporal lags less than $t$, and $K(b, d) - 2\pi b^2 d < 0$ instead indicates inhibition/regularity. We finally point out that there also exist higher order spatio-temporal summary statistics Cronie and Van Lieshout (2015) which may capture interactions beyond pairwise interactions, which is the case for K-functions.

3.4. Marginal point processes and separability

We say that an STPP is (first-order) separable if

$$\lambda(s, t) = \lambda_0(s)\lambda_1(t),$$

where it is customary to let either $\lambda_0$ or $\lambda_1$ be a density function. In other words, we assume that the univariate spatio-temporal properties of the STPP are obtained as a spatial scaling of the temporal univariate properties, or vice versa. Although this assumption mainly is suited for data which exhibit a seasonal behaviour (the spatial mass is not redistributed over time), it is often a needed pragmatic working assumption.

According to the definition of a point process, since we consider a restriction $g$ in a bounded domain $R \times [0, T]$ (as is always the case in practice), we may also consider the well-defined projected point processes $g_s = \{s_i\}_{i=1}^n \subseteq R \subseteq \mathbb{R}^2$ and $g_T = \{t_i\}_{i=1}^n \subseteq [0, T] \subseteq \mathbb{R}$. These point processes have intensity functions

$$\lambda_s(s) = \lim_{|\delta s| \to 0} \frac{\mathbb{E}[N(\delta s \times [0, T])]}{|\delta s|}, \quad s \in R, \quad \lambda_T(t) = \lim_{|\delta t| \to 0} \frac{\mathbb{E}[N(R \times \delta t)]}{|\delta t|}, \quad t \in [0, T],$$

and it is tempting to believe that they correspond to $\lambda_0$ and $\lambda_1$ when $g$ is separable. However, we have that (Møller and Ghorbani, 2012; González et al., 2016)

$$\lambda_s(s) = \int_0^T \lambda(s, t) dt = \lambda_0(s) \int_0^T \lambda_1(t) dt, \quad \lambda_T(t) = \int_0^T \lambda(s, t) ds = \lambda_1(t) \int_0^T \lambda_0(s) ds,$$

so only one of the integrals on the right hand sides will vanish if we assume that either $\lambda_0$ or $\lambda_1$ is a density function. E.g., if $\lambda_0(s)$ is a density on $R$, then

$$\lambda(s, t) = \lambda_0(s)\lambda_1(t) = \lambda_s(s)\lambda_T(t) \left( \frac{1}{\int_0^T \lambda_1(t) dt} \right)^1.$$
In this work, we assume that \( Z \) by-(\((\text{2005}) \), in the case of an LGCP we specifically have that \( Z \) and \( Z \) about spatial clustering of the events. The properties of \( \text{space–time Gaussian noise} \) that a \( \text{family of models to employ is spatio-temporal log-Gaussian Cox processes (LGCPs)} \) (Diggle, 2013). For instance, if \( Z \) \( Z \) such as weather and time of the day are possible drivers for the variations in the call risk), a reasonable \( \text{seem to come from some latent risk, which is the case with our ambulance call data (underlying factors} \) \( \text{first- and second-order intensity functions are given by} \) \( \text{LGCP} \). For \( \text{process in} \mathbb{R}^2 \) \( \text{distribution of} \), we may also consider the spatial point process time series \( \text{process with mean and the covariance functions} \). Defining \( \text{spatio-temporal LGCP is a spatio-temporal Cox process with the property that the logarithm of} \), its \( \text{first-} \) and \( \text{second-order intensity functions are given by} \) \( \text{of} \), \( \text{spatio-temporal Cox process with random intensity} \), \( \text{A spatio-temporal LGCP is a spatio-temporal Cox process with the property that the logarithm of} \), \( \text{family of models to employ is spatio-temporal log-Gaussian Cox processes (LGCPs)} \) (Diggle, 2013). Its \( \text{first- and second-order intensity functions are given by} \). Let \( \Lambda = \{ \Lambda(s, t) \in \mathbb{R}^+ \mid (s, t) \in \mathbb{R}^2 \times \mathbb{R} \} \) be a spatio-temporal random field, referred to as the random intensity process, such that (with probability one) \( \int_\mathbb{R} \Lambda(s, t)ds dt < \infty \) for any bounded subset \( \Lambda \) of \( \mathbb{R}^2 \times \mathbb{R} \). Its \( \text{its first- and second-order intensity functions are given by} \) \( \text{A spatio-temporal Cox process with random intensity} \) \( \text{generated by conditioning on} \) \( \text{and then, in turn, generating a Poisson with intensity function} \) \( \rho \). A spatio-temporal LGCP is a spatio-temporal Cox process with the property that the logarithm of its stochastic intensity function is a Gaussian process (Møller et al., 1998). Following e.g. Diggle et al. (2005), in the case of an LGCP we specifically have that

\[
\Lambda(s, t) = \lambda_0(s) \lambda_1(t) \exp [Z(s, t)],
\]

where \( Z = \{ Z(s, t) \in \mathbb{R} \mid (s, t) \in R \times [0, T] \} \) is an (unobservable) spatio-temporal Gaussian random field/process with mean and the covariance functions

\[
\mu(s, t) = \mathbb{E}[Z(s, t)] \quad \text{and} \quad \text{Cov}(Z(s_1, t_1), Z(s_2, t_2)) = \text{Cov}(Z(s_1, t_1), Z(s_2, t_2)).
\]

The distribution of \( Z \) is completely determined by the functions \( \mu(s, t) \) and \( \text{distribution with mean vector} \), \( \text{covariance matrix given by the} \) \( \text{the spatio-temporal random function} \) \( \text{represents unobservable space–time Gaussian noise that affects the occurrence of events and its inference provides information about spatial clustering of the events. The properties of} \), \( \text{of its corresponding LGCP. For instance, if} \). Defining \( \sigma^2(s, t) = \mathbb{V}(\text{ar}(Z(s, t))) = C((s, t), (s, t)) \), the first- and second-order intensity functions are given by (Møller et al., 1998)

\[
\lambda(s, t) = \lambda_0(s) \lambda_1(t) \mathbb{E}[\exp(Z(s, t))] = \lambda_0(s) \lambda_1(t) \exp[\mu(s, t) + \sigma^2(s, t)/2],
\]

\[
\lambda^2((s, t), (s', t')) = \lambda_0(s) \lambda_0(s') \lambda_1(t) \lambda_1(t') \mathbb{E}[\exp(Z(s, t) + Z(s', t'))] = \lambda_0(s) \lambda_0(s') \lambda_1(t) \lambda_1(t') \exp[\mu(s, t) + \mu(s', t') + \sigma^2(s, t)/2 + \sigma^2(s', t')/2 + C((s, t), (s', t'))].
\]

In this work, we assume that \( \mu(s, t) = -0.5 \sigma^2 \) and \( \sigma^2(s, t) = \sigma^2 > 0 \) so that \( \mathbb{E}[\exp(Z(s, t))] = 1 \) and we
note that the constant variance $\sigma^2$ determines the point-wise variability of $Z$ and it scales the log-intensity. This assumption implies that $\lambda(s, t) = \mathbb{E}[N(s, t)] = \lambda_0(s)\lambda_1(t)$, and we obtain that the LGCP is separable (recall Section 3.4). We further have that the deterministic temporal component $\lambda_1(t) \geq 0$ represents the expected number of events that occur in an infinitesimal neighbourhood of $t$ and when we consider a restriction of the type (1), i.e. when our LGCP is given on $R \times [0, T]$, we let the fixed spatial component $\lambda_0(s) \geq 0$ be a density, i.e. $\int_R \lambda_0(s) ds = 1$.

The above (separable) construction helps to provide separate spatial and temporal meaning to the expectation of the stochastic intensity. We further assume that $Z$ is stationary and isotropic, whereby the covariance function depends only on the separation lags between the evaluation points.

In addition, we assume a separable covariance structure, i.e. the covariance between two points in space-time can be factored into purely spatial and purely temporal components:

$$C((s_i, t_i), (s_j, t_j)) = C(s_i - s_j, t_f - t_i) = \sigma^2 \phi(||s_i - s_j||) \exp(r_0(||t_f - t_i||)),$$  

(7)

where $||\cdot||$ represents the Euclidean norm in $\mathbb{R}^2$, $|\cdot|$ denotes absolute value, and $\phi$, $\sigma^2$ and $\theta \in \mathbb{R}^+$ are fixed parameters (to be estimated). The scale parameters $\phi$ and $\theta$ control the rate of decay of the spatial and temporal correlation functions. We interpret temporal dependence as the covariance between the number of events in the study region at two points in time. Several forms for isotropic spatial correlation functions exist, e.g. the spatial correlation function in exponential and Matérn covariance functions.

Consulting the first- and second-order intensity functions in (6), we find that the pair correlation function and the $K$-function of the LGCP with driving covariance function (7) are given by

$$g(s_f - s_i, t_f - t_i) = \frac{\lambda^2((s_i, t_i), (s'_f, t'_f))}{\lambda(s_i, t_i)\lambda(s'_f, t'_f)} = \exp(C(s_f - s_i, t_f - t_i)) = \exp \left\{ \sigma^2 r_\phi (||s_f - s_i||) \exp (r_0 (||t_f - t_i||)) \right\},$$

$$K(r, t) = 2\pi \int_0^r u \int_0^u g(u, v) dvdu = 2\pi \int_0^r u \int_0^u \exp(\sigma^2 r_\phi(u)r_\theta(v)) dvdu;$$

the assumed LGCP is SOIRS, provided that the intensity is positive on $\mathbb{R}^2 \times \mathbb{R}$.

4. Statistical methodology

We next turn to the statistical inference based on $g = \{(s_i, t_i)\}_{i=1}^N \subseteq R \times [0, T]$, which is the restriction of the kind (1) of the LGCPs proposed in Section 3.5.

4.1. Spatial kernel intensity estimation

To non-parametrically deal with inhomogeneity of point patterns, there are at least two approaches that can be utilised. The first approach is to use methods of homogeneous point processes (constant intensity function) on subregions of the study region (Allard et al., 2001; Brix et al., 2001) while the second approach is to directly deal with the point patterns’ inhomogeneity by employing some functional approach which smooths over the whole domain (Cronie and Van Lieshout, 2018; Moradi et al., 2019; van Lieshout, 2012), e.g. kernel estimation.

Here we choose to use Quartic kernel smoothing methods to non-parametrically describe the purely spatial intensity function $\lambda_0(s)$, which generates a spatially smooth estimate of the local intensity of events. More precisely, we consider the kernel $\kappa(u) = (1 - u^2/2)I[u | u | \leq \sqrt{2}]$ and the estimator of the purely spatial intensity function is be given by

$$\hat{\lambda}_0(s) = \hat{\lambda}_0(s; h) = \frac{1}{h} \sum_{i=1}^n \kappa \left( \frac{s - s_i}{h} \right), \quad s \in R.$$

To incorporate edge correction into the spatial intensity estimate, the methods described in Diggle (1985) and Berman and Diggle (1989) may be applied. Note, however, that in the case of our ambulance data (recall Figure 1 and a map of Sweden and its border), edge effects may not actually be very present.

Although it is often claimed that the choice of kernel used is irrelevant (in comparison to the bandwidth choice), this is not entirely true (Cronie and Van Lieshout, 2018) and we note, in particular, that
kernels with bounded supports, such as the Quartic kernel, may generate large areas in the study region where the intensity estimate is 0, as opposed to using e.g. the Gaussian kernel which has unbounded support. The upside is that we, as in the case of our ambulance data, may better control where we place the mass of the estimated intensity. The issue that we are facing in the context of SOIRS models, however, is that our estimated intensity will not be positive everywhere, which is one of the conditions of SOIRS. Here we choose to ignore that latter fact and simply proceed with the Quartic kernel.

4.1.1. Bandwidth selection

As previously indicated, selection of an appropriate bandwidth, \( h \), which controls the smoothness of a kernel estimate, is the main challenge in kernel smoothing. In optimal bandwidth selection, we need to optimise the balance between under and over-smoothing of the kernel smoothing method. In the context of point processes, we are also faced with the additional challenge that the points of the underlying point process may be dependent. A range of different methods have been proposed in the literature (see e.g. Baddeley et al. (2015); Davies et al. (2018) for overviews), and most noteworthy are perhaps the recent method of Cronie and Van Lieshout (2018) and the Poisson processes likelihood cross-validation method are both based on integral expressions, where the integrals run over the whole spatial domain \( R \). Since essentially all of the ambulance calls have spatial locations which are close to the Swedish road network (see Figure 1), we have big holes in \( R \) where there are no events. This is exactly the scenario where the aforementioned methods tend to not work well (Cronie and Van Lieshout, 2018). The \( K \)-means clustering idea, on the other hand, operates on a more local scale and thus works better for our ambulance data. Note that the \( K \)-means clustering idea, similarly to the Poisson processes likelihood cross-validation method, does not fully compensate for dependence and, consequently, densely packed groups of events are interpreted as regions with high intensity function.

A few things need to be highlighted here. The Cronie and Van Lieshout (2018) method and the Poisson processes likelihood cross-validation method are both based on integral expressions, where the integrals run over the whole spatial domain \( R \). Since essentially all of the ambulance calls have spatial locations which are close to the Swedish road network (see Figure 1), we have big holes in \( R \) where there are no events. This is exactly the scenario where the aforementioned methods tend to not work well (Cronie and Van Lieshout, 2018). The \( K \)-means clustering idea, on the other hand, operates on a more local scale and thus works better for our ambulance data. Note that the \( K \)-means clustering idea, similarly to the Poisson processes likelihood cross-validation method, does not fully compensate for dependence and, consequently, densely packed groups of events are interpreted as regions with high intensity function.

### Algorithm 1. \( K \)-means clustering for bandwidth selection

1. Initialize \( K \) centroids of the clusters: \( \varpi_{k}^{(v)}, v = 0, k = 1, \ldots, K \).
2. Optimize 1-of-\( K \) class indicator variable for each observation \( i \):

\[
\begin{align*}
\iota_{ik}^{(v)} &= 1 \left\{ k = \arg \min_{q \in \{1, \ldots, K\}} \| s_i - \varpi_{q}^{(v)} \|^{2} \right\}, & i = 1, \ldots, n; k = 1, \ldots, K. \\
\end{align*}
\]

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

\[
\varpi_{k}^{(v+1)} = \frac{\sum_{i=1}^{n} \iota_{ik}^{(v)} s_i}{\sum_{i=1}^{n} \iota_{ik}^{(v)}}, & k = 1, \ldots, K.
\]

4. Repeat steps 2 and 3 for \( v = 1, 2, \ldots \) unless \( \sum_{k=1}^{K} \| \varpi_{k}^{(v+1)} - \varpi_{k}^{(v)} \|^{2} \leq 10^{-5} \).
5. Denote the obtained centroids of the clusters by \( \hat{\varpi}_{k}, k = 1, \ldots, K \).
6. Compute an optimal selection of the bandwidth:

\[
\hat{h} = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \| s_i - \hat{\varpi}_{k} \|^{2}.
\]
where \( t \) time mean and variance of \( N \), \( \lambda \) log given by be the expected number of emergency calls in expected number of calls over the study region can be understood from the regression model. Let \( N_{ff} \), number of events, i.e. the emergency alarm calls, over time. For instance, the effect of seasons on the expected number of calls over the study region can be understood from the regression model. Let \( N_{e}(t) \) be the expected number of emergency calls in \( R \) at time \( t \); under a Poisson distribution assumption, the mean and variance of \( N_{e}(t) \) are equal and here given by \( \lambda_{1}(t) \). The obtained Poisson regression model is given by

\[
\log \lambda_{1}(t) = \alpha_{0} + \sum_{i=1}^{7} \alpha_{i} 1(d(t) = i) + \sum_{j=1}^{4} \gamma_{j} 1(sn(t) = j) + \beta_{1} \sin(\tau t) + \beta_{2} \cos(\tau t) + \beta_{3} \sin(2\tau t) + \beta_{4} \cos(2\tau t) + \delta t,
\]

where \( \{\alpha_{0}, \alpha_{1}, \ldots, \alpha_{7}, \gamma_{1}, \ldots, \gamma_{4}, \beta_{1}, \ldots, \beta_{4}, \delta\} \) are the set of parameters, \( d(t) \) is the day of the week at time \( t \), \( sn(t) \) is the season of the year at time \( t \), \( \delta \) is the trend parameter, \( \tau = \frac{2\pi}{365} \), and \( 1(d(t) = i) \) and \( 1(sn(t) = i) \) are indicators of the day of the week and the season of the year at time \( t \). The days of a week, from Monday to Sunday, have been numbered ranging from 1 to 7 while the four seasons of a year, Spring, Summer, Fall, and Winter, have been numbered ranging from 1 to 4, respectively. The annual periodicity of the events can be taken into account in the model setting through the parameters \( \beta_{i}, \) \( i = 1, \ldots, 4 \).

4.2 Poisson regression based temporal intensity modelling

Recall from Section 3.5 that, by assumption, \( \int_{R} \lambda_{0}(s)ds = 1 \) so the temporal component \( \lambda_{1}(t) \) coincides with \( \lambda_{R}(t) \), which reflects the expected number of events in the spatial study region at time \( t \). Following Diggle et al. (2005), it seems to make sense to use a Poisson regression approach to model the temporal intensity component of the stochastic intensity function \( \Lambda(s, t) \). Day of the week, season of the year, and the periodic nature of the data can be used as covariate information for the modelling of the expected number of events, i.e. the emergency alarm calls, over time. For instance, the effect of seasons on the expected number of calls over the study region can be understood from the regression model. Let \( N_{e}(t) \) be the expected number of emergency calls in \( R \) at time \( t \); under a Poisson distribution assumption, the mean and variance of \( N_{e}(t) \) are equal and here given by \( \lambda_{1}(t) \). The obtained Poisson regression model is given by

\[
\log \lambda_{1}(t) = \alpha_{0} + \sum_{i=1}^{7} \alpha_{i} 1(d(t) = i) + \sum_{j=1}^{4} \gamma_{j} 1(sn(t) = j) + \beta_{1} \sin(\tau t) + \beta_{2} \cos(\tau t) + \beta_{3} \sin(2\tau t) + \beta_{4} \cos(2\tau t) + \delta t,
\]

where \( \{\alpha_{0}, \alpha_{1}, \ldots, \alpha_{7}, \gamma_{1}, \ldots, \gamma_{4}, \beta_{1}, \ldots, \beta_{4}, \delta\} \) are the set of parameters, \( d(t) \) is the day of the week at time \( t \), \( sn(t) \) is the season of the year at time \( t \), \( \delta \) is the trend parameter, \( \tau = \frac{2\pi}{365} \), and \( 1(d(t) = i) \) and \( 1(sn(t) = i) \) are indicators of the day of the week and the season of the year at time \( t \). The days of a week, from Monday to Sunday, have been numbered ranging from 1 to 7 while the four seasons of a year, Spring, Summer, Fall, and Winter, have been numbered ranging from 1 to 4, respectively. The annual periodicity of the events can be taken into account in the model setting through the parameters \( \beta_{i}, \) \( i = 1, \ldots, 4 \).

4.3 Non-parametric second-order summary statistic estimation

In the current context of LGCPs and their statistical inference, second-order summary statistics and their non-parametric estimators play a crucial role. Not only do they help in the non-parametric estimation of the underlying spatio-temporal dependence structures of the data, but they also play a crucial role in the minimum contrast-based parameter estimation for the latent Gaussian field \( Z \) in an LGCP’s random intensity function.

There are some issues that may arise in the estimation of spatio-temporal inhomogeneous \( K \)-functions or pair correlation functions. First off, they contain the true, unknown intensity function, but this we may replace by an intensity estimate (recall the previous sections). However, such a replacement should be done with caution, as it often adds bias to the estimation (Baddeley et al., 2000). This may for instance be the case when we estimate both a varying intensity function and a second-order characteristics using the same point pattern. We can overcome such problem by modelling the intensity using explanatory variables (Diggle et al., 2007; Mrkvička et al., 2014). Another way of dealing with the problem is to model the intensity parametrically as suggested by Arbia et al. (2012) and Diggle et al. (2007); e.g., Liu et al. (2007) and Moller and Waagepetersen (2003) have used parametric models to model intensity functions.

A further problem is related to edge effects, i.e. the influence of unobserved events outside the study region on the estimation of spatio-temporal interaction structures. In comparison to the purely spatial setting, edge effects are more difficult to overcome here where the dimension is greater than two (Baddeley, 1999). Various edge-correction methods have been proposed in the spatial setting (see e.g. Baddeley (1999); Cressie (1993); Diggle (2003); Illian et al. (2008); Law et al. (2009); Li and Zhang (2007); Ripley (1988); Goreaud and Pélissier (1999); Haase (1995); Pomerening and Stoyan (2006); Yamada and Rogerson (2003)) and in the spatio-temporal context, Gabriel (2014) has extended three classical spatial edge-correction factors to the spatio-temporal setting while Cronie and Särkkä (2011) proposed methods for parametric models.
Following Gabriel and Diggle (2009); Gabriel (2014) and assuming SOIRS with isotropy as in Gabriel and Diggle (2009), we employ the unbiased non-parametric estimator

$$ \hat{K}(r, t) = \frac{1}{T} \sum_{(s_1 \in \mathbb{R}^d) \cap [0, T]} \sum_{(s_2 \in \mathbb{R}^d) \cap [0, T]} \frac{1{[\|s_1 - s_2\| \leq r]}1{[\|t_1 - t_2\| \leq t]} \omega_R(s_1, s_2) \omega_f(t_1, t_2)}{\lambda(s_1, t_1) \lambda(s_2, t_2)} \quad (9) $$

of the $K$-function in equation (3). Here the notation $1{[\cdot]}$ is used for indicator functions, $\omega_R(s_1, s_2)$ is Ripley’s edge correction factor in $R$ (Ripley, 1977) and $\omega_f(t_1, t_2)$ is the one-dimensional analogue of Ripley’s edge correction factor in $[0, T]$.

Next, recall the point process time series $g(t), t = 1, \ldots, T$, in Section 3.4. Assuming separability and following Davies and Hazelton (2013), we may consider the purely spatial and temporally averaged non-parametric pair correlation function and $K$-function estimators

$$ \hat{g}(r) = \frac{1}{T} \sum_{t=1}^{T} \hat{g}(r) = \frac{1}{2\pi r T} \sum_{t=1}^{T} \frac{1}{\lambda_1^2(t)} \sum_{s_1 \in g(t)} \sum_{s_2 \in g(t)} \frac{\kappa_s(u - \|s_1 - s_2\|; h_s)}{\lambda_0(s_1) \lambda_0(s_2)} \omega_R(s_1, s_2), \quad (10) $$

$$ \hat{K}(r) = \frac{1}{T} \sum_{t=1}^{T} \hat{K}(r) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\lambda_1^2(t)} \sum_{s_1 \in g(t)} \sum_{s_2 \in g(t)} \frac{1{[\|s_1 - s_2\| \leq u]} \omega_R(s_1, s_2)}{\lambda_0(s_1) \lambda_0(s_2)}, $$

where $\omega_R$ is Ripley’s isotropic edge correction with respect to the region $R$, $\kappa_s$ is a univariate smoothing kernel with bandwidth $h_s$, and $|R|$ is the area of the spatial region $R$. Note that the above forms are derived from the fact that the expected number of events with (integer) event time $t$ is given by $\lambda_1(t) \int \lambda_0(s)ds$. For the bandwidth selection in the pair correlation function estimation in (10), we use Stoyan’s rule of thumb (Stoyan and Stoyan, 1994). Note that, in practice, in each of the estimators above we have to plug in an estimate of the intensity function.

4.4. Parameter estimation for the Gaussian random field parts

We next turn to the statistical inference for the whole LGCP model in Section 3.5. The key here is to note that we estimate $\lambda_0 = \{\lambda_0(s) \mid s \in R\}$ and $\lambda_1 = \{\lambda_1(t) \mid t \in [0, T]\}$ (recall Section 3.4) separately, non-parametrically and parametrically, respectively, so we are only left with estimating the covariance functions of the assumed LGCP, which are governed by the parameters $\sigma^2, \phi, \theta$. In other words, we are here only concerned with estimating the parameters of the latent Gaussian random field $Z$ in the formulation of the random intensity of the LGCP in Section 3.5.

It is common practice to exploit a regular grid discretisation $\{(s_i, t_j)\}_{i,j=1}^m \subseteq R \times [0, T]$ of the study region (Møller et al., 1998; Brix and Diggle, 2001; Diggle et al., 2005), which yields a discretised version $\mathbf{z} = \{(z(s_i, t_j))\}_{i,j=1}^m$ of the Gaussian process $Z$ over the grid (recall Section 3.5). Conditionally on $\mathbf{z}$, the number of events located within the grid cell centred at $(s_i, t_j)$ may be treated as an observation of a Poisson random variable $\Lambda(s_i, t_j)$. Assuming that the stochastic intensity function $\Lambda$ is constant over the cells (in practice this may be approximately be achieved by assuming that the cells are small), conditionally on the collection $\{(\Lambda(s_i, t_j))\}_{i,j=1}^m$, the collection $\{(\Lambda(s_i, t_j))\}_{i,j=1}^m$ follows a multivariate Poisson distribution. Note that vector-quantiles may be obtained by ordering the cells. The event times of our ambulance call data are recorded daily and we have thereby integer encoded them. Hence, we may consider three-dimensional time slices $x_t = \text{lexo}((x(s_i, t_j))_{i,j=1}^m), t \geq 1$, which are column vectors obtained by lexicographical ordering (lexo) of the cell-aggregated data at time $t$. Here the short notation $x_{1:t}$ represents the cell-counts at time slices $1, \ldots, t$. The notations $\mathbf{x}_t$ and $x_{1:t}$, which pertain to the discretisation of the random field $Z$, are interpreted in a similar manner as the notations $\mathbf{x}$ and $x_{1:T}$.

The natural starting-point for parameter estimation is maximum likelihood estimation. However, in general, the likelihood of an LGCP is analytically intractable since it is expressed as a high-dimensional integral with respect to the distribution of the unobserved Gaussian process $Z$. Under the current parametrization, the likelihood of the LGCP can be expressed as

$$ \mathcal{L}(\sigma^2, \theta, \phi, \lambda_0, \lambda_1 \mid y) = \mathbb{E}_\Lambda[ \mathcal{L}(\sigma^2, \theta, \phi, \lambda_0, \lambda_1 \mid y, \Lambda) ] = \mathbb{E}_\Lambda \left[ \exp \left\{ \int_0^T \int_w (1 - \Lambda(s, t)) ds dt \right\} \prod_{i,j=1}^m \Lambda(s_i, t_j) \right], $$

10
where \( \mathbf{y} = \{(s_i, t_i)\}_{i=1}^{m} \subseteq \mathbb{R} \times [0, T] \) is an observed spatio-temporal point pattern. In principle, importance sampling can be used to carry out maximum likelihood estimation of LGCP (Geyer, 1994; Møller et al., 1998). However, in order to use importance sampling for maximum likelihood estimation of the LGCP, we need to sample from the conditional distribution of the whole of \( \mathbf{z}_{1:m} \), given all the observations \( \mathbf{x}_{1:m} \), and this is unfortunately infeasible in big data settings such as the current one (Brix and Diggle, 2001).

Instead, the common approach is to carry out minimum contrast estimation, based on summary statistics. Here, the essential idea is to minimise a functional distance between the parametric form of a functional summary statistic, which has been derived from the assumed model (here the LGCP-structure in Section 3.5), and a non-parametric estimate of the summary statistic in question. Within the current set-up, the spatial and temporal components are estimated separately.

#### 4.4.1. Minimum contrast estimation of the spatial components

We start by dealing with \((\sigma^2, \phi)\), i.e. with the variance and the parameter(s) of the spatial part of the separable covariance function (7). More specifically, they are estimated according to

\[
(\hat{\sigma}^2, \hat{\phi}) = \arg \min_{\sigma^2, \phi} \left\{ \int_{r_{\text{min}}}^{r_{\text{max}}} \theta(r) \left( F \left[ S(r) - F \left[ S \left( r, \sigma^2, \phi \right) \right] \right] \right)^2 dr \right\},
\]

\[
\approx \arg \min_{\sigma^2, \phi} \left\{ \frac{1}{\Delta r} \sum_{r \in U_j} \theta(r) \left( F \left[ \hat{S}(r) - F \left[ S \left( r, \sigma^2, \phi \right) \right] \right] \right)^2 \right\},
\]

where \( U_j \) constitutes a fine and evenly spaced sequence \( 0 \leq r_{\text{min}} < r_1 < \cdots < r_{\text{max}} \) with \( \Delta r = r_{i+1} - r_i, \theta(\cdot) \) is an optional lag-dependent weight, \( S \) is the model-based parametric form of the summary statistic used, \( \hat{S} \) is its non-parametrically estimated counterpart and \( F \left[ \cdot \right] \) is a transformation applied to the functional summary statistics. For the summary statistics we may here use the non-parametric estimators in (10) and their theoretical LGCP-based counterparts (recall Section 3.5 and Section 3.4):

\[
g_t(s_r - s_i) = \exp \left\{ \sigma^2 r \phi (\|s_r - s_i\|) \right\},
\]

\[
K_t(r) = 2\pi \int_0^r u g_0(u) du = 2\pi \int_0^r u \exp(\sigma^2 r_\phi(u)) du;
\]

note that these are constant in \( t \). In the case of our ambulance data, we specifically employ an exponential correlation function \( r_\phi(u) = \exp \{-u/\phi\} \). Moreover, we have here chosen to use the pair correlation for our minimum contrast estimation.

#### 4.4.2. Minimum contrast estimation of the temporal components

The parameter \( \theta \) describes temporal dependence between different (spatial point process) components of the series \( y(t), t = 1, \ldots, T \). More specifically, it determines the covariance between the number of events in the spatial domain \( R \) at two points in time. Recalling the count function \( N_t(B), B \subseteq R, \) of \( y(t), t = 1, \ldots, T \) (see Section 3.4), by Diggle et al. (2005); Davies and Hazelton (2013),

\[
C(t; v; \theta, \sigma^2, \phi) = \text{Cov}(N_t(R), N_{t-v}(R)),
\]

\[
= \lambda_1(t) \lambda_1(t-v) \left( \int_R \int_R \lambda_0(s_1) \lambda_0(s_2) \exp \left\{ \sigma^2 r_\phi(u) r_\phi(v) \right\} ds_1 ds_2 - 1 \right), \quad v > 0,
\]

and an empirical auto-covariance function is a natural non-parametric estimator for the temporal covariance (12) and it is given by

\[
\hat{C}(t, v) = N_t(R) N_{t-v}(R) - \lambda_1(t) \lambda_1(t-v).
\]

We may now obtain a minimum contrast estimate of \( \theta \) by finding

\[
\hat{\theta} = \arg \min_{\theta} \left\{ \sum_{v=1}^{\text{max}} \sum_{t=v+1}^{\text{max}} \left[ \hat{C}(t, v) - C(t; v; \theta, \sigma^2, \phi) \right]^2 \right\}.
\]
where \( v \) is a temporal lag and \( v_{\text{max}} \) is a maximum temporal lag. We refer to Davies and Hazelton (2013) for a complete overview of a minimum contrast estimation of covariance function parameters in spatial and spatio-temporal LGCPs. Hence, using expressions (11) and (13), the parameters of the covariance function in equation (7) may be estimated. The centroids of a regular grid cells of the discretised study region and the estimated parameters can be exploited to approximate the covariance matrix \( \Sigma \) by computing its elements using equation (7).

By definition, the discretisation \( z \) of the unobservable Gaussian process \( Z \) on the study region follows a multivariate Gaussian distribution. It follows that simulation can be used to do inference for the unobservable random vector \( z \), by sampling from the conditional distribution of the unobservable random vector given its corresponding observable random vector (Brix and Diggle, 2001). Letting [\( \cdot \mid \cdot \)] denote the probability distribution of a random quantity, the conditional distribution of the unobserved time series of random vectors \( z_{1:T} \), given its corresponding observed time series of random vectors \( x_{1:T} \), satisfies

\[
[z_{1:T} \mid x_{1:T}] \sim [x_{1:T} \mid z_{1:T}].
\]

Note that here \( [x_{1:T} \mid z_{1:T}] \) essentially describes the distribution of a (discretised) Poisson process, given that its intensity is obtained through a realisation of a (discretised) random field, and \( [z_{1:T}] \) represents the unconditional distribution of the (discretised) field. Hence, to infer about \( z \), we need to be able to sample \( z_{1:T} \) from the conditional distribution. We may approximate the right hand side of (14) by \( [x_{1:T} \mid x_{1:T}] [z_{1:T}] \), where \( T - v \) is a small integer, since remote past observations tend to have a negligible effect on the current state. The distribution \( [z_{1:T}] \) can be evaluated using the estimates of the parameters \( \phi^2, \phi, \text{ and } \theta \) obtained through equations (11) and (13).

Finally, recall that in the case of the ambulance data we assume that the spatial correlation function is given by the exponential form \( r_{\phi}(v) = \exp \{-v/\phi\} \). We here also assume that the temporal correlation function has an exponential form, i.e. \( r_{\sigma}(v) = \exp \{-v/\theta\} \). Note that the scale parameters controls the smoothness of the underlying (marginal) fields and indicate the predictability, i.e. essentially how close to each other two events have to be "to influence each other".

In the Supplementary material we provide details on the Fast Fourier transform for covariance matrix computation and the Metropolis-adjusted Langevin algorithm for simulation.

### 4.4.3. Forecasting

In this work, the emergency alarm call locations at the last time point have been reserved and the emergency alarm call locations at the last time point have been forecasted to visually evaluate the practicability of the spatio-temporal log-Gaussian Cox process model. Furthermore, we are aimed at forecasting the emergency alarm call locations beyond the last time point \( T \) of data observation.

Using the last time point \( T \), the random intensity function in equation (5) can be rewritten as

\[
\Lambda(s, T + \Delta) = \lambda_0(s)\lambda_1(T + \Delta) \exp \{Z(s, T + \Delta)\},
\]

where \( \Delta \) is the number of time units (days in the case of the ambulance data) beyond the last time point \( T \). Let \( q = MN \), where \( M \) and \( N \) are the number of rows and columns of a discretisation of the spatial domain; see expression (3) in the Supplementary material. On the extended lattice (with dimension \( q \times q \)), the time series \( z^{xt} \), \( t = 1, \ldots, T \), is assumed to be a stationary Gauss–Markov process. It follows that this is a discrete-time sample of an Ornstein–Uhlenbeck process (Brix and Diggle, 2001), which solves the stochastic differential equation

\[
dz^{xt} = \theta (\mu - z^{xt}) dt + \sigma \, d\epsilon_t,
\]

where \( \theta \) is \( q \times q \) invertible real matrix, \( \mu \) is \( q \)-dimensional long-run mean, \( \sigma \) is \( q \times q \) positive semidefinite matrix, and \( \epsilon_t, t \geq 0 \), is a \( q \)-dimensional Wiener process (Jacobsen et al., 1993). The diffusion coefficient \( \sigma \) measures the size of the noise, \( \mu \) is the asymptotic mean level for the process and \( \theta \) determines how fast the process reacts to perturbations. Solving (16) and using the stationary Gauss–Markov process property of the process, we have that

\[
z^{xt}_{T+\Delta} | z^{xt}_T \sim N(\varphi(\Delta)z^{xt}_T + (1 - \varphi(\Delta))\mu, \int_T^{T+\Delta} \varphi(T + \Delta - u)\sigma \sigma' \varphi(T + \Delta - u) \, du),
\]
where $\varphi(\Delta) = \exp(-\theta\Delta)$ and $I$ is the $q \times q$ identity matrix. Given the observed data $x_{r:t}$, the forecast distribution of $z_{T+\Delta}$ can be formulated as

$$[z_{T+\Delta}^\text{ext} \mid x_{r:t}] = \int [z_{T+\Delta}^\text{ext} \mid z_T^\text{ext}] [z_T^\text{ext} \mid x_{r:t}] dz_T^\text{ext},$$

and based on equation (17), the expected value of the forecast distribution $z_{T+\Delta}^\text{ext} \mid x_{r:t}$ can be obtained as follows:

$$E[z_{T+\Delta}^\text{ext} \mid x_{r:t}] = E[E[z_{T+\Delta}^\text{ext} \mid z_T^\text{ext}] \mid x_{r:t}],
= \varphi(\Delta) E[z_T^\text{ext} \mid x_{r:T}] + (1 - \varphi(\Delta)) \mu, \quad (18)$$

Using the expected values in equations (17) and (18), the variance of the forecast distribution $z_{T+\Delta}^\text{ext} \mid x_{r:T}$ may be obtained as

$$\text{Var}(z_{T+\Delta}^\text{ext} \mid x_{r:T}) = \varphi(\Delta) \text{Var}(z_T^\text{ext} \mid x_{r:T}) \varphi(\Delta) + \int_{T+\Delta}^{T+\Delta} \varphi(T + \Delta - u) \sigma u \sigma u' \varphi(T + \Delta - u') du. \quad (19)$$

A simplifying assumption should be used to estimate the forecast distribution. One of the key assumptions is to assume the parameter matrices $\theta$ and $\sigma$ to be $I$ and $L r$. Taylor et al. (2013) have implemented the estimation of the forecast distribution under these simplified assumptions. Following this assumption, equations (17), (18), and (19) can be simplified and Monte Carlo estimation can be used to estimate the forecast expectation in equation (18). Let $z_T^\text{ext}(i)$ denote the $i^{th}$ sample obtained during the simulation of the Gaussian random field, which can be taken as a draw from $[z_T^\text{ext} \mid x_{r:T}]$ once the chain has reached stationarity. Using equation (17), we can then generate a draw $z_{T+\Delta}^\text{ext}(i)$ from $[z_{T+\Delta}^\text{ext} \mid z_T^\text{ext}]$ given the sample $z_T^\text{ext}(i)$. Considering a burn-in time $b$, the mean of the samples $z_{T+\Delta}^\text{ext}(i), i = b, b + 1, \ldots, q$, is an unbiased estimator of the expectation $E[z_{T+\Delta}^\text{ext} \mid x_{r:T}]$. Once we estimate the expected value in equation (18) and predict the expected number of emergency alarm calls by the estimated Poisson regression model, equation (15) can be computed to simulate spatio-temporal inhomogeneous Poisson process.

### 5. Spatio-temporal ambulance call data analysis

We have finally reached the point where we apply the above LGCP framework to the spatio-temporal ambulance call data in Section 2. Recall that the main task here is to simulate realistic future emergency alarm call locations, given a future time point (here, a day). A large part of the analysis has been carried out using the R package lgecp (Taylor et al., 2013).

#### 5.1. Non-parametric estimation of the spatial component of the intensity

The first thing that we do is to obtain a non-parametric estimate of the spatial intensity component $\lambda_0(s), s \in R$, which we do in accordance with Section 4.1. More specifically, we employ the Quartic kernel estimator (8), where we select the bandwidth using the $K$-means method outlined in Section 4.1.1; the outcome is visually illustrated in Figure 2. To select the number of clusters, $K$, to use, we have turned to visual inspection; we found that estimates with $K > 5$ tended to accentuate regions which (visually) have a low spatial intensity too much. In addition, $K = 5$ was also the choice which gave rise to the simulated forecasted point patterns that most resembled the true data (see the part on forecasting below).

#### 5.2. Non-parametric spatio-temporal interaction analysis

Before choosing a particular model to fit to a given point pattern, it is paramount to carry out a non-parametric analysis of spatio-temporal interaction, to better understand the data and to see what type of model may be appropriate for the data. Here the idea is to quantify the interaction (clustering/aggregation or inhibition/regularity) present at/within different inter-point distances. As previously indicated, we here employ the inhomogeneous $K$-function estimator $\hat{K}(r, t)$ in (9), which is compared with the theoretical $K$-function $K(t, r) = 2\pi \sigma^2 r$ obtained for an inhomogeneous spatio-temporal Poisson process (with any intensity function), to see whether there is excess spatio-temporal interaction in the underlying STPP. Recall that we plug an intensity estimate $\hat{\lambda}(s, t)$ into expression (9), which we here let be given by a
product of the spatial intensity estimate obtained in Section 5.1 and a temporal kernel estimate with an Epanechnikov kernel. Given the scale of the data, the effects of the miss-specification that our data are discrete in time while (9) is, in fact, defined for temporally continuous data are negligible. Figure 2 shows a plot of $\hat{K}(r,t) - 2\pi r^2 t$, where all values have been scaled by the maximum value of the estimator $\hat{K}(r,t)$ in order to obtain values between 0 and 1. We see that there is light aggregation present which, we argue, advocates for employing an LGCP-model (a special case of an LGCP is a Poisson process). There is a caveat, however, which is that since we assume first-order separability, $\hat{K}(r,t)$ reflects both spatio-temporal aggregation and non-separable first-order effects. It should further be noted that there may be stronger temporal interactions present, which operate on time lags smaller than one time unit, i.e. one day.

![Spatial kernel intensity estimation plot](image1)
![Inhomogeneous K-function estimation plot](image2)

Fig. 2. Spatial kernel intensity estimation plot (left) and inhomogeneous $K$-function estimation plot (right) for the ambulance call data. In the right plot, the x-axis label denotes spatial lags in meters while the y-axis label represents temporal lags in days.

5.3. Estimation of the temporal intensity component

We next turn to modelling the temporal evolution of the calls, which is done in accordance with Section 4.2. To fit the Poisson regression model for the temporal component $\lambda_1(t)$, we use the iteratively reweighted least squares method and the obtained results can be found in Table 1. We see that our model is neither overestimating nor underestimating as the median deviance residual is close to zero, see Table 1. By looking at the residual deviance we see that the inclusion of covariates in the Poisson regression model generates a better fit as compared to how well we would have performed with a model that only includes the intercept (reflected by the null deviance). This suggests that the fitted model is indeed appropriate. The table also reflects that the daily expected number of calls has a strong relationship with the covariate day-of-the-week and season-of-the-year, e.g. summer and fall. This may indicate that summer and fall, as we have defined them here, stand out while the rest of a year has a similar behaviour.

A plot of the fitted Poisson regression model and the observed number of emergency alarm calls are shown in Figure 3. We can here visually verify that the fitted model has captured the pattern in the observed data quite well. The significance of summer and fall can be visually detected by the small offsets observed in the harmonic behaviour in Figure 3.
Table 1. Estimated Poisson regression model for the temporal component of the spatio-temporal intensity function.

| Variables      | estimate | std.error ($\times 10^{-3}$) | z-value | p (> |z|) |
|----------------|----------|-------------------------------|---------|-------|------|
| Monday         | 5.433    | 5.840                         | 930.261 | 0.000 |
| Tuesday        | 5.396    | 5.875                         | 918.443 | 0.000 |
| Wednesday      | 5.398    | 5.867                         | 920.032 | 0.000 |
| Thursday       | 5.403    | 5.862                         | 921.831 | 0.000 |
| Friday         | 5.447    | 5.818                         | 936.104 | 0.000 |
| Saturday       | 5.491    | 5.743                         | 956.086 | 0.000 |
| Sunday         | 5.442    | 5.813                         | 936.217 | 0.000 |
| Summer         | 0.021    | 9.522                         | 2.191   | 0.028 |
| Fall           | 0.019    | 7.939                         | 2.349   | 0.019 |
| Cos(wt)        | 0.059    | 4.663                         | 12.550  | 0.000 |
| Sin(wt)        | 0.035    | 3.920                         | 8.852   | 0.000 |
| Cos(2wt)       | 0.024    | 3.160                         | 7.681   | 0.000 |
| Sin(2wt)       | 0.019    | 2.614                         | 7.357   | 0.000 |
| time of events | 0.000    | 0.003                         | 20.404  | 0.000 |

### Computed value | Degrees of freedom
--- | ---
Median deviance residual & -0.023 & 1826
Null deviance & 4002192 & 1812
Residual deviance & 3282 & 1812
Akaike information criterion & 16689 &

Fig. 3. A plot of the successive weekly averages of the observed (solid dots) and the fitted (solid line) number of ambulance calls.

5.4. Covariance fitting, forecasting and simulation of future ambulance call locations

One of the main aims of forecasting is to obtain a dynamical model which stepwisely takes in new data and outputs updated forecast, i.e. on-line spatio-temporal risk-mapping. Hence, we here leave out the data at the last time point in the fitting of the model (training data) and then generate the forecast model for that time point to simulate realistic “future” spatial event locations.

By employing the minimum contrast approach outlined in Section 4.4, which deals with the estimation of the variance parameter $\sigma^2$, the spatial correlation function $r_s(u) = \exp(-u/\phi)$ and the temporal correlation function $r_t(v) = \exp(-v/\theta)$, we obtain the estimates $(\hat{\sigma^2}, \hat{\phi}, \hat{\theta}) = (4.933, 3494.705, 0.182)$. Hence, given the current separable setting, we have indications that there are intraday dependencies at play, given the fitted spatial dependence structure (recall from (13) that the temporal covariance component is fitted conditionally on the estimated spatial one), and the variance parameter estimate indicates
that we have more structure than in a Poisson process. At the same time, the fairly large spatial scale parameter estimate indicates that there is quite weak dependence in space.

Next, the fitted/trained model has been used to forecast the Gaussian random field to obtain forecast intensities, which have been exploited to simulate Poisson process realisations. Figure 4 illustrates the observed ambulance call locations (spatial test data) together with a simulated realisation, i.e. simulated call locations. Visually, it seems that the forecasted model quite well picks up on many of the different hot-spots found in the actual data. It does, however, not fully capture a few of the different regions where there are just a few calls, which most likely is an effect of the spatial kernel intensity estimate (see Figure 2). Also, the counts in the simulated realisations are slightly smaller than the what we see in the true data, which may be due to the discretisation used for the random field. On the other hand, it seems to catch the different hot-spots quite well. Taking things a step further, Figure 5 shows simulated realisations beyond the last time point, for the days January 01 (a Tuesday) until January 07 (a Monday), 2019. We see that the spatial event distribution tends to change over time.

Fig. 4. Observed ambulance call location point pattern (left) and a simulated realisation of its forecasted counterpart (right).
6. Discussion

The main focus of this paper is to study whether spatio-temporal log-Gaussian Cox process (LGCP) models can successfully model ambulance call data and, in addition, whether the associated spatio-temporal forecasts have the ability of generating realistic simulated future ambulance call scenarios. This is relevant e.g. to design optimal ambulance dispatching rules/strategies. Our rather unique spatio-temporal point pattern dataset under consideration consists of 444,283 ambulance/emergency alarm calls, with associated dispatch addresses and event day recordings, in the four northernmost regions of Sweden over the years 2014 to 2018.

In accordance with e.g. Diggle et al. (2005), we consider an LGCP which consists of three different components: a spatial component, \( \lambda_0(s) \), a temporal component, \( \lambda_1(t) \), and a separable log-Gaussian random intensity field, \( \exp(Z(s,t)) \). The former two are purely deterministic components which are designed to take purely spatial and purely temporal variation of the events into account. The stochastic component \( \exp(Z(s,t)) \), on the other hand, models dependence and spatio-temporal variation of the events. Moreover, computational merit can be obtained using separable covariance structure and exponential form of correlation function for the temporal component (Diggle et al., 2005). In addition, assuming an exponential covariance function for the temporal part of the Gaussian random field \( Z \) makes it Markovian in time.

The parameters of the spatial and temporal parts of the covariance function of \( Z \) have been estimated using minimum contrast estimation (Møller et al., 1998; Davies and Hazelton, 2013), whereas Quartic kernel smoothing has been utilised to estimate \( \lambda_0(s) \) for the ambulance data. The bandwidth in the Quartic kernel smoothing has been selected using \( K \)-means clustering, which requires selecting the number of clusters to be used during the estimation of the bandwidth. The proposed \( K \)-means approach better managed to capture the high concentration of events along the relatively highly populated east coast and the main roads of the region, than did many of the existing (adaptive) intensity estimation approaches (Davies et al., 2018; Moradi et al., 2019; Ogata et al., 2003; Ogata, 2004). The number of clusters has been selected by visually evaluating whether the spatial intensity generated by the Quartic kernel smoothing method shows the spatial variation of the ambulance calls. A Poisson regression model, which incorporates different calendar covariates such as day-of-the-week and season-of-the-year, has been adopted to model the temporal variation of the calls and the fitting of the parameters has been carried out using
the iteratively reweighted least squares method. We find that the fit is good, and a few covariates are significantly governing the behaviour of the model. An estimate of the inhomogeneous spatio-temporal $K$-function of Gabriel and Diggle (2009) has been used to verify that an LGCP is indeed a sensible choice of model – there seems to be low/moderate spatio-temporal clustering present in the ambulance data but we hypothesise that there likely is within-day-dependence present (which we cannot quantify here since the temporal sampling of our data is per day). Regarding the computational aspects, a two-dimensional discrete Fourier transform and a fast-Fourier transform have been adopted to reduce the computational cost in simulating the Gaussian random field $Z$. In addition, the Gaussian random field has been simulated using the Metropolis-adjusted Langevin algorithm. The fitted LGCP model can be utilised to obtain a forecast distribution of the Gaussian random field and, consequently, a Poisson intensity from which we can simulate realistic future event locations. To carry out the forecasting, we have excluded the last time point in the data and carried out forecasting of the spatial structure for this time point. We find that the patterns of the observed and the simulated spatial locations are very similar. However, the number of spatial locations in the simulations are slightly smaller, which may be due to the discretisation used for the random field.

The core challenges here stem from the nature of the current set of spatio-temporal (big) data. First of all, the geographic region is vast and, consequently, the changes in the data vary substantially over the spatial region. This may partly be remedied by splitting up the data set into samples over smaller sub-regions; this will, in effect, have the consequence that the final model will have a spatially varying covariance structure (which seems quite reasonable, given the scope of the data). Additionally, since most of the events are located in regions with a higher population density, having access to relevant spatial, e.g. demographic, covariates (something we are currently working on obtaining) may also improve the fit substantially. A further idea, which likely requires a substantial amount of methodological development, is to transition to linear network point processes (Ang et al., 2012; Baddeley et al., 2015, 2020; Moradi and Mateu, 2019).

In conclusion, this preliminary study has shown that spatio-temporal log-Gaussian Cox processes are practically viable for the spatio-temporal modelling and forecasting of ambulance calls. Simulations from such forecasts may, in turn, be exploited in optimising prehospital care resources, e.g. by designing ambulance dispatching strategies in an optimal way.

Acknowledgments

This work was supported by Vinnova [grant number 2018-00422] and the regions Västerbotten, Norrbotten, Västernorrland and Jämtland-Härjedalen. We are grateful to our project partners SOS Alarm and the regions Västerbotten, Norrbotten, Västernorrland and Jämtland-Härjedalen for fruitful discussions on the Swedish prehospital care.

Disclosure of Conflicts of Interest

The authors have no relevant conflicts of interest to disclose.

References

Allard, D., Brix, A., and Chadoeuf, J. (2001). Testing local independence between two point processes. *Biometrics*, 57(2):508–517.

Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and computing*, 18(4):343–373.

Ang, Q. W., Baddeley, A., and Nair, G. (2012). Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scandinavian Journal of Statistics*, 39(4):591–617.

Arbia, G., Espa, G., Giuliani, D., and Mazzitelli, A. (2012). Clusters of firms in an inhomogeneous space: The high-tech industries in Milan. *Economic Modelling*, 29(1):3–11.

Aringhieri, R., Bruni, M. E., Khodaparasti, S., and van Essen, J. T. (2017). Emergency medical services and beyond: Addressing new challenges through a wide literature review. *Computers & Operations Research*, 78:349–368.

Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G., and Davies, T. M. (2020). Analysing point patterns on networks - a review. *Spatial Statistics*.

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

Baddeley, A. J. (1999). Spatial sampling and censoring. *Stochastic geometry: likelihood and computation*, 2:37–78.

Baddeley, A. J., Møller, J., and Waagepetersen, R. (2000). Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350.
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.

O’keeffe, C., Nicholl, J., Turner, J., and Goodacre, S. (2011). Role of ambulance response times in the survival of patients with out-of-hospital cardiac arrest. *Emergency medicine journal*, 28(8):703–706.

Pell, J. P., Sirel, J. M., Marsden, A. K., Ford, I., and Cobbe, S. M. (2001). Effect of reducing ambulance response times on deaths from out of hospital cardiac arrest: cohort study. *Bmj*, 322(7299):1385–1388.

Pommerening, A. and Stoyan, D. (2006). Edge-correction needs in estimating indices of spatial forest structure. *Canadian Journal of Forest Research*, 36(7):1723–1739.

Ripley, B. (1988). *Statistical inference for spatial processes*. Cambridge University Press, Cambridge.

Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(2):172–192.

Roberts, G. O. and Rosenthal, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268.

Roberts, G. O., Tweedie, R. L., et al. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363.

Stoyan, D. and Stoyan, H. (1994). *Fractals, random shapes, and point fields: methods of geometrical statistics*, volume 302. John Wiley & Sons Inc.

Taylor, B. M., Davies, T. M., Rowlingson, B. S., Diggle, P. J., et al. (2013). lgcp: an r package for inference with spatial and spatio-temporal log-gaussian cox processes. *Journal of Statistical Software*, 52(4):1–40.

van Lieshout, M.-C. N. (2012). On estimation of the intensity function of a point process. *Methodology and Computing in Applied Probability*, 14(3):567–578.

Wood, A. T. and Chan, G. (1994). Simulation of stationary Gaussian processes in [0, 1] d. *Journal of computational and graphical statistics*, 3(4):409–432.

Yamada, I. and Rogerson, P. A. (2003). An empirical comparison of edge effect correction methods applied to K-function analysis. *Geographical Analysis*, 35(2):97–109.

Zhou, Z., Matteson, D. S., et al. (2016). Predicting melbourne ambulance demand using kernel warping. *The Annals of Applied Statistics*, 10(4):1977–1996.

Zhou, Z., Matteson, D. S., Woodard, D. B., Henderson, S. G., and Micheas, A. C. (2015). A spatio-temporal point process model for ambulance demand. *Journal of the American Statistical Association*, 110(509):6–15.
Supplementary material

Fast Fourier transform for covariance matrix computation

To implement methods for log-Gaussian Cox processes, it is necessary to discretise the spatial study region $R$. The most common such discretisation strategy is to construct an approximation grid over an encapsulating rectangle $W$ of the study region since $R$ can be an irregular region in $\mathbb{R}^2$. Here, the cell-centroids are given by

$$C = \left\{(s^x_{\min} + (i - 0.5) \Delta_x, s^y_{\min} + (j - 0.5) \Delta_y) \mid i = 1, 2, \cdots, m; j = 1, 2, \cdots, p \right\},$$

(1)

where $\Delta_x$ is the maximum range spanning $R$ along the x-axis divided by the number of disjoint cells $m$ along the x-axis, $\Delta_y$ is the maximum range spanning $R$ along the y-axis divided by the number of disjoint cells $p$ on the y-axes, $s^x_{\min}$ and $s^y_{\min}$ are the minimum x-coordinate and y-coordinate of $W$ and the index $(i, j)$ represents the cell-centroids of the regular spatial grid over $W$. For any given time, the spatially continuous Gaussian process $Z$ can be approximated by a collection of random variables at cell-centroids of the regular spatial grid over $W$ covering the observation window $R$. This collection of random variables can be ordered lexicographically to obtain a long column vector of random variables $z = \text{lexo}(\{z_{i,j}\})_{i=1,m}^{p}$.

In this way, the simulation of the spatially continuous Gaussian process $Z$ can be approximately translated to the simulation of a high-dimensional Gaussian random vector $z$. Under this assumption, the stationary Gaussian random field $z$ has mean $\mu$ and covariance structure $\Sigma$ with dimensions $mp \times 1$ and $mp \times mp$, respectively. The $(i, j)^{th}$ entry of the covariance matrix $\Sigma$ may be obtained as follows:

$$\Sigma_{ij} = \sigma^2 r_\phi \left( d_{ij} \right),$$

(2)

where $i, j \in I = \{1, 2, \cdots, mp\}$, $d_{ij} = ||C_{z(i)} - C_{z(j)}||$, and $r_\phi$ is the spatial correlation function with parameter $\phi$. The quantities $C_{z(i)}$ and $C_{z(j)} \in C$ are the cell-centroids corresponding to the $i^{th}$ and $j^{th}$ element in the lexicographic ordering. A massive computation is required to simulate the Gaussian random field directly even for a coarse regular grid approximation of the study area. Given the discretization, the size of the covariance structure grows dramatically and it is necessary to consider an eigendecomposition of the covariance structure to simulate a correlated multivariate Gaussian random field. A method using a two-dimensional discrete Fourier transform and a fast-Fourier transform to reduce the computational cost in simulating a Gaussian random field has been proposed (Møller et al., 1998; Wood and Chan, 1994). A fast-Fourier transform is a quick evaluation algorithm that helps to obtain the eigenvalues of block circulant matrices at a dramatically reduced computational cost. In its current form, the covariance structure $\Sigma$ is not a block circulant matrix. However, a fast-Fourier transform requires the matrix to be block circulant. Therefore, circulant embedding has to be performed using an extended encapsulating rectangular grid, which must be at least twice the original size along both axes of the study region. The cell-centroids of the extended encapsulating regular spatial grid of the study region can be given by

$$C^{ext} = \left\{(s^x_{\min} + (i - 0.5) \Delta_x, s^y_{\min} + (j - 0.5) \Delta_y) \mid i = 1, 2, \cdots, M; j = 1, 2, \cdots, N \right\},$$

(3)

where $M \geq 2m$ and $N \geq 2p$. The number of equally-spaced disjoint cells of the extended encapsulating rectangle of the study region is $MN$. The extended cell-centroid based Euclidean distance definition must correspond to wrapping the extended lattice on a torus and then computing the shortest Euclidean distances to satisfy the block circulant requirement of the covariance structure. Let $(u, v) = C^{ext}_{x(u,i)}$ be an index in $C^{ext}$ corresponding to the $i^{th}$ entry in $z^{ext}$, which is a column vector obtained by a lexicographic ordering of the Gaussian random fields at the cell-centroids in $C^{ext}$. Denote the shortest distance on the torus between $(u_1, v_1) = C^{ext}_{x(u,i)}$ and $(u_2, v_2) = C^{ext}_{x(u,i)}$ by

$$d \left( (u_1, v_1), (u_2, v_2) \right) = \sqrt{d_{u_1u_2}^2 + d_{v_1v_2}^2},$$

where
where \( d_{u_1u_2} = \min\{|u_1 - u_2|, R_x - |u_1 - u_2|\} \) and \( d_{v_1v_2} = \min\{|v_1 - v_2|, R_y - |v_1 - v_2|\} \). The ranges in the x- and y-directions \( R_x = \text{range(\text{Index}^x)} \) and \( R_y = \text{range(\text{Index}^y)} \) are obtained from the extended index set
\[
\text{Index} = \left\{ (s^x_{\text{min}} + i\Delta_x, s^y_{\text{min}} + j\Delta_y) \mid i = 0,1,2,\cdots M; j = 0,1,2,\cdots N \right\}.
\]
Accordingly, the \((i,j)^{th}\) entry of the covariance matrix \( \Sigma \) can be obtained as follows:
\[
\Sigma^{\text{ext}}_{ij} = \sigma^2 I_y (d ((u_1, v_1), (u_2, v_2))).
\tag{4}
\]
where \( i, j \in I = \{1,2,\cdots,MN\} \). The covariance structure obtained from equation (4) is a block circulant matrix and therefore, its spectral decomposition can be computed efficiently using a fast-Fourier transform. Let \( \hat{F}_M \) represents a normalised \( M \times M \) discrete Fourier transform matrix with entries given by
\[
\hat{F}_L (i,j) = \frac{\exp \{-2\pi ij \sqrt{-1}/M\}}{\sqrt{M}},
\]
where \( i, j \in \{0,1,\cdots,M-1\} \). The block circulant matrix property of \( \Sigma^{\text{ext}} \) helps to drive its eigenvalues from an \( M \times N \) matrix \( \Sigma_x \), which is obtained only from the first row of \( \Sigma^{\text{ext}} \). The eigenvalues corresponding to \( \Sigma^{\text{ext}} \) can be quickly computed as follows:
\[
\Omega = \sqrt{MN} \hat{F}_M \Sigma_x \hat{F}_N,
\]
where \( \hat{F}_M \) and \( \hat{F}_N \) are \( M \times M \) and \( N \times N \) complex conjugates of \( F_M \) and \( F_N \), and \( \Sigma_x \) is an \( M \times N \) matrix obtained only from the first row of \( \Sigma^{\text{ext}} \). On the extended lattice, compute an \( M \times N \) matrix \( \Gamma \) as follows:
\[
\Gamma = \hat{F}_M \left[ [F_M \Sigma_x F_N] \odot \Omega^{1/2} \right] F_N,
\]
where \( U \) is an \( M \times N \) matrix containing independent standard normal variables, the symbol \( \odot \) denotes the Hadamard product, and \( \Omega^{1/2} \) represents element-wise power of the eigenvalues in eigenvalue matrix \( \Omega \) to the scalar 1/2. Using a lexicographically ordered form \( \Gamma^{\text{ext}} \) of \( \Gamma \), a realisation of the Gaussian random field \( z^{\text{ext}} \) on the extended lattice can be obtained by
\[
z^{\text{ext}} = \Gamma^{\text{ext}} + \mu^{\text{ext}},
\]
Since the square root of the eigenvalues in \( \Omega^{1/2} \) is required, the successfullness of the discrete Fourier transform depends on the positive semidefiniteness of \( \Sigma^{\text{ext}} \). In practice, there is no guarantee for the positive semidefiniteness of \( \Sigma^{\text{ext}} \). However, this problem occurs rarely provided that suitable values of \( M \) and \( N \) are used (Møller et al., 1998). In terms of computational speed and efficiency, the fast-Fourier transform algorithms perform their best when the grid cell resolutions \( M \) and \( N \) are factorised as powers of two (Møller et al., 1998). To obtain the realised Gaussian random field on the study region, we simply discard the elements of \( z^{\text{ext}} \) that correspond to the centroids falling outside the study region \( R \).

The Metropolis-adjusted Langevin algorithm

The Metropolis-adjusted Langevin algorithm has been suggested by Besag (1994) and it was later studied in detail by Roberts et al. (1996). The Metropolis-adjusted Langevin algorithm is a Metropolis-Hasting type of Markov chain Monte Carlo algorithm and it differs from Metropolis-Hasting algorithm in that its proposal distribution for each transition exploits gradient information about the target distribution. At each iteration of Metropolis-adjusted Langevin algorithm, multiple spatial intensities are considered in the spatio-temporal log-Gaussian Cox process due to the effect of the temporal correlation on the spatial variation. Monte Carlo simulation from (14) can be made faster by working with a linear transformation of \( z \). As was stated earlier, fast-Fourier transform algorithms perform their best when the extended grid cell resolution are \( M = 2m \) and \( N = 2p \), where \( m \) and \( p \) are some positive integer powers of two. Let \( D \) be the rank of the covariance matrix \( \Sigma^{\text{ext}} \). Set
\[
z_i^{\text{ext}} = A\gamma_i + \nu^{\text{ext}},
\]
where $\gamma_t$ follows a $D$-dimensional standard multivariate normal distribution at time $t$, $A$ is the matrix that diagonalizes the covariance matrix $\Sigma^{st}$, and the restriction of $\nu^{st}$ to $C$ agrees with the mean of $z_t$.

Our goal is to simulate ambulance call locations at the last time point $T$ in the data, given the observed data up to that time. In comparison to $[z_t^{st} | x_t^{st}]$, it is easier to simulate from $[z_t^{st} | x_t^{st}]$ as the distribution is known up to a proportionality constant (Brix and Diggle, 2001). In big data settings, such as in our case, computational efficiency problems can be encountered when simulating $T$ Gaussian random fields at each iteration of the chain. This problem can be minimised by considering only data corresponding to the most recent times (Brix and Diggle, 2001). Let $\zeta = T - \nu + 1$ be a reasonably small positive integer. To overcome the computational efficiency problem, the spatial data at $\zeta$ time points can be utilised to simulate Gaussian random fields from the distribution $[Z_{\nu:T} | x_{\nu:T}]$. The strength of the estimated temporal correlation parameter $\theta$ can be used to suggest the number of time points $\zeta$ to be exploited in the simulation. Under the assumption that $[Z_t^{st} : t \in \{\nu, \cdots, T\}]$ is a Markov process and $x_t^{st}$ is conditionally independent of $x_t^{ext}$ and $z_t^{st}$ given $Z_{\nu:t}$ for all $t \neq \nu$, the target density using $\gamma_t$ can be given by

$$
\begin{bmatrix} \gamma_{\nu:T} | x_{\nu:T}^{st} \end{bmatrix} \propto \begin{bmatrix} x_{\nu:T}^{st} | \gamma_{\nu:T} \end{bmatrix} = \prod_{t=\nu}^{T} \begin{bmatrix} x_t^{st} | x_t^{ext} \end{bmatrix} \begin{bmatrix} \gamma_t \end{bmatrix} \prod_{t=\nu+1}^{T} \begin{bmatrix} \gamma_t | \gamma_{t-1} \end{bmatrix}.
$$

Since $x_t$ is a Gaussian random field, the probability distributions $[\gamma_t]$ and $[\gamma_t | \gamma_{t-1}]$ are Gaussian for each time $t$. The logarithm of the target density $\begin{bmatrix} \gamma_{\nu:T} | x_{\nu:T}^{st} \end{bmatrix}$ can be expressed explicitly using equation (5.5). Taking the gradient information of the target density into account during simulation of the Gaussian random field can assist in directing the proposal candidates toward dense regions of the target density. Besides, it can help to achieve superior convergence rates of the Markov chains (Møller et al., 1998). In the Metropolis-Hastings algorithm, Roberts et al. (1996) suggested a gradient information based proposal density given by

$$
\begin{bmatrix} \gamma_{\nu:T} | \gamma_{\nu:T}^{(k)} \end{bmatrix} = \mathcal{N} \left( \cdot | \gamma_{\nu:T}^{(k)} + 0.5\xi^2 \nabla \log \begin{bmatrix} \gamma_{\nu:T}^{(k)} | x_{\nu:T}^{st} \end{bmatrix}, \xi^2 I \right),
$$

where $\mathcal{N} (\cdot)$ represents a Gaussian distribution, $\xi^2$ denotes the variance of a mutually independent multivariate Gaussian distribution, $\nabla \log \begin{bmatrix} \gamma_{\nu:T}^{(k)} | x_{\nu:T}^{st} \end{bmatrix}$ is a $D \times \zeta$ dimensional gradient of the logarithm of the target density, and $\gamma_{\nu:T}^{(k)}$ is the current state of the chain. A realisation of $\gamma_{\nu:T}$ can be generated from the Gaussian distribution in equation (6) to exploit it as a proposal candidate in gradient information based Metropolis-Hastings algorithm. The Metropolis-adjusted Langevin algorithm can be summarized as follows:

Algorithm 2. The Metropolis-adjusted Langevin algorithm

1. Select the current state of the chain $\gamma_{\nu:T}^{(k)}$, $k = 0,$
2. Generate a proposal candidate from the proposal distribution:

$$
\gamma' \sim \begin{bmatrix} \gamma_{\nu:T} | \gamma_{\nu:T}^{(k)} \end{bmatrix} = \mathcal{N} \left( \cdot | \gamma_{\nu:T}^{(k)} + 0.5\xi^2 \nabla \log \begin{bmatrix} \gamma_{\nu:T}^{(k)} | x_{\nu:T}^{st} \end{bmatrix}, \xi^2 I \right),
$$

3. Compute the quantity:

$$
\alpha (\gamma', \gamma_{\nu:T}^{(k)}) = \min \left\{ 1, \frac{\begin{bmatrix} \gamma_{\nu:T}^{(k)} | x_{\nu:T}^{st} \end{bmatrix} \begin{bmatrix} \gamma_{\nu:T}^{(k)} | \gamma_{\nu:T}^{(k)} \end{bmatrix}}{\begin{bmatrix} \gamma_{\nu:T} | x_{\nu:T}^{st} \end{bmatrix} \begin{bmatrix} \gamma_{\nu:T} | \gamma_{\nu:T}^{(k)} \end{bmatrix}} \right\},
$$

4. Draw $\ell \sim U [0, 1]$, a uniform distribution on the interval $[0, 1]$,
5. Accept or reject the proposal candidate according to:

$$
\gamma_{\nu:T}^{(k+1)} = \begin{cases} 
\gamma' & \text{if } \ell \leq \alpha (\gamma', \gamma_{\nu:T}^{(k)}), \\
\gamma_{\nu:T}^{(k)} & \text{otherwise},
\end{cases}
$$

6. Repeat steps 2 and 5 until the desired length of the chain is reached.

It is not easy to set the tuning parameter $\xi^2$ in the Metropolis-adjusted Langevin algorithm. There is
evidence how to choose the variance $\xi^2$ of the proposal distribution to obtain the fastest rate of convergence and an overall acceptance rate of around 0.574 (Roberts and Rosenthal, 1998). By examining some shorter preliminary runs of the algorithm, we can set the value of the tuning parameter. An algorithm of Andrieu and Thoms (2008) can be exploited for automatic choice of $\xi^2$ so that the acceptance rate 0.574 can be achieved without disturbing the ergodic property of the chain.