Received: 11 May 2019 | Revised: 10 March 2020 | Accepted: 29 April 2020

DOI: 10.1111/sjos.12480

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Semiparametric estimation with spatially correlated recurrent events

Akim Adekpedjou[1] | **Sophie Dabo-Niang**[2]

[1]Department of Mathematics and Statistics, Missouri University of Science and Technology

[2]Laboratoire Paul Painvelé UMR CNRS 8524, INRIA-MODAL, University of Lille

Correspondence
Akim Adekpedjou, Department of Mathematics and Statistics, Missouri University of Science and Technology, Rolla, MO 65409.
Email: akima@mst.edu

Abstract

This article pertains to the analysis of recurrent event data in the presence of spatial correlation. Consider units located at n possibly spatially correlated geographical areas described by their longitude and latitude and monitored for the occurrence of an event that can recur. We propose a new class of semiparametric models for recurrent events that simultaneously account for risk factors and correlation among the spatial locations, and that subsumes the current methods. Since the parameters involved in the models are not directly estimable because of the high dimension of the likelihood, we use composite likelihood approach for estimation. The approach leads to estimates with population interpretation where their large sample properties are obtained under a reasonable set of regularity conditions. Simulation studies suggest that the resulting estimators have a very good finite sampling properties. The methods are illustrated using spatial data on recurrent esophageal cancer in the northern region of France and recurrent wildfire data in the province of Alberta, Canada.

KEYWORDS

composite likelihood, estimating functions, Gaussian random fields, increasing domain asymptotic, recurrent events, spatial correlation

1 INTRODUCTION

Recurrent event data are encountered in various settings such as biomedical, reliability, actuarial science, sociology, and public health to name a few (Cook & Lawless, 2002, 2007; Therneau & Hamilton, 1997) and they are part of a class of data called survival or failure time data. Let S_{ij} be the time of the jth event of unit i, and T_{ij} the gap between the jth and $(j-1)$th events. Interest, in general, lies in estimating the distribution function $F_{ij}(t)=P(T_{ij} \leq t)$, with probability density function $f_{ij}(t)$ or the hazard function $\lambda_{ij}(t) = f_{ij}(t)[1 - F_{ij}(t)]^{-1}$. Parametric, semiparametric, nonparametric, as well as regression-type models for F_{ij} have been the subject of intense research in the past decades. Some assumed renewal-type processes, wherein the gap time are independent. Models for correlated gap time within a unit have also been proposed, mostly modeled using frailty (Murphy, 1995; Parner, 1998; Vaupel, 1990). Under those two settings, some references are Therneau and Hamilton (1997), Hougaard (2000), Therneau and Grambsch (2000), Peña, Strawderman, and Hollander (2001), Huang and Wang (2004), Peña, Slate, and Gonzalez (2007), Liu, Lu, and Zhang (2014), and Adekpedjou and Stocker (2015) and references therein. In those manuscripts, the units are assumed independent, and the impact of potential environmental factors contributing to onset of event and future recurrences are not modeled. Such environmental factors in a given location may affect nearby locations thereby inducing the so-called *spatial correlation*, which is a correlation between the geographical locations of two units. To motivate the problem of interest in the single event, here is an example of Asthma data where onset of the disease on children is subject to spatial correlation.

Example: *East Boston Asthma Study: cf. Li and Ryan* (2002).

A total of 753 subjects are enrolled in a Community Health Clinic in the east Boston area. Questionnaire data pertaining to residential addresses, demographic variables, asthma status, geographic coordinates, and other environmental factors were collected during regularly scheduled visits. Geocoding the data set allows linkage with various community-level covariates to individuals in the east Boston data from U.S. census data at the census block level. Because children residing in nearby census blocks were often exposed to unmeasured similar physical and social environments, the investigators suspected there might exists spatial correlation across different communities. The goal of the study was to identify significant risk factors associated with age at asthma onset while accounting for the possible spatial correlation among the locations.

In the above example, residents of east Boston are mainly relatively low income with similar social and economical backgrounds and often exposed to similar physical and social environments. Although the children are independent, their different geographical locations in east Boston are spatially correlated since adjacent neighborhoods usually have a lot in common and the potential for spatial dependence exists. Hence, correct inference on the association of the main covariates with the event-specific survival times relies on careful consideration of the underlying spatial correlation. Therefore, estimation of the distribution of the time of onset of an event requires a combination of modern survival analysis and geostatistics techniques in order to identify risk factors at onset and further recurrences of the event.

To that end, consider n units located at areas described by their longitude and latitude in a two-dimensional surface, are monitored for the occurrence of some event such as onset of disease, epidemic, claims filed as a result of property losses, cancer, migration of individuals from one area to another to seek better living conditions. There exist nuisance parameters such as environmental factors, social and physical environments, population density, or weather conditions out of control of the investigators that can have substantial impact on the occurrence of events between two areas via their spatial coordinates.

This article is concerned with the development of models for the distribution function of event time subject to recurrence. Data on event time with spatial impact can be encountered in many areas such as epidemiology, e-commerce, clinical trials, politics, ecology, and population migration (Bronnenberg, 2005; Engen, 2007; Hunt, 1978; Paik & Ying, 2012). As indicated earlier, of interest is estimator of the distribution function of the time of onset and further recurrences. This is important in the sense that if the spatial impact leads to drastic consequences, local authorities could take necessary preventive actions to reduce damage. It is therefore of considerable importance to develop models for estimating the distribution function of time to event while accounting for spatial correlation. The topic has emerged as an area of active research, especially in the case of no recurrence. The models of interest are part of a multivariate survival models containing a parameter modeling the association between event times T_i and T_j, $i \neq j$ of two independent units. Such models include bivariate frailty, copulas, marginal models, cluster models, and spatial correlation-type models via the covariation process using a martingale representation. Similarities and differences between copulas and frailties are discussed in Goethals, Janssen, and Duchateau (2008). Interest in spatial correlation modeling dates back to the pioneering work of Krige, and recently Matheron (1962). Works in the single event can be found in Li and Ryan (2002), Henderson, Shimakura, and Gorst (2002), Banerjee, Wall, and Carlin (2003), Banerjee and Dey (2005), Li and Lin (2006), Diva, Banerjee, and Dey (2007), Diva, Dey, and Banerjee (2008), Paik and Ying (2012), and Pan, Cai, Wang, and Lin (2014) and references therein.

Frailty, cluster, marginal, and copula models do not properly account for spatial correlation that is inherent with these data. As a consequence, sophisticated techniques of geostatistics, coupled with modern failure time modeling are needed. In recognition of that, Li and Lin (2006), in the single event, assumed a Cox model for failure time and used a probit-type transformation of the failure times yielding a multivariate Gaussian random field. Furthermore, they imposed a spatial structure on the associated random fields that properly captures the spatial patterns among regions. Likelihood construction was then facilitated using composite likelihood. In the recurrent event arena, Scheike, Eriksson, and Tribler (2019) proposes a nonparametric estimator of the correlation of the mean number of events for two recurrent event processes. Lee and Cook (2019) uses a Gaussian copula to model association between type-specific random effects. Their models have marginal features and composite likelihood was used to deal with high-dimensional integral. Since recurrent event data are an unbalanced data, which is the number of recurrences varies from one unit to another, Barthel, Geerdens, Czado, and Janssen (2019) uses an ordered D-Vine copula to better capture model association. Golzy and Carter (2019) proposes a general class of models based on a nonlinear mixed effects to model association between gap time with recurrent event data.

Although some works in the modeling of association between failure time in the recurrent setting exist, the models do not properly embed spatial patterns in their approaches. The aims of this article are to develop statistical models for spatially correlated recurrent events where regression parameters have a region- and/or area-level interpretation and in which spatial correlation is properly incorporated. Ideas in the work of Li and Lin (2006) in the single event are borrowed by transforming the set of gap times using a probit-type transformation allowing the vector of gap times to follow a Gaussian process. The data sets that motivate this work are the recurrent esophageal cancer in various spatial locations in France Northern Region called Nord-Pas de

Calais (NPC hereon), and a data set on recurrent wildfire in the province of Alberta, Canada. A description of both data is given in section 5.

The layout of the article is as follows. In Section 2, we introduce the spatially correlated recurrent event data and discuss the choice of the spatial correlation model. Section 3 deals with the joint modeling of recurrent events in two regions and models estimation, whereas Section 4 is devoted to the large sample properties. The last section pertains to the simulation studies and an illustration of our models in the biomedical and environmental fields. A brief discussion and technical proofs close the article.

2 | SPATIALLY CORRELATED RECURRENT EVENTS

The first critical step in the modeling is to identify a suitable dependence model between spatial locations. As noted earlier, we will focus on geostatistical formulation that relies on the fitting of covariance and cross-covariance structures for Gaussian random fields for mathematical and computational convenience.

2.1 | Pairwise recurrent event data

Consider n units located at n different spatial locations (areas in a town, counties, zip codes) described by a two-dimensional coordinates $\{\mathbf{l}_i = (l_{i,1}, l_{i,2}) : i = 1, \dots, n\}$ where $l_{i,1}$ and $l_{i,2}$ represent longitude and latitude of the ith location, respectively. The geographical locations are subject to spatial correlation. Suppose further that the units are monitored for the occurrences of an event that can recur up to a random time τ_i. For unit i, let $S_{i,j}$ be the time of occurrence of the jth event, and $T_{i,j}$ the gap between the $(j-1)$ and jth occurrences. For every unit in location $\mathbf{l}_i$, a p-dimensional covariates $\mathbf{x}_i$ is recorded. We assume that the $T_{i,j}$s have a distribution function $F_i(\cdot)$ that depends on the covariates. Location $\mathbf{l}_i$ is assumed to be spatially correlated with $\mathbf{l}_j$, $i \neq j$ and denote the spatial correlation between the two by $\rho_{ij} := \rho(\|\mathbf{l}_i - \mathbf{l}_j\|)$, where $\|\mathbf{l}_i - \mathbf{l}_j\|$ is the Euclidean distance between $\mathbf{l}_i$ and $\mathbf{l}_j$. If K_i is the random number of occurrences by τ_i, the observables for unit i at spatial location $\mathbf{l}_i$ at the conclusion of monitoring are

$$\mathcal{O}(\mathrm{I}_i) = (K_i, \tau_i, T_{i,1}, \dots, T_{i,K_i}, \tau_i - S_{i,K_i}, \mathbf{x}_i). \tag{1}$$

Since we want to account for spatial correlation between pair of locations, the observables will therefore be considered pairwise. Consequently, the spatially correlated recurrent events is given by

$$\mathcal{O} = \{\{[\mathcal{O}(\mathrm{I}_i), \mathcal{O}(\mathrm{I}_j)], \rho_{ij}\} : (i,j), i \neq j, i = 1, \dots, n\}\}. \tag{2}$$

2.2 | Gaussian random fields with recurrent events

Gaussian random fields (GRF) and their multivariate counterpart (MGRF) have a dominant role in spatial modeling, especially in geostatistics. We will not dwell on the technicalities of GRF, but refer the reader to standard stochastic processes textbooks. Here, normalizing the interevent times will lead to the construction of MGRF. The motivation behind this approach is threefold: (i) the marginal distribution of the interevent times will follow a model that accounts for covariates, (ii) prediction of event occurrences at new sites is faster with GRF using existing software packages and kriging techniques, and (iii) the approach facilitates construction of composite likelihood process, estimation of parameters, as well as large sample properties.

To that end, let $N_i^{\dagger}(s)$ be the number of events up to time s and $\mathcal{E}_i(s)$ the effective age of a unit i in location $\mathbf{l}_i$. $\mathcal{E}_i(s)$ models the impact of interventions performed to restore a unit back to operation after event occurrences. It can take various forms, such as minimal intervention or repair (when $\mathcal{E}_i(s) = s$, no improvement in future lifetime), perfect intervention (when $\mathcal{E}_i(s) = s - S_{N_i^{\dagger}(s)}$, this case reverts the effective age of the unit to a new one), and an effective age between the minimal and perfect, called imperfect intervention. We assume throughout the article that $\mathcal{E}_i(s) = s - S_{N_i^{\dagger}(s)}$, a perfect repair. Let $\lambda_i(s|\mathbf{x}_i) = \lambda_0(\mathcal{E}_i(s))\exp(\boldsymbol{\beta}'\mathbf{x}_i)$ be the hazard function at time s, a Cox-type regression model that accounts for covariates, with $\mathbf{a}'$ denoting the transpose of a column vector $\mathbf{a}$. If $\Lambda_i(s|\mathbf{x}_i)$ is the true marginal cumulative hazard function of T_{ij}, then $\Lambda_i(T_{ij})$ follows an EXP(1) distribution. Consequently, with $\Phi(\cdot)$ being the cumulative distribution function of N(0,1), we have

$$T_{i,j}^{\dagger} := \Phi^{-1}[1 - \exp\{-\Lambda_i(T_{i,j})\}] \sim \mathrm{N}(0,1). \tag{3}$$

At each location, given $K_i = k_i$, define the vector $\mathbf{T}_i^{\dagger} = (T_{i,1}^{\dagger}, \dots, T_{i,k_i}^{\dagger})$, with $T_{i,j}^{\dagger}$ independent. Thus, $\mathbf{T}_i^{\dagger}$ follows a multivariate normal (MVN) distribution, and the hazard function of $T_{i,j}^{\dagger}$ follows a Cox model marginally.

With a view toward the choice of the spatial correlation function, given $K_i = k_i$, let $\mathbf{T}_i^{\dagger} = (T_{i,1}^{\dagger}, \dots, T_{i,K_i}^{\dagger}|K_i = k_i)$ be the resulting multivariate vector of gap times. Then $\mathbf{T}_i^{\dagger} = (T_{i,1}^{\dagger}, \dots, T_{i,k_i}^{\dagger})$ is a k_i-dimensional MVN and it follows from the logit transformation that $\mathbf{T}_i^{\dagger} \sim \mathrm{MVN}_{k_i}(\mathbf{0}_{k_i}, \boldsymbol{\Sigma}_i^{\dagger})$, where $\boldsymbol{\Sigma}_i^{\dagger} = \mathrm{Var}(\mathbf{T}_i^{\dagger}) = \sigma^2 \mathrm{I}_{k_i}$, I_{k_i} is the $k_i \times k_i$ identity matrix, $\sigma^2 > 0$. Let $\boldsymbol{\Sigma}_{ii}^{\dagger} = \boldsymbol{\Sigma}_i^{\dagger}$ and $\boldsymbol{\Sigma}_{ij}^{\dagger} = \mathrm{cov}(\mathbf{T}_i^{\dagger}, \mathbf{T}_j^{\dagger})$, $i \neq j$. Expression of $\boldsymbol{\Sigma}_{ij}^{\dagger}$ is constructed via simple stationary isotropic covariance function. That is, the covariance function $\mathrm{cov}(\mathbf{T}_i^{\dagger}, \mathbf{T}_j^{\dagger})$, depends only on the separation spatial vector $\mathbf{l}_i - \mathbf{l}_j$ and some unknown parameters. Specifically, let $\boldsymbol{\Sigma}_{ij}^{\dagger} = \mathrm{cov}(\mathbf{T}_i^{\dagger}, \mathbf{T}_j^{\dagger}) = \{C_{uv}(\mathbf{l}_i, \mathbf{l}_j)\}_{u,v}$ where $C_{uv}(\mathbf{l}_i, \mathbf{l}_j) = \mathrm{cov}(T_i, u_{\dagger}, T_j, v_{\dagger})$, $C_{uu}(\mathbf{l}_i, \mathbf{l}_j)$ is the univariate covariance function while $C_{uv}(\mathbf{l}_i, \mathbf{l}_j) = 0$, $u \neq v$ is the cross-covariance function between gap times. The condition $C_{uv}(\mathbf{l}_i, \mathbf{l}_j) = 0$ means that for two individuals i and j, only the gap times of the same rank are dependent. However, this condition may be restrictive otherwise and falls under the umbrella of space time-dependent gap time, which may be considered in future research. Under the assumption, we have

$$\boldsymbol{\Sigma}_{ii}^{\dagger} = \begin{pmatrix} C_{11}(\mathbf{l}_i, \mathbf{l}_i) & \dots & C_{1k_i}(\mathbf{l}_i, \mathbf{l}_i) \\ \vdots & \ddots & \vdots \\ C_{k_i1}(\mathbf{l}_i, \mathbf{l}_i) & \dots & C_{k_ik_i}(\mathbf{l}_i, \mathbf{l}_i) \end{pmatrix} := \boldsymbol{\Sigma}_i^{\dagger}, \quad \boldsymbol{\Sigma}_{ij}^{\dagger} = \begin{pmatrix} C_{11}(\mathbf{l}_i, \mathbf{l}_j) & \dots & C_{1k_j}(\mathbf{l}_i, \mathbf{l}_j) \\ \vdots & \ddots & \vdots \\ C_{k_i1}(\mathbf{l}_i, \mathbf{l}_j) & \dots & C_{k_ik_j}(\mathbf{l}_i, \mathbf{l}_j) \end{pmatrix}. \tag{4}$$

For instance, if $k_i = 3$ and $k_j = 5$, one obtains,

$$\boldsymbol{\Sigma}_i^{\dagger} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_j^{\dagger} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_{ij}^{\dagger} = \begin{pmatrix} 0.5 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0 \end{pmatrix}. \tag{5}$$

2.3 | Spatial correlation model

As indicated earlier, the critical part in identifying significant risk factors that trigger an initial occurrence and future recurrences is to identify the best spatial correlation function. In the single-event setting, copula models have been proposed by Lawless and Yilmaz (2011) and Yilmaz and Lawless (2011) and frailty models by Banerjee et al. (2003). A pairwise joint distribution that depends on the distance between locations has been investigated by Paik and Ying (2012). Diva et al. (2007) extended the work of Banerjee and Carlin (2003) by generalizing their multivariate conditional autoregressive models. Henderson et al. (2002) proposed a multivariate gamma frailty model incorporating spatial dependence between locations. The problem with the use of frailty or copula is that the former models within cluster correlation using frailties or random effects, and the latter or models joint distribution of two failure times, and consequently do not really model spatial correlation. We seek a spatial correlation that is a function of distance between spatial locations, so called *isotropic spatial covariance* functions. They have received a great deal of attention recently, specifically the Matérn family (Gneiting, Kleiber, & Schlather, 2010; Guttorp & Gneiting, 2006; Matérn, 1986) given by

$$C(\mathbf{h}) = \sigma^2 M(\mathbf{h}; \nu, a) = \sigma^2 \left(\frac{2^{1-\nu}}{\Gamma(\nu)} (a\|\mathbf{h}\|)^{\nu} K_{\nu}(a\|\mathbf{h}\|) \right), \tag{6}$$

where σ^2 is the marginal variance or *sill*, which is the variance if $\|\mathbf{h}\| = \|\mathbf{l}_i - \mathbf{l}_j\| = 0$. $\nu > 0$ is a smoothing parameter that controls the differentiability of a Gaussian process with this covariance; $a > 0$ is a *range* parameter that measures the correlation decay as the separation between two sites increases. $K_\nu(\cdot)$ and $\Gamma(\cdot)$ are the Bessel and gamma functions, respectively. When $\nu = 0.5$ and $+\infty$, we recover the exponential and Gaussian covariance given by

$$C(\mathbf{h}) = \sigma^2 \exp(-a\|\mathbf{h}\|), \tag{7}$$

$$C(\mathbf{h}) = \sigma^2 \exp(-a^2\|\mathbf{h}\|^2), \tag{8}$$

respectively. More details on sill and range can be found in section 3 of Handcock and Stein (1993) or section 1 of Gneiting et al. (2010). The Matérn family turns out to be a good choice because of its flexibility in modeling various types of spatial correlation structure in many fields and possesses a good interpretability of the parameters. The importance of the family is also highlighted in Stein (1999, p. 14).

Let $C_{uu}(\mathbf{l}_i, \mathbf{l}_j) = \sigma^2 M(\mathbf{h}; \nu, a)$, $a > 0$, $\nu > 0, \mathbf{h} = \|\mathbf{l}_i - \mathbf{l}_j\|$ and $C_{uv}(\mathbf{l}_i, \mathbf{l}_j) = 0$, $u \neq v$. In what follows, we assume the covariance function depends on the r-dimensional parameter $\boldsymbol{\delta} = (\delta_1, \dots, \delta_r)$ each describing various elements of the family. In this article, we assume a Matérn-type family for spatial correlation on the transformed failure times. In that case, $\boldsymbol{\delta} = (\text{range}, \text{sill})$. The transformation leads to a MGRF, where the marginal failure times follow a Cox model with a population-level interpretation for the regression parameters, and facilitates estimation of the spatial as well as regression parameters. To that end, we define

$$\boldsymbol{\Sigma}^{\dagger}_{ij} = \text{Cov}(\mathbf{T}^{\dagger}_i, \mathbf{T}^{\dagger}_j) = \rho_{ij} = \rho(\mathbf{l}_i, \mathbf{l}_j; \boldsymbol{\delta}).$$

For compactness, we will use the notation $\rho_{ij}(\boldsymbol{\delta})$ for $\rho(\mathbf{l}_i, \mathbf{l}_j; \boldsymbol{\delta})$.

3 | MODELS AND ESTIMATION

3.1 | Joint modeling

The notations and details on stochastic processes formulation in this section can be found in Peña et al. (2001) for interested readers. We assume the observables are as given in (2). The τ_is are assumed to be noninformative about the $S_{i,j}$s and that they satisfy the independent censoring condition. Consider a study with n units located at n different spatial locations $\mathbf{l}_i$, where a unit i at location $\mathbf{l}_i$ is monitored over the interval $[0, \tau_i]$ for the occurrence of a recurrent event. Models are described using counting process and martingale formulation. In what follows, s represents calendar time, whereas t represents gap time. The process $N_i{}^{\dagger}(s)$ determines the event occurrences up to time s, whereas the $Y_i{}^{\dagger}(s)$ process determines if the unit is at-risk for a recurrent event. From stochastic integration theory, the compensator process of $N_i{}^{\dagger}(s)$ is $A_i{}^{\dagger}(s)$ given by $A_i^{\dagger}(s|\boldsymbol{\beta}) = \int_0^s Y_i^{\dagger}(v)\lambda_0[\mathcal{E}_i(v)]\exp[\boldsymbol{\beta}'\mathbf{x}_i]dv$, where $\boldsymbol{\beta}$ is a q-dimensional regression parameter and $\lambda_0(\cdot)$ is the baseline hazard function, whose argument is the effective age process $\mathcal{E}_i(s)$. Notice that the argument in $A_i^{\dagger}(s|\boldsymbol{\beta})$ is the process $\mathcal{E}_i(s)$. Transforming it, we obtain a doubly indexed processes $N_i(s,t)$ and $A_i(s,t|\boldsymbol{\beta})$ that are functions of both the calendar and gap times. The process $N_i(s,t)$ is the number of events that occurred by calendar time s for unit i whose effective age is at most gap time t. Hence, for fixed t, $M_i(s,t|\boldsymbol{\beta}) = N_i(s,t) - A_i(s,t|\boldsymbol{\beta})$ is a zero-mean square integrable martingale. Following the notation in the previous section, let $\mathbf{l}_i$ and $\mathbf{l}_j$ be two spatial locations. Fix a calendar time s and recall that $N_i{}^{\dagger}(s)$ and $N_j{}^{\dagger}(s)$ are the number of occurrences by time s in locations $\mathbf{l}_i$ and $\mathbf{l}_j$, respectively. Fix t_1 and t_2 and denote by $N_i(s,t_1)$ (resp. $N_j(s,t_2)$) the number of gap times T_{ik} (resp. T_{jm}) that exceed t_1 (resp. t_2) by time s, and $Y_i(s,t_1)$ and $Y_j(s,t_2)$ their corresponding generalized at-risk processes. Thus, $N_{ij}(s;t_1,t_2)$ is the overall number of gap times over the two spatial locations exceeding t_1 and t_2 simultaneously by time s and given by

$$N_{ij}(s;t_1,t_2) = \sum_{k=1}^{N_i^{\dagger}(s-)} \sum_{m=1}^{N_j^{\dagger}(s-)} I\{T_{ik} \geq t_1, T_{jm} \geq t_2\}.$$

Let $A_i(s,t)$ and $A_j(s,t)$ be the compensator of $N_i(s,t)$ and $N_j(s,t)$, respectively. Define $M_i(s,t) = N_i(s,t) - A_i(s,t)$, and similarly for $M_j(s,t)$. Introduce the filtration $\mathbf{F} = \{\mathfrak{F}_s : s \geq 0\}$, where $\mathfrak{F}_s = \sigma\{[N_i^{\dagger}(s), Y_i^{\dagger}(s)], [N_j^{\dagger}(s), Y_j^{\dagger}(s)], i \neq j, s \geq 0\}$. Then, for fixed t_1 and t_2, the covariance function cov($M_i(s,t_1)$,$M_j(s,t_2)$) is defined by

$$E(M_i(s,t_1)M_j(s,t_2)|T_{ik} > t_1, T_{jm} > t_2) = A_{ij}(s;t_1,t_2) = \langle M_i(t_1), M_j(t_2)\rangle(s).$$

Using stochastic integration theory, we have

$$E\left(M_i(s,t_1)M_j(s,t_2) - \int_0^{t_1}\int_0^{t_2} Y_i(s,u_1)Y_j(s,u_2)A_{ij}(s;du_1,du_2)\right) = 0.$$

The covariance process $E(M_i(s,t_1)M_j(s,t_2)|T_{ik} > t_1, T_{jm} > t_2)$ is bounded. This is true since $E(N_k{}^{\dagger}(s))$, $k \in \{i,j\}$ is finite at each location by virtue of the finiteness of the renewal function $\rho_k(\tau_k)$ at each location. Arguments such as those in the proof of proposition 2 of Peña et al. (2001) can be used to prove that fact. In the spatial setting, the covariance process $A_{ij}(s;t_1,t_2)$ is actually a function of the spatial correlation between locations I_i and I_j, expression of which is given in the next proposition.

To that end, let $\overline{F}^{\dagger}_{ij}(t_1, t_2; \rho_{ij}(\boldsymbol{\delta})) = P(T^{\dagger}_i > t_1, T^{\dagger}_j > t_2; \rho_{ij}(\boldsymbol{\delta}))$ be the bivariate survivor function of $T_ik^{\dagger}$ and $T_{jm}{}^{\dagger}$ (viewed here as individual kth and mth transformed gap time, respectively) and $\overline{F}_{ij}(t_1, t_2; \rho_{ij}(\boldsymbol{\delta})) = P(T_{ik} > t_1, T_{jm} > t_2; \rho_{ij}(\boldsymbol{\delta}))$ that of the original gap times. If $F_i(t_1)$ and $F_j(t_2)$ are the marginal distribution functions of the original gap times, then

$$\overline{F}_{ij}(t_1, t_2; \rho_{ij}(\boldsymbol{\delta})) = \overline{F}^{\dagger}_{ij}(\Phi^{-1}(F_i(t_1)), \Phi^{-1}(F_j(t_2)); \rho_{ij}(\boldsymbol{\delta})).$$

We seek an expression for the covariation process $A_{ij}(s; t_1, t_2; \rho_{ij}(\boldsymbol{\delta}))$. Let the partial derivatives of $\overline{F}_{ij}(t_1, t_2; \rho_{ij}(\boldsymbol{\delta}))$ with respect to t_1 and t_2, t_1, and t_2 be defined by

$$\frac{\partial^2 \overline{F}_{ij}(t_1, t_2; \rho_{ij}(\boldsymbol{\delta}))}{\partial t_1 \partial t_2}, \quad \frac{\partial \overline{F}_{ij}(t_1, t_2; \rho_{ij}(\boldsymbol{\delta}))}{\partial t_1}, \quad \text{and} \quad \frac{\partial \overline{F}_{ij}(t_1, t_2; \rho_{ij}(\boldsymbol{\delta}))}{\partial t_2},$$

respectively. The following proposition gives an expression of $A_{ij}(s; t_1, t_2; \rho_{ij}(\boldsymbol{\delta}))$.

Proposition 1. *For fixed t_1 and t_2, we have*

$$\begin{aligned} A_{ij}(s; t_1, t_2; \rho_{ij}(\boldsymbol{\delta})) = {} & Y_i(s, t_1) Y_j(s, t_2) \lambda_i(t_1) \lambda_j(t_2) \overline{F}_{ij}^{-1}(t_1, t_2; \rho_{ij}(\boldsymbol{\delta})) \\ & \times \left(\frac{\partial^2 \overline{F}_{ij}(t_1, t_2; \rho_{ij}(\boldsymbol{\delta}))}{\partial t_1 \partial t_2} + \frac{\partial \overline{F}_{ij}(t_1, t_2; \rho_{ij}(\boldsymbol{\delta}))}{\partial t_1} + \frac{\partial \overline{F}_{ij}(t_1, t_2; \rho_{ij}(\boldsymbol{\delta}))}{\partial t_2} + \overline{F}_{ij}(t_1, t_2; \rho_{ij}(\boldsymbol{\delta})) \right). \end{aligned}$$

The proof follows along the lines of a similar result in Prentice and Cai (1992) for single event and is omitted due to pages limitation.

3.2 | Estimation

Two models are considered here. The Cox model which is used to identify risk factors and contains two parameters, namely, the infinite dimensional baseline hazard $\lambda_0(t)$ and the regression parameter $\boldsymbol{\beta}$. The other model is the Matérn spatial correlation with an r-dimensional parameter $\boldsymbol{\delta}$ given in Proposition 1. Though, in the case of the Matérn family $r = 2$, we develop the theory for an unknown r. The infinite dimensional parameter $\lambda_0(t)$, although unknown, will be substituted by the generalized Nelson–Aalen estimator $\hat{\Lambda}_0(s, t|\boldsymbol{\beta})$ given in Adekpedjou and Stocker (2015). So, the parameter to be estimated is $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\delta}_0)$, a $(p + r)$-dimensional.

3.2.1 | Process decomposition

Estimation of parameters with spatially correlated random censorship data poses challenges because (i) the high dimension of the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\delta}; \Lambda_0)$ and (ii) the full-likelihood $\mathrm{L}(\boldsymbol{\theta}|\text{Data})$ is computationally burdensome and intractable. The estimation problem is even more complex with recurrent event data. Since it is quite difficult to apply direct maximum likelihood method, we adopt the *pairwise likelihood* approach as an alternative. The main reference is Lindsay (1988). See also Varin, Reid, and Firth (2011) for an overview of composite likelihood applications in various fields. Lindsay, Yi, and Sun (2011) also discusses issues and strategies for the selection of composite likelihood. The idea is to form pairwise likelihoods, a product of likelihoods for data in two spatial locations that can be the basis of an unbiased estimating function, and then be used for parameter estimation. It is a special case of a more general class of pseudolikelihoods called *composite likelihoods*. It is a technique allowing addition of likelihoods in a situation where the components do not represent independent replicates. It has good theoretical properties and behaves well in many applications concerning spatial statistics (Heagerty & Lele, 1998; Hjort & Omre, 1994; Lele & Taper, 2002; Varin, Høst, & Skare, 2005; Varin & Vidoni, 2005). Moreover, it is robust to model misspecification, is computationally advantageous when dealing with data that has a complex structure, and the estimated parameter is the same as in the complete model (Lindsay et al., 2011). Weights between two areas $\mathbf{l}_i$ and $\mathbf{l}_j$, denoted by $w_{i,j}(\boldsymbol{\delta})$ can be included to improve efficiency and rate of convergence.

From now, we adopt a slight change in notation for convenience, from i,j to u,v. Consider a pair of spatial locations u,v and let $M_u(s,t|\boldsymbol{\beta})$ and $M_v(s,t|\boldsymbol{\beta})$ be the martingale processes associated with each location. Let $\mathbf{M}(s,t|\boldsymbol{\beta}) = (M_u(s,t|\boldsymbol{\beta}), M_v(s,t|\boldsymbol{\beta}))$ and $\mathbf{W}(\boldsymbol{\delta}) = (w_{ij}(\boldsymbol{\delta}))$, where $ij \in \{uu, uv, vu, vv\}$ be the (2×2) matrix of weights. Following Adekpedjou and Stocker (2015), we introduce the general process

$$W^{[uv]}(s,t|\boldsymbol{\theta}_0) = \int_0^t \mathbf{H}(s,z|\boldsymbol{\beta}_0)\mathbf{W}(\boldsymbol{\delta}_0)\mathbf{M}'(s,dz|\boldsymbol{\beta}_0), \tag{9}$$

where $\mathbf{H}(s,z|\boldsymbol{\beta}) = (H_u(s,z|\boldsymbol{\beta}), H_v(s,z|\boldsymbol{\beta}))$ is a vector of predictable processes depending on the risk factors and the regression parameter $\boldsymbol{\beta}$. For more details on those expressions, we refer the readers to Peña et al. (2007) and Adekpedjou and Stocker (2015). See also Andersen and Gill (1982) for the random censorship model case. Examination of $W^{[uv]}(s,t|\boldsymbol{\theta}_0)$ shows that it is the sum of four terms, each of which given by:

$$\begin{aligned}
W^{uu}(s,t|\boldsymbol{\theta}_0) &= \int_0^t w_{uu}(\boldsymbol{\delta}_0)H_u(s,z|\boldsymbol{\beta}_0)M_u(s,dz|\boldsymbol{\beta}_0),\\
W^{uv}(s,t|\boldsymbol{\theta}_0) &= \int_0^t w_{uv}(\boldsymbol{\delta}_0)H_u(s,z|\boldsymbol{\beta}_0)M_v(s,dz|\boldsymbol{\beta}_0),\\
W^{vu}(s,t|\boldsymbol{\theta}_0) &= \int_0^t w_{vu}(\boldsymbol{\delta}_0)H_v(s,z|\boldsymbol{\beta}_0)M_u(s,dz|\boldsymbol{\beta}_0),\\
W^{vv}(s,t|\boldsymbol{\theta}_0) &= \int_0^t w_{vv}(\boldsymbol{\delta}_0)H_v(s,z|\boldsymbol{\beta}_0)M_v(s,dz|\boldsymbol{\beta}_0).
\end{aligned} \tag{10}$$

Consider now all the pairwise spatial locations formed with the n locations, then the aggregate process over all pair of locations is given by

$$W(s,t|\boldsymbol{\theta}_0) = \sum_{(u,v),u\le v} W^{[uv]}(s,t|\boldsymbol{\theta}_0). \tag{11}$$

3.2.2 | Estimating functions

For the purpose of estimating $\boldsymbol{\theta}_0$, we seek functions $f(\cdot, \text{Data})$ of $(\boldsymbol{\delta}, \text{Data})$, and $(\boldsymbol{\beta}, \text{Data})$ such that

$$E(f(\boldsymbol{\delta}, \text{Data})) = 0 \quad \text{and} \quad E(f(\boldsymbol{\beta}, \text{Data})) = 0. \tag{12}$$

Equation (12) implies that $f(\boldsymbol{\theta}, \text{Data})$ is an unbiased estimator of 0 if $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$. We begin with estimating function of $\boldsymbol{\delta}_0$. Recall that

$$E(M_u(s,t_1)M_v(s,t_2)|T_{uk} > t_1, T_{vm} > t_2) = A_{uv}(s; t_1, t_2; \rho(\boldsymbol{\delta})) = \langle M_u(t_1), M_v(t_2)\rangle(s),$$

where k and m represent the kth and mth gap time at location u and v, respectively. Therefore, $E\{M_u(s,t)M_v(s,t) - A_{uv}(s;t,t;\rho(\boldsymbol{\delta}))\} = 0$. So any mean-zero function of $M_u(s,t)M_v(s,t) - A_{uv}(s,t;\rho(\boldsymbol{\delta}))$, say $W_{\boldsymbol{\delta}}^{uv}(s,t)$ can serve as an unbiased estimating function for $\boldsymbol{\delta}$ over two spatial locations. To obtain a proper estimating function that has the flavor of score equations, define the (2×2) matrix

$$\mathbf{A}(s,t;\rho(\boldsymbol{\delta})) = (A_{ij}(s,t;\rho(\boldsymbol{\delta})))_{ij\in\{uu,uv,vu,vv\}}.$$

Let $\mathbf{A}_{\delta_i}(s,t;\rho(\boldsymbol{\delta})) = -\frac{\partial}{\partial\delta_i}\mathbf{A}(s,t;\rho(\boldsymbol{\delta}))$, $i = 1, \ldots, r$, be a (2×2) matrix of elementwise derivatives. Let $\Pi_i = \mathbf{A}^{-1}\mathbf{A}_{\delta_i}\mathbf{A}^{-1}$, where we use $\mathbf{A}$ for $\mathbf{A}(s,t;\rho(\boldsymbol{\delta}))$ for compactness. Then, following Cressie (1993, p. 483), we can show that $E(\mathbf{M}(s,t)\Pi_i\mathbf{M}(s,t)) + \mathrm{tr}(\Pi_i\mathbf{A}) = 0$, where $\mathrm{tr}(\cdot)$ is the trace of a matrix. Consequently, the ith component of $W_{\boldsymbol{\delta}}^{uv}(s,t)$, say $W_{\delta_i}^{uv}(s,t)$ is given by

$$W_{\delta_i}^{uv}(s,t) = \mathbf{M}(s,t)\Pi_i\mathbf{M}'(s,t) + \mathrm{tr}(\Pi_i\mathbf{A}) := \mathbf{M}(s,t)\Pi_i\mathbf{M}'(s,t) + \mathrm{tr}(\mathbf{A}^{-1}\mathbf{A}_{\delta_i}). \tag{13}$$

The expression in (13) can be viewed as a score process and its sum over all pair (u,v) can serve as an unbiased estimating function for δ_i. So, the estimating functions over all pairs of spatial locations for $\boldsymbol{\delta}$ is the $r\times 1$ vector $W_{\boldsymbol{\delta}}(s,t) = (W_{\delta_i}(s,t), i = 1, \ldots, r)$, where $W_{\delta_i}(s,t)$ is given by

$$W_{\delta_i}(s,t) = \sum_{(u,v),u\leq v} W_{\delta_i}^{uv}(s,t).$$

Once $\boldsymbol{\delta}$ is estimated, following Adekpedjou and Stocker (2015), an estimating function for $\boldsymbol{\beta}$ over two spatial locations I_u and I_v is

$$W_{\boldsymbol{\beta}}(s,t|\hat{\boldsymbol{\delta}}) = \sum_{(u,v),u\leq v} W_{\boldsymbol{\beta}}^{[uv]}(s,t|\hat{\boldsymbol{\delta}}) = \sum_{(u,v),u\leq v}\int_0^t \mathbf{H}(s,z|\boldsymbol{\beta})\mathbf{W}(\hat{\boldsymbol{\delta}})\mathbf{N}(s,dz), \tag{14}$$

with $\mathbf{N}(s,dz) = (N_u(s,dz), N_v(s,dz))'$, where $\hat{\boldsymbol{\delta}}$ is the estimator of $\boldsymbol{\delta}_0$ over all pairwise regions, where with

$$W_{\boldsymbol{\beta}}^{[uv]}(s,t|\hat{\boldsymbol{\delta}}) = \sum_{ij\in\{uu,uv,vu,vv\}} W_{\boldsymbol{\beta}}^{ij}(s,t|\hat{\boldsymbol{\delta}}).$$

The decomposition in (10) leads to

$$W_{\boldsymbol{\beta}}^{uu}(s,t|\hat{\boldsymbol{\delta}}) = \sum_{v\leq u}\int_0^t w_{uu}(\hat{\boldsymbol{\delta}})H_u(s,z|\boldsymbol{\beta})N_u(s,dz),$$

$$W_{\boldsymbol{\beta}}^{vu}(s,t|\hat{\boldsymbol{\delta}}) = \sum_{v\leq u}\int_0^t w_{vu}(\hat{\boldsymbol{\delta}})H_v(s,z|\boldsymbol{\beta})N_u(s,dz),$$

$$W_{\boldsymbol{\beta}}^{uv}(s,t|\hat{\boldsymbol{\delta}}) = \sum_{v\leq u}\int_0^t w_{uv}(\hat{\boldsymbol{\delta}})H_u(s,z|\boldsymbol{\beta})N_v(s,dz),$$

$$W_{\beta}^{vv}(s,t|\hat{\delta}) = \sum_{v\le u}\int_0^t w_{vv}(\hat{\delta})H_v(s,z|\beta)N_v(s,dz). \tag{15}$$

4 | LARGE SAMPLE PROPERTIES

We now turn to the existence and asymptotic properties of the estimators. In geostatistics, asymptotic properties can be investigated in two different ways: the *Increasing domain asymptotic* or the *Infill asymptotic*. The increasing domain asymptotic is a sampling structure in spatial statistics where new observations are added at the boundary points of an area, whereas the infill asymptotic consists of a sampling structure where new observations are added in between existing locations. The latter is appropriate when the spatial locations are in a bounded domain. In this article, we will use the increasing domain asymptotic since interest is placed in adding new spatial locations for the purpose of predicting events or kriging at those new locations. Therefore, in what follows, the statement "$n \to \infty$" means *Increasing Domain Asymptotic*. Readers are referred to Cressie (1993, section 7.3.1, p. 480) for details.

To proceed with the asymptotic properties, we assume that regularity conditions on p. 53 of Adekpedjou and Stocker (2015) are in effect and refrain from stating them here due to pages limitation. All random entities are assumed to be defined on a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$. As indicated earlier, we assume a perfect intervention with $\mathcal{E}_i(s) = s - S_{N_i^{\dagger}(s)}$ being the backward recurrence time process. In the perfect intervention case, the generalized at-risk process $Y_i(s,t|\beta)$ for a unit at spatial location $\mathbf{l}_i$ is given by

$$Y_i(s,t|\beta) = \left[\sum_{j=1}^{N_i^{\dagger}((s\wedge\tau_i)-)} \mathrm{I}(T_{i,j}\ge t) + \mathrm{I}((s\wedge\tau_i) - S_{i,N_i^{\dagger}((s\wedge\tau_i)-)} \ge t)\right]\mathbf{x}_i^{\otimes k}\exp(\beta'\mathbf{x}_i)$$

and the kth partial derivative of $Y_i(s,t|\beta)$ is

$$\begin{aligned} Y_i^{(k)}(s,t|\beta) &= \frac{\partial Y_i(s,t|\beta)}{\partial \beta^k} \\ &= \left[\sum_{j=1}^{N_i^{\dagger}((s\wedge\tau_i)-)} \mathrm{I}(T_{i,j}\ge t) + \mathrm{I}((s\wedge\tau_i) - S_{i,N_i^{\dagger}((s\wedge\tau_i)-)} \ge t)\right]\mathbf{x}_i^{\otimes k}\exp(\beta'\mathbf{x}_i). \end{aligned}$$

Let $\mathcal{B}$ be a neighborhood of β_0. For $i=1,2,\dots,n$, $s^{\star} = \max_{1\le i\le n}\tau_i$; $\mathbf{H}_i(s^*,t|\beta_0) = [H_{i,j}(s^*,t|\beta_{0,j}) : j=1,2,\dots,q]'$ is a vector processes whose components are predictable and bounded for fixed t. Furthermore, let θ_0 be the true value of θ and θ the free value in $\Theta \subseteq \Re^{r+p}$. As before, for any pair of spatial locations u and v, let $\mathbf{W}^{uv}(\delta) = (w_{ij}^{uv}(\delta))$ with $ij \in \{11,12,21,22\}$, where ij represents the components of the 2×2 matrix of weights. We introduce new processes connected to the $Y_i(s,t|\beta)$, namely, $Y_i^{\mathrm{w}}(s,t|\beta)$ (here the w indicates weighted generalized at-risk processes) and given by

$$Y_i^{\mathrm{w}}(s,t|\beta) := Y_i(s,t;w_{ij}^{uv}) = w_{ij}^{uv}(\delta)Y_i(s,t|\beta).$$

Also, let $Y(s,t|\beta) = \sum_{i=1}^n Y_i(s,t|\beta)$, $Y(s,t;w^{uv}|\beta) = \sum_{i=1}^n Y_i(s,t;w^{uv}|\beta)$, $Y^{(k)}(s,t|\beta) = \sum_{i=1}^n Y_i^{(k)}(s,t|\beta)$, and $Y^{(k)}(s,t;w^{uv}|\beta) = \sum_{i=1}^n Y_i^{(k)}(s,t;w^{uv}|\beta)$. The proofs of the asymptotic properties and the set of regularity conditions are relegated to the Appendix section.

4.1 | Unbiased estimating functions

We begin with the asymptotic properties of processes of the form

$$\mathbf{W}(s,t|\boldsymbol{\theta}_0)=\frac{1}{n}\sum_{(u,v),u\leq v}\int_0^t \mathbf{H}(s,z|\boldsymbol{\beta}_0)\mathbf{W}(\boldsymbol{\delta}_0)\mathbf{M}'(s,dz|\boldsymbol{\beta}_0), \tag{16}$$

where $\mathbf{H}=(H_u,H_v)$, $\mathbf{M}=(M_u,M_v)'$, and $\mathbf{W}(\boldsymbol{\delta}_0)=(w_{ij}^{uv}(\boldsymbol{\delta}_0))$, with $ij\in\{11,12,21,22\}$.

The following lemma extends lemma 1 of Adekpedjou and Stocker (2015) and the proof is omitted.

Lemma 1. *As $n\to\infty$, the process $\{\mathbf{W}(s^*,t|\boldsymbol{\theta}_0)\,:\,t\in[0,t^*]\}$ converges weakly to a zero-mean Gaussian process $\{\mathbf{W}^{(\infty)}(s^*,t|\boldsymbol{\theta}_0)\,:\,t\in[0,t^*]\}$ with covariance function given by*

$$\mathrm{Cov}(\mathbf{W}^{(\infty)}(s^*,t_1|\boldsymbol{\theta}_0),\mathbf{W}^{(\infty)}(s^*,t_2|\boldsymbol{\theta}_0))=\boldsymbol{\Sigma}_V(s^*,t_1\wedge t_2|\boldsymbol{\theta}_0)).$$

Expression of $\boldsymbol{\Sigma}_V(s^*,t_1\wedge t_2|\boldsymbol{\theta}_0))$ is given in the Appendix section under regularity conditions.

We now show the asymptotic unbiasedness of the two estimating functions $W_{\boldsymbol{\beta}}^{[uv]}(s,t|\hat{\boldsymbol{\delta}})$ and $W_{\boldsymbol{\delta}}(s,t)$. We seek estimators of $\boldsymbol{\delta}_0$ and $\boldsymbol{\beta}_0$ that satisfy, asymptotically, $n^{-1}\sum_{u\leq v}W_{\delta_i}^{uv}(s,t)\xrightarrow{p}0$ for $i=1,\dots,r$ and $n^{-1}W_{\boldsymbol{\beta}}^{[uv]}(s,t|\hat{\boldsymbol{\delta}})\xrightarrow{p}0$. We begin with $n^{-1}\sum_{u\leq v}W_{\delta_i}^{uv}(s,t)$. We use the concept of ergodicity which is a generalized version of mixing. Since the spatial covariance function has the second-order property and is intrinsic, we can show unbiasedness using the fact that $\langle M_u(s,t_1),M_v(s,t_2)\rangle$ is ergodic. Discussion on ergodicity and mixing properties with geostatistical data can be found in Cressie (1993, pp. 55–57). See also Kolmogorov and Rozanov (1960) and Bradley (2005).

Theorem 1. *Under Condition III, as $n\to\infty$, for $j=1,\dots,r$, we have*

$$W_{\delta_j}(s,t)=\frac{1}{n}\sum_{(u,v),u\leq v}W_{\delta_j}^{uv}(s,t)\xrightarrow{p}0.$$

Theorem 2. *Under Condition I, as $n\to\infty$,* $\sup_{t\in[0,t^\star]}\frac{1}{n}\sum_{(u,v),u\leq v}|W_{\boldsymbol{\beta}}^{[u,v]}(s,t|\hat{\boldsymbol{\delta}})|\xrightarrow{p}0$.

4.2 | Asymptotic properties

We now turn to the asymptotic properties of estimators. The first theorem below establishes the consistency of the estimators, whereas the second deals with the question of their convergence in distribution when they are properly standardized.

Theorem 3.

(a) There exists a sequence of solutions $(\hat{\boldsymbol{\beta}}_n(s^*,t^\star)\,:\,n=1,2,\dots)$ and $(\hat{\boldsymbol{\delta}}_n(s^*,t^\star)\,:\,n=1,2,\dots)$ to the sequence of estimating equations $W_{\boldsymbol{\delta}}(s^\star,t)=\mathbf{0}$ and $W_{\boldsymbol{\beta}}(s^\star,t|\hat{\boldsymbol{\delta}})=\mathbf{0}$, respectively.

(b) *Under Conditions I and II, as $n \to \infty$, $\hat{\boldsymbol{\beta}}_n(s^*, t^\star) \xrightarrow{p} \boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\delta}}_n(s^*, t^\star) \xrightarrow{p} \boldsymbol{\delta}_0$.*

Theorem 4. *Under regularity Conditions I and IV, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}_{p+r}, \Xi(\boldsymbol{\theta}_0; s^\star, t))$, where $\Xi(\boldsymbol{\theta}_0; s^\star, t)$ is a $(p+r)\times(p+r)$ matrix given by*

$$\Xi(\boldsymbol{\theta}_0; s^\star, t) = [J^{-1}(\boldsymbol{\theta}_0; s^\star, t)]\Sigma_\infty(\boldsymbol{\theta}_0; s^\star, t)[J^{-1}(\boldsymbol{\theta}_0; s^\star, t)]'.$$

Expressions of $J(\boldsymbol{\theta}_0; s^\star, t)$ and $\Sigma_\infty(\boldsymbol{\theta}_0; s^\star, t)$ are given under regularity conditions in the Appendix section.

5 | SIMULATION STUDY AND APPLICATION

5.1 | Simulation study

A simulation study is performed in the R software package to examine the finite sample performance of the proposed estimators with a combination of sample sizes and spatial correlation levels. We consider n bidimensional spatial locations of subjects as a uniform sample over the region $[0, 10]^2$. Subjects are monitored for 2 years until censoring. The censoring distribution is a Uniform distribution on $[0, B]$ with B chosen so that each unit experiences between 2 and 20 events. The recurrent event data were generated via the method described in Amorin and Cai (2015, p. 326). We use the Weibull baseline hazard with shape parameter 2 and scale parameter 1. Other shape parameters are considered but not reported. The covariates x_1 and x_2 were generated from the binomial $\mathrm{B}(n, 0.5)$ and normal $\mathrm{N}(0, 1)$ distribution, respectively. The regression parameter is $\boldsymbol{\beta} = (\beta_1, \beta_2) = (0.9, 0.2)$. For units i and k at locations $\mathbf{I}_i$ and $\mathbf{I}_k$, respectively, the gap times of order j, namely, $T_{i,j}$ and $T_{k,j}$ are correlated. The spatial dependence between two locations $\mathbf{I}_i$ and $\mathbf{I}_j$ is the exponential and Gaussian given in Equations (7) and (8), respectively. We also considered the spherical covariance function between subjects i and j located at $\mathbf{I}_i$ and $\mathbf{I}_j$ given by

$$C(d_{i,j}) = \sigma^2 \left(1 - \frac{3}{2} a d_{i,j} + \frac{1}{2}(a d_{i,j})^3\right),$$

where $d_{i,j}$ is the distance between the two locations, and a is the range. The graphs of the different spatial correlation models are given in Figure 1.

The following algorithm was used to generate the spatially correlated recurrent events.

- Generate $\{S^\dagger_{i,1}; i = 1, \ldots, n\}$, n event times from a centered multivariate normal distribution with marginal variance σ^2 and each of the three spatial correlation models
- Transform $\{S^\dagger_{i,1}; i = 1, \ldots, n\}$ into uniform random variables $U_{i,1}$,
- Transform $\{U_{i,1},\ i = 1, \ldots, n\}$ into dependent Weibull variables $S_{i,1} = \Lambda^{-1}(-\log(U_{i,1}))$, with $S_{i,1} = T_{i,1}$,
- As in the previous steps, generate dependent normal variables $S^\dagger_{i,2}$ and get $U_{i,2}$. So $S_{i,2} = S_{i,1} + d\Lambda^{-1}_{S_{i,1}}(-\log(U_{i,2}))$, where $d\Lambda_t = \Lambda(t+h) - \Lambda(t)$,
- Repeat the previous steps until censoring occurs for each unit.

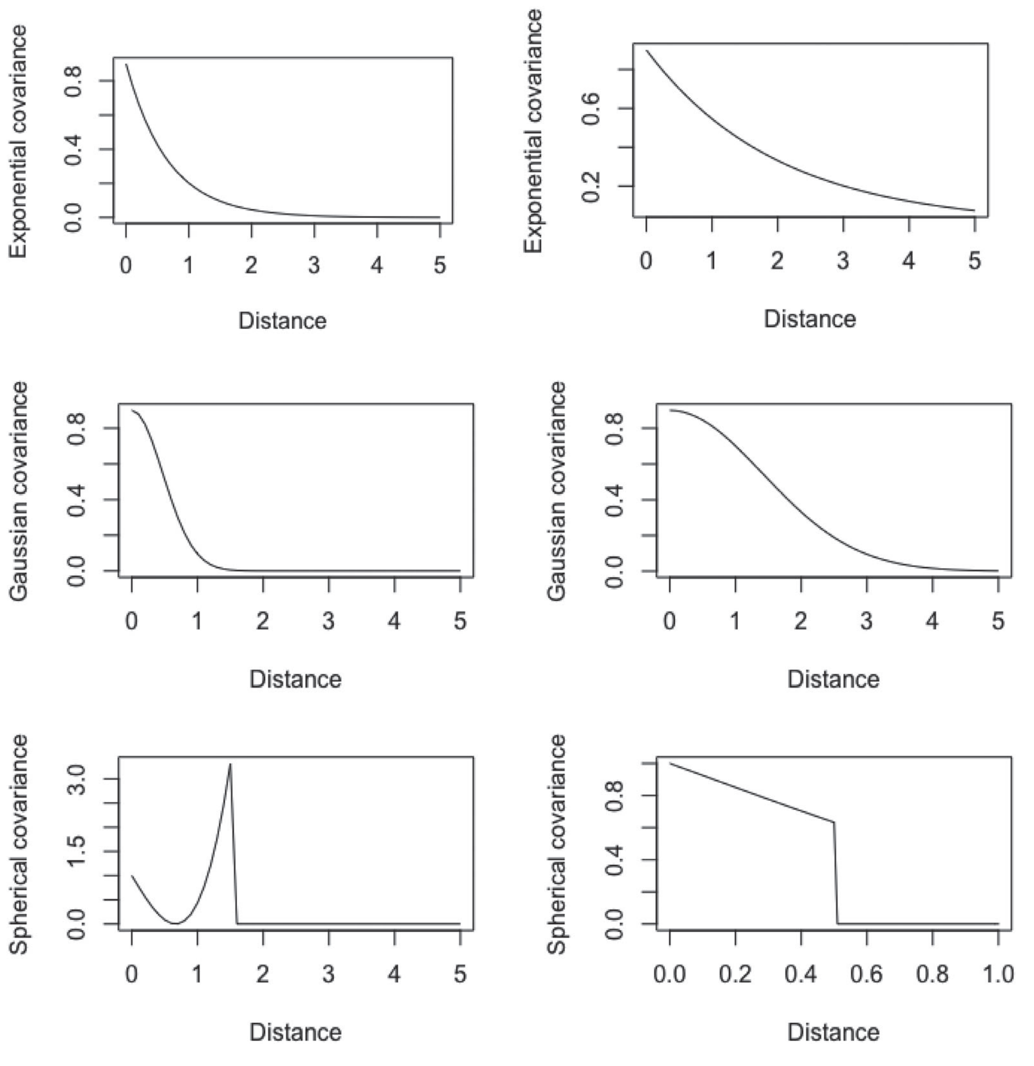


FIGURE 1 Spatial covariance functions: (a) Exponential spatial dependence $\sigma^2 = 1$ and $a = 0.5$ (top panel—left) and $\sigma^2 = 1$, $a = 1.5$ (top panel—right). (b) Gaussian spatial dependence with parameters $\sigma^2 = 1$ and $a = 0.5$ (center panel—left) and $\sigma^2 = 1$, $a = 1.5$ (center panel—right). (c) Spherical spatial dependence with parameters $\sigma^2 = 1$ and $a = 0.5$ (bottom panel—left) and $\sigma^2 = 1$, $a = 1$ (bottom panel—right)

Simulation setting: We use three different correlation decay models: Model A with $\sigma^2 = 0.9$ and $a = 0.5$ corresponding to a small range, but stronger correlation; Model B with $\sigma^2 = 0.9$ and $a = 1.5$ corresponding to a larger range, but weaker correlation; and Model C with $\sigma^2 = 1$ and $a = 1$. Model C is only used for the spherical correlation. We consider sample sizes n in $\{30,50,100\}$ and examine the effect of sample sizes on the properties of $\hat{\boldsymbol{\beta}}$ when the spatial dependence parameters are included or not, denoted by Spatial and Independent (Ind) case, respectively in Tables 1–3. A total of 200 replications were generated for each parameter and sample size combination. The average estimates, mean square standard errors (M^i_{β}, $i = 1, 2$), and standard deviations (Std) of $\hat{\boldsymbol{\beta}}$ were calculated. The results of the study indicate that the estimators of the spatial correlation as well as $\hat{\boldsymbol{\beta}}$ perform well for Models A, B, and C. The tables show that the estimates and their

TABLE 1 Simulation results for Models A and B for spatial and independent cases

Method	Model	n	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{a}$	$\hat{\sigma}^2$	M_β^1	M_β^2	S_a	S_{σ^2}
Spatial (Gaussian)	A	30	0.988 (0.055)	0.191 (0.052)	0.507 (0.051)	0.914 (0.060)	3.2e−3	2.8e−3	3.6e−3	2.8e−3
		50	1.000 (0.048)	0.200 (0.044)	0.496 (0.047)	0.906 (0.047)	2.3e−3	1.9e−3	2.2e−3	2.3e−3
		100	0.989 (0.038)	0.189 (0.040)	0.498 (0.042)	0.914 (0.041)	1.6e−3	1.7e−3	1.8e−3	1.9e−3
	B	30	0.989 (0.056)	0.201 (0.057)	1.481 (0.116)	0.950 (0.049)	3.2e−3	3.1e−3	2.7e−3	2.9e−3
		50	0.985 (0.055)	0.184 (0.049)	1.497 (0.056)	0.896 (0.057)	3.2e−3	2.5e−3	3.1e−3	3.2e−3
		100	1.000 (0.049)	0.192 (0.058)	1.509 (0.052)	0.915 (0.053)	2.4e−3	3.4e−3	2.7e−3	3e−3
Ind	A	30	1.011 (0.109)	0.198 (0.043)	.	.	1.2e−2	1.9e−3	.	.
		50	1.005 (0.081)	0.197 (0.038)	.	.	6e−3	1.5e−3	.	.
		100	1.002 (0.066)	0.201 (0.027)	.	.	4.3e−3	7.2e−4	.	.
	B	30	0.983 (0.116)	0.208 (0.052)	.	.	1.3e−2	2.7e−3	.	.
		50	1.025 (0.079)	0.194 (0.040)	.	.	6.7e−2	1.6e−3	.	.
		100	1.020 (0.028)	0.197 (0.047)	.	.	2.6e−2	7.9e−4	.	.

Note: Standard deviations are in parentheses. MSE is M_{β^i}, $i = 1, 2$. Exponential covariance parameters are $\hat{a}$ and $\hat{\sigma}^2$.

associated standard errors have relatively little bias across the sample sizes and that the regression coefficient estimates perform well particularly for large sample sizes. When correlation is ignored, that is in the independent case, the estimates are unstable resulting in very high standard deviations, especially for $\hat{\beta}_1$. They are particularly very unstable in Model A (stronger correlation) than in Model B (weaker correlation, which is closer to independence). The missing cells in the Tables 1–3 mean no estimates are provided since those cells correspond to the independent case, consequently, there are no spatial correlation parameters to report.

5.2 | Application

5.2.1 | Checking regularity conditions

For our application, we use the exponential spatial covariance function given by $C(\mathbf{h}) = \sigma^2 \exp(-a\|\mathbf{h}\|)$. The T_{ij}s are assumed to follow a Weibull distribution, so close form expression of

TABLE 2 Simulation results for Models A and B

Method	Model	n	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{a}$	$\hat{\sigma}^2$	M^1_β	M^2_β	S_a	M_{σ^2}
Spatial (exponential)	A	30	0.992 (0.049)	0.194 (0.054)	0.499 (0.053)	0.896 (0.056)	2.5e−3	3e−3	2.7e−3	3.1e−3
		50	0.994 (0.049)	0.199 (0.048)	0.494 (0.051)	0.911 (0.053)	2.4e−3	2.3e−3	2.6e−3	2.9e−3
		100	0.989 (0.036)	0.187 (0.037)	0.483 (0.037)	0.947 (0.040)	1.4e−3	1.5e−3	1.6e−3	1.7e−3
	B	30	0.971 (0.049)	0.178 (0.054)	1.494 (0.056)	0.918 (0.056)	3.1e−3	3.2e−3	3e−3	3e−3
		50	0.993 (0.054)	0.197 (0.056)	1.490 (0.059)	0.909 (0.064)	2.9e−3	3.1e−3	3.6e−3	4.1e−3
		100	0.990 (0.049)	0.197 (0.047)	1.493 (0.054)	0.910 (0.058)	2.5e−3	2.2e−3	2.9e−3	3.3e−3
Ind	A	30	0.995 (0.121)	0.191 (0.053)	.	.	1.4e−2	2.5e−3	.	.
		50	0.999 (0.083)	0.200 (0.042)	.	.	6.8e−3	1.7e−3	.	.
		100	1.001 (0.060)	0.200 (0.026)	.	.	3.6e−3	7.1e−4	.	.
	B	30	0.952 (0.110)	0.207 (0.058)	.	.	1.3e−2	3.2e−3	.	.
		50	1.011 (0.099)	0.189 (0.035)	.	.	9.7e−3	1.3e−3	.	.
		100	1.001 (0.055)	0.201 (0.023)	.	.	3e−2	5e−4	.	.

Note: Models with and without spatial dependence (independent case). Standard deviations are in parentheses. MSE is M^i_β, $i = 1,2$. Gaussian covariance parameters are $\hat{a}$ and $\hat{\sigma}^2$.

$\lambda(t)$ exists. Consider a pair $\{u,v\}$ of regions and let $\|\mathbf{h}\| = \|u - v\|$. To check the regularity conditions, it suffices to check uniform convergence of the weighted generalized processes. To this end, consider the process in (9). We need to check that the weighted generalized at risk processes are uniformly convergent. To do so, we need an exact expression of the weight matrix $\mathbf{W}$. Using variable transformation, it can easily be shown that the joint survivor function $\overline{F}_{ij}(t_1, t_2)$ is a function of $\overline{F}_i(t_1)$, $\overline{F}_j(t_2)$, and $C(\mathbf{h})$. Let $D = \text{diag}\{A_{11}, \dots, A_{nn}\}$, and $\mathbf{A}$ be the $(n \times n)$ matrix of compensators. Following Heagerty and Lele (1998) and Cressie (1993), we define the weight matrix $\mathbf{W}$ by $\mathbf{W} = [D^{-0.5}\mathbf{A}D^{-0.5}]^{-1}$. We have $A_{ij} = \sigma^2$ for $ij \in \{uu,vv\}$, and $A_{ij} = \sigma^2 \exp(-a\|u - v\|)$ for $ij \in \{uv,vu\}$. In the preceding, $\mathbf{A}$ and $\mathbf{W}$ are $(n \times n)$ matrices. Simple algebra calculations give, for $n = 3$ spatial locations

$$w_{ij}(\boldsymbol{\delta}) = \begin{cases} 1, & \text{if } i = j, \\ A_{ii}^{-0.5} A_{ij} A_{jj}^{-0.5}, & ij \in \{12, 21, 13, 31, 23, 32\}. \end{cases}$$

TABLE 3 Simulation results for Models A and C

Method	Model	n	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{a}$	$\hat{\sigma}^2$	M_{β_1}	M_{β_2}	S_a	S_{σ^2}
Spatial (spherical)	A	30	0.987 (0.050)	0.191 (0.053)	0.479 (0.046)	0.901 (0.051)	2.6e−3	2.8e−3	2.5e−3	2.6e−3
		50	0.998 (0.051)	0.201 (0.049)	0.495 (0.053)	0.902 (0.054)	2.6e−3	2.4e−3	2.8e−3	2.8e−3
		100	0.992 (0.044)	0.194 (0.047)	0.505 (0.042)	0.893 (0.046)	2e−3	2.2e−3	1.8e−3	2.1e−3
	C	30	1.004 (0.049)	0.216 (0.049)	0.996 (0.055)	0.895 (0.042)	2.4e−3	1.8e−3	1.8e−3	2.9e−3
		50	0.988 (0.057)	0.198 (0.050)	0.996 (0.052)	0.909 (0.054)	3.3e−3	2.4e−3	2.7e−3	2.9e−3
		100	0.990 (0.044)	0.184 (0.047)	1.005 (0.048)	0.902 (0.049)	2.1e−3	2.4e−3	2.3e−3	2.4e−3
Ind	A	30	1.021 (0.109)	0.204 (0.051)	.	.	1.2e−2	2.6e−3	.	.
		50	1.025 (0.077)	0.208 (0.036)	.	.	6.5e−3	1.4e−3	.	.
		100	1.001 (0.054)	0.197 (0.026)	.	.	2.9e−3	6.6e−4	.	.
	C	30	1.004 (0.103)	0.185 (0.048)	.	.	1e−2	2.5e−3	.	.
		50	1.000 (0.083)	0.201 (0.039)	.	.	6.7e−3	1.4e−3	.	.
		100	0.996 (0.061)	0.201 (0.03)	.	.	3.7e−2	9e−4	.	.

Note: Models with and without spatial dependence (independent case). The table gives the mean estimates for each combination, with their standard deviation in parentheses, and the MSE M^i_{β}, $i = 1,2$. Spherical covariance parameters are $\hat{a}$ and $\hat{\sigma}^2$.

The components of $\mathbf{W}$ are deterministic and bounded functions of (σ^2, a) and therefore regularity Conditions I (i) are satisfied using the uniform law of large numbers of Pollard (1990). Parts of regularity Condition I (ii) is a consequence again of the uniform convergence of average of partial derivative of $Y^{(k)}(s, t; w^{uv}|\boldsymbol{\beta})$. As for Condition I (iv), observe that, for $\mathbf{t} = (t_1, \dots, t_n)$, $\mathbf{M}(s,\mathbf{t}) = (M_1(s,t_1), \dots, M_n(s,t_n))$, properly scaled has a multivariate normal limiting distribution with zero mean. In view of regularity Conditions (I), Condition (IV) is trivially satisfied.

5.2.2 | Applications

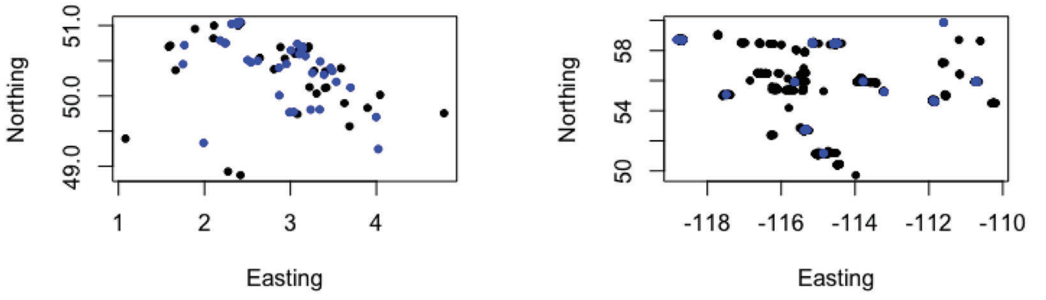
We now provide two different applications for illustration. One in the biomedical area pertaining to the recurrent esophagus cancer and the other in environment pertaining to recurrent wildfire.

Application to esophagus cancer

The data set in this application pertains to the recurrence of esophagus cancer in patients in NPC. The data were provided by the CHRU (University Hospital Center) of Lille and concerns mainly patients from Lille. We want to model relapses in esophagus cancer patients with the data set. NPC has one of the highest rate of esophageal and stomach cancers in France. According to the "*Registre General des Cancers de Lille et de sa Region*" aka General Register of Cancer of Lille and Greater Lille of 2008–2010, there is a 20% and 16% rate of these with men and women respectively in the area. Those high rates are associated with a high percentage of recurrences with an estimated 30% of chance of recurrence within 5 years in both cases. Moreover, there is between 20% and 80% likelihood of recurrence of stomach cancer within the first 5 years and 2–3 years for the esophageal. Past informal studies suggested that socioeconomic background, physical and social factors may have influenced the initial occurrence and further recurrences of cancer in the area. The objective of the study is to identify, while taking into account proven risk factors, the impact of new socioeconomic factors, both at the regional and national level, in order to make recommendations to public authorities on the policies to be implemented to reduce the impact of the socioeconomic causes of cancer. A research project has been launched in Lille (North of France) to study the recurrence rate of esophageal cancer. The main objective is to identify the frequency of recurrence of esophageal cancer at 36 months for patients who have undergone curative surgery. A secondary objective of the project is to identify risk factors that will allow patients to be classified by risk groups (low, medium, high) for recurrence.

For many years, researchers have been interested in the impact of socioeconomic inequalities between individuals on cancer incidence as well as the probability of survival. These inequalities can be measured by different factors (Galobardes, Shaw, Lawlor, Lynch, & Smith, 2006a, 2006b). Other studies include Menvielle, Luce, Geoffroy-Perez, Chastang, and Leclerc (2005) on the impact of an individual's level of education and type of occupation on mortality from several type of cancers (pharynx, larynx, lung, esophagus, stomach, and rectum). Ouédraogo et al. (2014) investigated, among other, the effect of distance to the treatment center on the probability of following a breast cancer screening. Fiva, Hægeland, Rønning, and Syse (2014) studied the effect of access to treatment and educational inequalities on cancer survival in Norway. Other studies have tried to address these socioeconomic disparities between individuals by incorporating variables measured at a larger geographical scale. For instance, Pornet et al. (2012) have developed an economic precariousness index for France, measured at the level of IRIS (Ilots Regrouped for Statistical Information, corresponding to geographical grids with about 2,000 inhabitants). This index is based on a measure of the lack of access to goods perceived as fundamental by the population and was used by Ouédraogo et al. (2014) to study the impact of living in a precarious neighborhoods on the probability of having a mammogram. Some authors have studied the problem of recurrence of esophageal cancers and have shown a correlation with tumor, clinical, epidemiological, and behavioral risk factors such as alcoholism and smoking status (Mariette et al., 2003; Markar et al., 2015; Oppedijk et al., 2014).

To the best of our knowledge, the project launched in Lille is the first one with a goal of analyzing the impact of socioeconomic risk factors on these types of cancer. The methodology used for this project consists of a retrospective cohort study. Thus, a database including patients diagnosed with esophageal cancer and having undergone curative surgery in the Digestive and General Surgery Department of the CHRU of Lille was created and contains patients who had a surgery between 1990 and 2015. The prognostic factors are sex (females vs. males), postsurgery status based on histological nodes and tumor stages (ptNM Stage 0 vs. pN Stages 1-3), age of the patient at diagnosis, smoking status, alcohol status, geographical positions of the postal code,



(a) Year 2010 recurrent cancer cases in blue (b) Recurrent wildfire in various locations in blue

FIGURE 2 Maps of areas used for applications

TABLE 4 Percentages of esophagus relapses according to some criterions

Criterion	%Pat	%Relap
RTCT	63	56.52
ptNM(0)	13.6	1.00
ptNM(1)	31.5	21.7
ptNM(2)	19.1	57.1
ptNM(3)	35.6	76.9
Tobacco	64	54.3
Alcohol	33	58.3

Abbreviations: Relap, relapse; Pat, patients.

deprivation indices at the zip code area (EDI), and information on the relapses. The relevant outcome on esophageal cancer status is only available at the zip code level due to personal privacy constraints. The maps of the area considered in both applications are given in Figure 2. We only consider patients that had surgery for whom the tumor has been totally removed and survived at least a few days after the surgery. We applied our procedures for analyzing relapses and survival among patients. Since our methodology is only adapted to one individual per spatial location, we limit our study to subjects having surgery in 2010, where only one patient by postal code was recorded. A total of 73 patients were monitored after a first diagnosis and 46.57% of them had a relapse. The maximum number of relapses was 1, implying a maximum number of events to be 2 per patient. The main medical postsurgery information we consider is the pTNM variable with four stages (ptNM Stage 0, 1, 2, 3) of tumor, ptNM 0 is the best stage after surgery, whereas the worst case is ptNM 3. There were 84.9% males, 35.61% patients were diagnosed at ptNM Stage 3 after the surgery, 64% were smokers, 33% drink alcohol, and 63% have previously had chemotherapy and/or radiotherapy before the surgery, cf. Table 4.

We applied the spatial semiparametric normal transformation model to our data set of 73 patients having surgery in 2010. We used the Cox model with covariates: sex, status of tobacco consumption, deprivation index (EDI), chemotherapy status, and the stage of the tumor. The exponential spatial covariance model was considered. We estimated the regression coefficients

TABLE 5 Esophagus cancer data illustrated with an exponential spatial correlation

Criterion	Coef	Std	95% CI—Low	95% CI—Up
EDI	1.10	0.02	1.05	1.14
Smoking alcohol status	4.05	0.45	1.59	7.92
RTCT status	4.71	0.39	2.27	10.27
ptNM Stage (1)	5.64	0.62	1.63	19.10
ptNM Stage (2)	8.00	0.75	2.20	32.78
ptNM Stage (3)	28.12	0.68	7.92	95.58
	Spatial coef	Std	95% CI—Low	95% CI—Up
Sill	0.23	0.16	0.017	0.501
Range	1.08	1.21	0.03	4.13

Note: The table gives parameter estimates and their standard deviations for the covariates: sex, alcohol, smoking status, chemotherapy, radiotherapy, and ptNM stage.

and the spatial covariance parameters (sill and range) and run 500 bootstrap and calculated the associated standard deviations and their 95% confidence intervals (CIs). The estimated range and sill of the autocorrelation are $\hat{a} = 1.086$ (in order of 122 km) and $\hat{\sigma}^2 = 0.23$. The value of the range translates into patients living in areas within a radius of 122 km were more likely to have esophagus cancer than people living far away.

The risk factors hazard ratios (HR) and corresponding 95% confidence intervals for recurrences are reported in Table 5. We found that EDI has a significant impact on the likelihood of recurrence. Patients with high EDI have a larger risk of relapses and those diagnosed with a ptNM stage 0 have a lower risk of relapses than patients diagnosed at later stages of the disease, that is, ptNM Stages 1–3. For relapse patients, previous chemotherapy and radiotherapy (*RTCT* variable equals to 1 if a patient has a chemotherapy and radiotherapy before surgery and 0 otherwise) have significant impact. Patients that had a previous chemotherapy and radiotherapy have a higher risk of relapse. Patients who smoke and drink alcohol (*smokealcohol* variable equals to 1 if a patient smokes and drinks at least two cups per day 0 otherwise) have a higher risk of relapse.

Application to forest wildfires

This application is in environment and pertains to the recurrence of forest fires known as wildfire. The data obtained from the wildfire database in the Province of Alberta in Canada over the 2004–2018 period. It concerns fires having a point of origin inside the Forest Protection Area (FPA) either Provincial or Federal lands. Alberta is divided in many regions and the database gives several information of the wildfire for various regions such as region geographic coordinates, the year in which the wildfire occurred, an estimated time and date the wildfire started, the fire origin type of land, initial assessment of fire, weather conditions at the time of initial assessment, size of the wildfire at the time it is extinguished, among others.

A descriptive analysis of the data set indicates that several fires at the same location recurred during the period 2004–2018 and the methodology in the article can be used for illustration. Initial analysis led us to focus on fires originating from the Indian Reservation where we noted a dispersion of the recurrences from 0 to 13. It is composed of 2,363 fires locations with 76 recurrences. Locations are censored at the end of the study period (2004-01-01 to 2018-10-31). The covariates include weather conditions over the wildfire (clear: taken as baseline, cloudy, and raining); category of wildfires in hectares (ha) based on final area burned (A: [0,0.1) taken as baseline; B:

TABLE 6 Proportion of recurrences with respect to specific covariates

Criterion	%Fires	%Recur
Weather (Rainshowers)	2.2	2.3
Weather (Cloudy)	36.5	2.5
Weather (Clear)	61.3	3.10
FireSize (A)	56.2	3.3
FireSize (B)	37.9	3.3
FireSize (C)	4.80	1
FireSize (D)	0.80	0
FireSize (E)	0.30	0

TABLE 7 Wildfire data set illustrated with the exponential spatial correlation

Criterion	HR	Std (coef)	95% CI—Low	95% CI—Up
Weather (Rainshowers)	0.85	0.06	0.79	0.89
(ref. Clear)				
Weather (Cloudy)	0.90	0.04	0.86	0.92
FireSize (B)	0.87	0.05	0.82	0.91
(ref. A)				
FireSize (C)	0.59	0.20	0.46	0.68
FireSize (D)	6.8e−4	2.80	2.98e−5	5.5e−3
FireSize (E)	6.7e−4	2.80	2.95e−5	5.4e−3
	Spatial coef	Std	95% CI—Low	95% CI—Up
Sill	0.91	0.04	0.88	0.96
Range	2.30	0.33	1.81	2.53

Note: The table gives parameter estimates of weather condition and size class covariates and their standard.

[0.1,4), C: [4,40), D: [40,200), and E: $[200,\infty)$). The percentage of the fires that had at least one recurrence is 3.21% resulting in a total of 217 recurrences. The average number of recurrences is 1.06 and those vary from 0 to 13. Among those with at least one recurrence, 97.4% had at most 5 with a mean number of recurrences of 2.6. We truncated the data set after the fifth event due to the small number of locations with large number of recurrences and analyzed 191 recurrences, see Table 6.

The results in Table 7 present the hazard ratios (HR) and corresponding 95% confidence intervals for the risk factors for fire recurrences with 500 bootstrap and associated standard deviations of coefficients. The results point out that locations weather conditions and size class are recurrence factors. In fact, under raining weather or cloudy conditions, there is a reduction of 15% and 10% on the risk of recurrence, respectively, compared with fires under clear weather conditions (under raining; HR = 0.85, with 95% confidence interval: [0.79,0.89]). There is a reduction of 13% and 41% on the risk of recurrence respectively for large fire size, say B and C classes, compared with fire of A class which coincides to fire extinguished. For instance, under class B, HR = 0.87

and the 95% confidence interval is [0.82,0.91]. The risk of recurrence for larger classes is almost 0 and the confidence interval is almost a null set. Furthermore, under the exponential spatial correlation model, the estimated range and sill controlling the strength and the scale of autocorrelation are $\hat{a} = 2.3$ (in order of 258 km) and $\hat{\sigma}^2 = 0.91$.

6 | DISCUSSION

In this article, we have proposed models for identifying risk factors at onset and future recurrence of events. Due to the intractable nature of the likelihood and its high dimensionality, parameters in the proposed methodologies are estimated using composite likelihood. The approach facilitates the derivation of asymptotic properties of estimators, which otherwise would have been difficult. That also led to computationally efficient estimates through proper choices of weights. The proposed models provide a general framework and tools for researchers and practitioners when dealing with recurrent failures in the presence of spatial correlation. Although they provide significant contribution to the body of knowledge in the modeling and analysis of spatially correlated recurrent events, more need to be done. It would be interesting to develop criteria for an optimal choice of weights. Another model that could be used to account for covariates is the additive model. As one of the referees pointed out, having methodologies with more units per geographical area would be a very good contribution to the field, since, in almost all real-life applications that would be the case. Of interest in that case is how to develop models for correlation within a spatial site while accounting for spatial correlation between sites. Random effects models within a geographical area and spatial correlation between two areas would be a good place to start.

ACKNOWLEDGEMENTS

The authors graciously thank the editor, associate editor, and referees for their useful comments and perceptive suggestions, which led to considerable improvements of the article. The authors thank ONCO Lille, Professeur Christophe Mariette of CHU of Lille for giving us access to the NPC data.

ORCID

Akim Adekpedjou https://orcid.org/0000-0001-9584-4297
Sophie Dabo-Niang https://orcid.org/0000-0002-4000-6752

REFERENCES

Adekpedjou, A., & Stocker, R. (2015). A general class of semiparametric models for recurrent event data. *Journal of Statistical Planning and Inference*, *156*, 48–63.

Amorin, D. L., & Cai, J. (2015). Modelling recurrent events: A tutorial for analysis in epidemiology. *International Journal of Epidemiology*, *44*(1), 324–333.

Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, *10*(4), 1100–1120.

Banerjee, S., & Carlin, B. (2003). Semi-parametric spatiotemporal analysis. *Environmetrics*, *14*, 523–535.

Banerjee, S., & Dey, D. K. (2005). Semiparametric proportional odds models for spatially correlated survival data. *Lifetime Data Analysis*, *11*(2), 175–191.

Banerjee, S., Wall, M., & Carlin, B. (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*, *4*(1), 123–142.

Barthel, N., Geerdens, C., Czado, C., & Janssen, P. (2019). Dependence modeling for recurrent event times subject to right-censoring with D-vine copulas. *Biometrics*, *75*(2), 439–451.

Bradley, R. C. (2005). Basic properties of strong mixing conditions: A survey and some open questions. *Probability Surveys*, *2*, 107–144.

Bronnenberg, B. J. (2005). Spatial models in marketing research and practice. *Applied Stochastic Models in Business and Industry*, *21*(4-5), 335–343.

Cook, R., & Lawless, J. (2002). Analysis of repeated events. *Statistical Methods in Medical Research*, *11*, 141–166.

Cook, R. J., & Lawless, J. (2007). *The statistical analysis of recurrent events (Statistics for Biology and Health) Statistics for biology and health* (). New York, NY: Springer.

Cressie, N. A. C. (1993). *Statistics for spatial data Wiley series in probability and mathematical statistics: applied probability and statistics* (). New York, NY: John Wiley & Sons, Inc., A Wiley-Interscience Publication.

Crowder, M. (1986). On consistency and inconsistency of estimating equations. *Econometric Theory*, *2*, 305–330.

Diva, U., Banerjee, S., & Dey, D. K. (2007). Modelling spatially correlated survival data for individuals with multiple cancers. *Statistical Modelling*, *7*(2), 191–213.

Diva, U., Dey, D. K., & Banerjee, S. (2008). Parametric models for spatially correlated survival data for individuals with multiple cancers. *Statistics in Medicine*, *27*(12), 2127–2144.

Engen, S. (2007). Stochastic growth and extinction in a spatial geometric Brownian population model with migration and correlated noise. *Mathematical Biosciences*, *209*(1), 240–255.

Fiva, J. H., Hægeland, T., Rønning, M., & Syse, A. (2014). Access to treatment and educational inequalities in cancer survival. *Journal of Health Economics*, *36*, 98–111.

Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., & Smith, G. D. (2006a). Indicators of socioeconomic position (Part 1). *Journal of Epidemiology & Community Health*, *60*(1), 7–12.

Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., & Smith, G. D. (2006b). Indicators of socioeconomic position (Part 2). *Journal of Epidemiology & Community Health*, *60*(2), 95–101.

Gneiting, T., Kleiber, W., & Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, *105*(491), 1167–1177.

Goethals, K., Janssen, P., & Duchateau, L. (2008). Frailty models and copulas: Similarities and differences. *Journal of Applied Statistics*, *35*(9-10), 1071–1079.

Golzy, M., & Carter, R. L. (2019). Generalized frailty models for analysis of recurrent events. *Journal of Statistical Planning and Inference*, *200*, 213–222.

Guttorp, P., & Gneiting, T. (2006). Studies in the history of probability and statistics XLIX. On the Matérn correlation family. *Biometrika*, *93*(4), 989–995.

Guyon, X. (1995). *Random fields on a network Probability and its applications* (). New York, NY: Springer-Verlag.

Handcock, M. S., & Stein, M. L. (1993). A Bayesian analysis of Kriging. *Technometrics*, *35*(4), 403–410.

Heagerty, P. J., & Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, *93*(443), 1099–1111.

Henderson, R., Shimakura, S., & Gorst, D. (2002). Modeling spatial variation in leukemia survival data. *Journal of the American Statistical Association*, *97*(460), 965–972.

Hjort, N. L., & Omre, H. (1994). Topics in spatial statistics [with discussion, comments and rejoinder]. *Scandinavian Journal of Statistics*, *21*(4), 289–357.

Hougaard, P. (2000). *Analysis of multivariate survival data Statistics for biology and health* (). New York, NY: Springer-Verlag.

Huang, C.-Y., & Wang, M.-C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association*, *99*(468), 1153–1165.

Hunt, F. Y. (1978). Genetic and spatial variation in some selection-migration models (Ph.D. thesis). ProQuest LLC, Ann Arbor, MI: New York University.

Kolmogorov, A. N., & Rozanov, J. A. (1960). On a strong mixing condition for stationary Gaussian processes. *Theory of Probability & Its Applications*, *5*, 222–227.

Lawless, J. F., & Yilmaz, Y. E. (2011). Semiparametric estimation in copula models for bivariate sequential survival times. *Biometrical Journal*, *53*(5), 779–796.

Lee, J., & Cook, R. J. (2019). Dependence modeling for multi-type recurrent events via copulas. *Statistics in Medicine*, *38*(21), 4066–4082.

Lele, S., & Taper, M. L. (2002). A composite likelihood approach to (co)variance components estimation C. R. Rao 80th birthday felicitation volume, Part I. *Journal of Statistical Planning and Inference*, *103*(1-2), 117–135.

Li, Y., & Lin, X. (2006). Semiparametric normal transformation models for spatially correlated survival data. *Journal of the American Statistical Association*, *101*(474), 591–603.

Li, Y., & Ryan, L. (2002). Modeling spatial survival data using semiparametric frailty models. *Biometrics*, *58*(2), 287–297.

Lindsay, B. G. (1988). *Composite likelihood methods*. In *Statistical inference from stochastic processes (Ithaca, NY, 1987)* (Vol. *80*, pp. 221–239). Providence, RI: American Mathematical Society.

Lindsay, B. G., Yi, G. Y., & Sun, J. (2011). Issues and strategies in the selection of composite likelihoods. *Statistica Sinica*, *21*(1), 71–105.

Liu, B., Lu, W., & Zhang, J. (2014). Accelerated intensity frailty model for recurrent events data. *Biometrics*, *70*(3), 579–587.

Mariette, C., Balon, J.-M., Piessen, G., Fabre, S., Van Seuningen, I., & Triboulet, J.-P. (2003). Pattern of recurrence following complete resection of esophageal carcinoma and factors predictive of recurrent disease. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, *97*(7), 1616–1623.

Markar, S., Gronnier, C., Duhamel, A., Mabrut, J.-Y., Bail, J.-P., Carrere, N., et al. (2015). The impact of severe anastomotic leak on long-term survival and cancer recurrence after surgical resection for esophageal malignancy. *Annals of Surgery*, *262*(6), 972–980.

Matérn, B. (1986). *Spatial variation Lecture notes in statistics* (2nd ed.). Berlin, Germany: Springer-Verlag.

Matheron, G. (1962). *Traité de gostatistique appliquée Statistics for biology and health* (). France: Editions Technip.

Menvielle, G., Luce, D., Geoffroy-Perez, B., Chastang, J.-F., & Leclerc, A. (2005). Social inequalities and cancer mortality in France, 1975–1990. *Cancer Causes & Control*, *16*(5), 501–513.

Murphy, S. (1995). Asymptotic theory for the frailty model. *The Annals of Statistics*, *23*, 182–198.

Oppedijk, V., van der Gaast, A., van Lanschot, J. J., van Hagen, P., van Os, R., van Rij, C. M., et al. (2014). Patterns of recurrence after surgery alone versus preoperative chemoradiotherapy and surgery in the cross trials. *Journal of Clinical Oncology*, *32*(5), 385–391.

Ouédraogo, S., Dabakuyo-Yonli, T. S., Roussot, A., Pornet, C., Sarlin, N., Lunaud, P., et al. (2014). European transnational ecological deprivation index and participation in population-based breast cancer screening programmes in France. *Preventive Medicine*, *63*, 103–108.

Paik, J., & Ying, Z. (2012). A composite likelihood approach for spatially correlated survival data. *Computational Statistics & Data Analysis*, *56*(1), 209–216.

Pan, C., Cai, B., Wang, L., & Lin, X. (2014). Bayesian semiparametric model for spatially correlated interval-censored survival data. *Computational Statistics & Data Analysis*, *74*, 198–208.

Parner, E. (1998). Asymptotic theory for the correlated gamma frailty model. *The Annals of Statistics*, *26*, 183–214.

Peña, E. A., Slate, E. H., & Gonzalez, J. R. (2007). Semiparametric inference for a general class of models for recurrent events. *Journal of Statistical Planning and Inference*, *137*, 1727–1747.

Peña, E. A., Strawderman, R. L., & Hollander, M. (2001). Nonparametric estimation with recurrent event data. *Journal of the American Statistical Association*, *96*(456), 1299–1315.

Pollard, D. (1990). *Empirical processes: Theory and applications. NSF-CBMS Regional Conference Series in Probability and Statistics* (vol 2). Hayward, CA and Alexandria, VA: American Statistical Association, Institute of Mathematical Statistics.

Pornet, C., Delpierre, C., Dejardin, O., Grosclaude, P., Launay, L., Guittet, L., … Launoy, G. (2012). Construction of an adaptable Eropean transnational ecological deprivation index: the French version. *Journal of Epidemiol Community Health*, *66*(11), 982–989.

Prentice, R. L., & Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, *79*(3), 495–512.

Scheike, T. H., Eriksson, F., & Tribler, S. (2019). The mean, variance and correlation for bivariate recurrent event data with a terminal event. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *68*(4), 1029–1049.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory for Kriging Springer series in statistics* (). New York, NY: Springer-Verlag.

Therneau, T., & Hamilton, S. (1997). rhdnase as an example of recurrent event analysis. *Statistics in Medicine*, *16*, 271–288.

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: Extending the Cox model Statistics for biology and health* (). New York, NY: Springer-Verlag.

van der Vaart, A. W. (1998). *Asymptotic statistics Cambridge series in statistical and probabilistic mathematics* (Vol. *3*). Cambridge, MA: Cambridge University Press.
Varin, C., Høst, G., & Skare, O. i. (2005). Pairwise likelihood inference in spatial generalized linear mixed models. *Computational Statistics & Data Analysis*, *49*(4), 1173–1191.
Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, *21*(1), 5–42.
Varin, C., & Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, *92*(3), 519–528.
Vaupel, J. (1990). *Kindred lifetimes: Frailty models in population genetics*. In J. Adams, D. A. Lam, A. I. Herman, & P. E. Smouse (Eds.), *Convergent issues in genetics and demography* (pp. 156–170). London, UK: Oxford University Press.
Yilmaz, Y. E., & Lawless, J. F. (2011). Likelihood ratio procedures and tests of fit in parametric and semiparametric copula models with censored data. *Lifetime Data Analysis*, *17*(3), 386–408.

How to cite this article: Adekpedjou A, Dabo-Niang S. Semiparametric estimation with spatially correlated recurrent events. *Scand J Statist*. 2021;48:1097–1126. https://doi.org/10.1111/sjos.12480

APPENDIX A

A.1 Regularity conditions

I. Asymptotic stability

(i) There exist deterministic functions $y^{(k)}(s,t;w_{ij}|\boldsymbol{\beta})$, for $ij \in \{11,12,21,22\}$, $k=0,1,2$ such that

$$\sup_{\substack{t\in[0,t^*]\\ \theta\in\Theta}} \left\| \frac{1}{n} \sum_{(u,v),u\leq v} Y^{(k)}(s,t;w^{uv}|\boldsymbol{\beta}) - y^{(k)}(s,t;w_{ij}) \right\| \xrightarrow{p} 0,$$

as $n\to\infty$, and for any ij, and $k=0,1,2$.

(ii) Let $\mathbf{HW}(s,t|\boldsymbol{\theta})$ be a (1×2) vector, and suppose there exists a deterministic function $\mathbf{hw}(s^*,t|\boldsymbol{\theta}_0)$ such that

$$\max_{(i,j)} \sup_{t\in[0,t^*]} ||[\mathbf{HW}](s^*,t|\boldsymbol{\theta}_0) - \mathbf{hw}(s^\star,t|\boldsymbol{\theta}_0)|| \xrightarrow{p} 0,$$

as $n\to\infty$. Define

$$\mathbf{V}_n(s,t|\boldsymbol{\theta}_0) = n^{-1} \sum_{(u,v),u\leq v} \int_0^t [\mathbf{HW}]^{\otimes 2}(s,z|\boldsymbol{\theta}_0) Y(s,z;w^{uv}|\boldsymbol{\beta})\lambda_0(z)dz,$$

and let

$$\boldsymbol{\Sigma}_V(s,t|\boldsymbol{\theta}_0) = \int_0^t \mathbf{hw}(s,z|\boldsymbol{\theta}_0)^{\otimes 2} y^{(k)}(s,w;w_{ij})|\boldsymbol{\theta}_0)\lambda_0(z)dz.$$

Furthermore, assume that $\mathbf{V}_n(s^*,t|\boldsymbol{\theta}_0)$ converges uniformly in probability to $\boldsymbol{\Sigma}_V(s^*,t|\boldsymbol{\theta}_0)$ under the increasing domain asymptotic.

II. Consistency of spatial parameter

Let $W_{\delta_i}(s,t) = \sum_{(u,v),u\le v} W^{uv}_{\delta_i}(s,t)$, and $W_{\delta}(s,t) = (W_{\delta_i}(s,t), i = 1, \dots, r)$ and suppose there exists a positive definite matrix $\Sigma(s,t|\delta_0)$ such that

$$\sup_{t\in[0,t^\star]} \|n^{-1}\nabla_{\delta} W_{\delta}(s,t) - \Sigma(s,t|\delta_0)\| \xrightarrow{p} 0. \tag{A1}$$

III. Mixing condition for showing spatial parameter estimating function converges to zero
Let $\mathfrak{F}^{(0,i)}_s = \sigma\{\mathcal{O}_i(z) : 0 \le z \le s\}$ and $\mathfrak{F}^{(j,\infty)}_s = \sigma\{\mathcal{O}_j(z) : 0 \le z \le s\}$ and $\phi(\cdot)$ some mixing function.
   (i) $\phi(\cdot)$ is a mixing function such that $\phi(h) \to 0$ as $\|h\| \to \infty$.
   (ii) For any $A \in \mathfrak{F}^{(0,i)}_s$, and $B \in \mathfrak{F}^{(j,\infty)}_s$ collected on the corresponding units by time s in location $\mathbf{l}_i$ and $\mathbf{l}_j$, assume

$$\sup_{A,B} |P(A\cap B) - P(A)P(B))| \le \phi(\|i-j\|), \tag{A2}$$

where P is a probability measure on $\mathfrak{F}$.

IV. Guyon (1995) condition on convergence to normal distribution of estimators in spatial statistics. Stated for convenience for recurrent events.
Let $\mathbf{W}_{\delta}(s,t) = [W_{\delta_j}(s,t) : j = 1, \dots, r]$ and $\mathbf{W}_{\theta}(s,t) = n^{-1}(W'_{\beta}(s,t,), W'_{\delta}(s,t))'$. We define two matrices, namely, $\Sigma_n(\theta_0; s,t)$ and $J_n(\theta; s,t)$ by

$$\Sigma_n(\theta_0; s,t) = \frac{1}{n^2}\sum_{(p,q),(u,v)} [E\mathbf{W}^{uv}_{\theta_0}(s,t)\cdot(\mathbf{W}^{pq}_{\theta_0}(s,t))'],$$

and the matrix $J_n(\theta; s,t)$ is given by

$$J_n(\theta; s,t) = \frac{1}{n}\sum_{(u,v),u\le v} [E(\nabla_{\beta'} W^{uv}_{\beta}(s,t)), E(\nabla_{\delta'} W^{uv}_{\delta}(s,t))].$$

   (i) There exists a neighborhood $\mathcal{N}(\theta_0)$ of $\Re^{p+r}$ over which $\mathbf{W}_{\theta}(s,t)$ is continuously differentiable, and a positive definite matrix $J(\theta_0; s^\star, t)$ such that, as $n \to \infty$,

$$\sup_{t\in[0,t^*]} \|J_n(\theta_0; s^\star, t) - J(\theta_0; s^\star, t)\| \xrightarrow{p} 0,$$

   and that

$$\sqrt{n} J_n^{-1}(\theta_0; s^\star, t)\mathbf{W}_{\theta_0}(s,t) \xrightarrow{p} \mathrm{N}(\mathbf{0}, \mathrm{I}_{p+r}).$$

   (ii) There exists a matrix $\Sigma_\infty(\theta_0; s,t)$ such that $\Sigma_n(\theta_0; s,t) \xrightarrow{p} \Sigma_\infty(\theta_0; s,t)$.

A.2 Proofs of theoretical results

A.2.1 Proof of Theorem 1

Let $I_n(\iota_n) = \{i,j/\|i-j\| \ge \iota_n\}$. The set $I_n(\iota_n)$ gives the range beyond which a cutoff point beyond which the spatial correlation does not have any impact. The cutoff point ι_n depends on the number of locations. We use Chebyshev inequality along with condition II. For $\epsilon > 0$

$$
\begin{aligned}
P(W_{\delta_j}(s,t) > \epsilon) &= P(n^{-1}\mathbf{M}'(s,t)\Pi_j\mathbf{M}(s,t) \geq [\epsilon + \mathrm{tr}(\Pi_j\mathbf{A})]) \\
&\leq \frac{\langle \mathbf{M}'(s,t)\Pi_j\mathbf{M}(s,t)\rangle}{n^2[\epsilon + \mathrm{tr}(\Pi_j\mathbf{A})]^2} \\
&\leq \frac{\langle \mathbf{M}'(s,t)\mathbf{M}(s,t)\rangle}{n^2\epsilon^2} \\
&\leq [n^2\epsilon^2]^{-1} \sum_{i,j\in I_n(\iota_n)} (\langle M_i(s,t)M_j(s,t)\rangle) \\
&\quad + [n^2\epsilon^2]^{-1} \sum_{i,j\in I_n^C(\iota_n)} (\langle M_i(s,t)M_j(s,t)\rangle) \\
&\leq [n^2\epsilon^2]^{-1} \left[\sum_{i,j\in I_n^C(\iota_n)} \phi(\|i-j\|) + \sum_{i,j\in I_n(\iota_n)} \phi(\|i-j\|) \right] \\
&\leq [n^2\epsilon^2]^{-1} \left[n\iota_n^2 + n \sum_{\|i-j\|\geq \iota_n} \phi(\|i-j\|) \right].
\end{aligned}
$$

Under Condition III, and an appropriate choice of $\iota_n \to \infty$, the right-hand side in the previous display goes to zero as $n \to \infty$. This may be done by the polynomial mixing condition such as $\phi(t) \leq Ct^{-\theta}$. The condition on ι_n can be linked to the mixing coefficient $\theta > 0$. Exponential mixing condition may also be used.

A.2.2 Proof of Theorem 2

We begin by observing that

$$
\begin{aligned}
|W_{\beta}^{[u,v]}(s,t|\hat{\boldsymbol{\delta}})| \leq{}& |W_{\beta}^{uu}(s,t|\hat{\boldsymbol{\delta}})| + |W_{\beta}^{uv}(s,t|\hat{\boldsymbol{\delta}})| \\
&+ |W_{\beta}^{vu}(s,t|\hat{\boldsymbol{\delta}})| + |W_{\beta}^{vv}(s,t|\hat{\boldsymbol{\delta}})|. \qquad \text{(A3)}
\end{aligned}
$$

It suffices to show that each term in the RHS of (A3) converges to zero in probability. Without loss of generality, we only show that $W_{\beta}^{uv}(s,t|\hat{\boldsymbol{\delta}})$ does. Asymptotic negligibility of the remaining terms are obtained in the same manner. Define

$$
\begin{aligned}
I_1 &= \frac{1}{n}\sum_{u\leq v}\int_0^t H_u(s,z)w_{11}^{uv}M_u(s,dz), \\
I_2 &= \frac{1}{n}\sum_{u\leq v}\int_0^t H_u(s,z)Y_u(s,z;w_{11}^{uv}|\boldsymbol{\theta})\frac{\sum_{i=1}^n M_i(s,dz)}{Y(s,z|\boldsymbol{\beta})}, \\
I_3 &= \frac{1}{n}\sum_{i=1}^n\int_0^t \left[H_i(s,z)\left(\sum_{u\leq v} Y_u(s,z;w_{11}^{uv}||\boldsymbol{\theta}) - y(s,z;w_{11}|\boldsymbol{\beta})\right)\right] M_i(s,dz), \\
I_4 &= \frac{1}{n}\sum_{i=1}^n\int_0^t y(s,z;w_{11}|\boldsymbol{\beta})M_i(s,dz), \\
I_5 &= \frac{1}{n}\sum_{i=1}^n\int_0^t \frac{y(s,z;w_{11}|\boldsymbol{\beta})}{y(s,z|\boldsymbol{\beta})}M_i(s,dz).
\end{aligned}
$$

Observe that $W_{\beta}^{uu}(s,t|\hat{\delta}) = I_1 - I_2$. A decomposition of I_1 yields

$$I_1 = \sum_{i=1}^{n} \int_0^t \left[H_i(s,z) \left(\sum_{u \le v} Y_u(s,z;w_{11}^{uv}|\theta) - y(s,z;w_{11}|\beta) \right) \right] M_i(s,dz) + \frac{1}{n} \sum_{i=1}^{n} \int_0^t y(s,z;w_{11}|\beta) M_i(s,dz),$$

where the RHS is $I_3 + I_4$. Using Lemma 1 and Condition I, it can be shown that

$$\sup_{t\in[0,t^*]} \left| \frac{1}{n} \sum_{i=1}^{n} \int_0^t \left[H_i(s,z) \left(\sum_{u \le v} Y_u(s,z;w_{11}^{uv}||\theta) - y(s,z;w_{11}|\beta) \right) \right] M_i(s,dz) \right| = o_p(1).$$

Again, by Lemma 1, the process $\{n^{-1} \sum_{i=1}^{n} \int_0^t y(s,z;w_{11}|\beta) M_i(s,dz) : t \in [0,t^\star]\}$ converges weakly to a zero-mean Gaussian process, and from that fact, it can be shown directly that

$$\sup_{t\in[0,t^*]} \frac{1}{n} \sum_{i=1}^{n} \int_0^t y(s,z;w_{11}^{uv}|\beta) M_i(s,dz) = o_p(1).$$

We now consider I_2. Note that

$$I_2 = \frac{1}{n} \sum_{u \le v} \int_0^t \left[\frac{\sum_{u\le v} H_u(s,z) Y_u(s,z;w_{11}^{uv}|\beta)}{Y(s,z|\beta)} - \frac{y(s,z;w_{11}^{uv}|\beta)}{y(s,z|\beta_0)} \right] \sum_{i=1}^{n} M_i(s,dz) + \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{y(s,z;w_{11}^{uv}|\beta)}{y(s,z|\beta_0)} M_i(s,dz). \tag{A4}$$

Both terms in the RHS of (A4) are of $o_p(1)$ order. Combining the asymptotic negligibility of both terms in I_2, coupled with the fact that I_3, I_4, and I_5 are all $o_p(1)$ by Lemma 1, we may conclude that $\sup_{t\in[0,t^*]} |W_{\beta}^{uu}(s,t)| = o_p(1)$. Similarly, under condition I, it can be shown that $W_{\beta}^{xy}(s,t)$, $xy \in \{uv, vu, vv\}$ are all asymptotically negligible. Hence $\sup_{t\in[0,t^\star]} n^{-1} W_{\beta}^{[uv]}(s,t) \xrightarrow{p} 0$ as $n \to \infty$, completing the proof.

A.2.3 Proof of Theorem 3

(a) We use theorem 5.9 of van der Vaart (1998). We need to show that $W_{\beta}(s^\star, t|\hat{\delta})$ converges to a deterministic function that is negative definite. Consider the term $W_{\beta}^{uu}(s,t|\hat{\delta})$ in (15) and observe that

$$\frac{1}{n} W_{\beta}^{uu}(s,t|\hat{\delta}) = \frac{1}{n} \sum_{v \le u} \int_0^t y(s,z;w_{11}^{uv}|\beta) H_u(s,z) M_u(s,dz|\beta_0) + \frac{1}{n} \sum_{v \le u} \int_0^t w_{11}^{uv}(\hat{\delta}) \left[\frac{Y^{(1)}(s,z|\beta_0)}{Y(s,z|\beta_0)} - \frac{Y^{(1)}(s,z|\beta)}{Y(s,z|\beta)} \right] \lambda_0(z) dz. \tag{A5}$$

The first term in the RHS of (A5) is $o_p(1)$ by Lemma 1. Also, clearly, using Condition I and the decomposition used in the proof of Theorem 2, the second term in the RHS of (A5) converges in probability to

$$\int_0^t y(s,z;w_{11}^{uv}|\hat{\boldsymbol{\delta}}_0)\left[\frac{y^{(1)}(s,z|\boldsymbol{\beta}_0)}{y(s,z|\boldsymbol{\beta}_0)}-\frac{y^{(1)}(s,z|\boldsymbol{\beta})}{y(s,z|\boldsymbol{\beta})}\right]\lambda_0(z)dz, \tag{A6}$$

which is 0 at $\boldsymbol{\beta}=\boldsymbol{\beta}_0$. Moreover, the derivative of (A6) with respect to $\boldsymbol{\beta}$ is

$$-\int_0^t y(s,z;w_{11}^{uv}|\hat{\boldsymbol{\delta}}_0)\left[\frac{y^{(2)}(s,z|\boldsymbol{\beta}_0)}{y^{(1)}(s,z|\boldsymbol{\beta}_0)}-\left(\frac{y^{(1)}(s,z|\boldsymbol{\beta})}{y(s,z|\boldsymbol{\beta})}\right)^{\otimes 2}\right]\lambda_0(z)dz.$$

The preceding display is negative definite at $\boldsymbol{\beta}=\boldsymbol{\beta}_0$. One can similarly show, under Condition I that each term in $n^{-1}W_{\boldsymbol{\beta}}(s,t)$ converges in probability to a term with negative definite derivative. Since $\hat{\boldsymbol{\beta}}_n(s^*,t^\star)$ is a global maximum, by theorem 5.9 of van der Vaart (1998), the sequence of solutions $\hat{\boldsymbol{\beta}}_n(s^*,t^*)$ is consistent in a neighborhood $\mathcal{B}_0$ of $\boldsymbol{\beta}$. This proves Part (a) of the theorem.

(b) To obtain the consistency of $\hat{\boldsymbol{\delta}}_n(s^*,t^*)$, we need to verify the conditions of lemma 3.1 and theorem 3.2 of Crowder (1986). Let $\partial S(\boldsymbol{\delta},r)$ be the boundary of the sphere centered at $\boldsymbol{\delta}$ with radius r. The first condition of Crowder on continuity of $W_{\boldsymbol{\delta}}(s,t)$ is trivially satisfied since $W_{\delta_j}^{uv}(s,t)$ is continuous. Moreover, convergence of $W_{\delta_j}(s,t)$ to zero has been established in Theorem 1. To establish the remaining conditions, observe that a Taylor expansion of $E_{\boldsymbol{\delta}_0}[W_{\boldsymbol{\delta}}(s,t)]$ at $\boldsymbol{\delta}_0$ yields

$$(\boldsymbol{\delta}-\boldsymbol{\delta}_0)'E_{\boldsymbol{\delta}_0}[W_{\boldsymbol{\delta}_0}(s,t)]=(\boldsymbol{\delta}-\boldsymbol{\delta}_0)'E_{\boldsymbol{\delta}_0}\nabla_{\boldsymbol{\delta}'}W_{\boldsymbol{\delta}}(s,t)(\boldsymbol{\delta}-\boldsymbol{\delta}_0)+o(\|\boldsymbol{\delta}-\boldsymbol{\delta}_0\|^2). \tag{A7}$$

Now, Condition II together with the Taylor expansion in (A7) is enough to satisfy the last condition of Crowder (1986) on $\partial S(\boldsymbol{\delta},r)$, thereby establishing consistency of the sequence $\hat{\boldsymbol{\delta}}_n(s^*,t^*)$ in a neighborhood $\mathcal{N}_{\boldsymbol{\delta}_0}$ of $\boldsymbol{\delta}_0$.

A.2.4 Proof of Theorem 4

Condition IV (i) is similar to condition 3 in remarks (3), p. 112 of Guyon (1995) and is guaranteed from Condition III. It only remains to show Condition IV(ii). To show condition IV(ii), consider $W_{\boldsymbol{\beta}}^{uu}(s,t)$. We have

$$W_{\boldsymbol{\beta}}^{uu}(s,t)=\frac{1}{n}\sum_{v\le u}\int_0^t w_{11}^{uu}(\boldsymbol{\delta})H_u(s,z)[M_u(s,dz)+Y_u(s,z)d\Lambda_0(z)]. \tag{A8}$$

The derivative with respect to $\boldsymbol{\beta}$ of (A8) is

$$\frac{1}{n}\sum_{v\le u}\int_0^t w_{11}^{uu}(\boldsymbol{\delta})H_u(s,z)^{\otimes^2}M_u(s,dz)+\frac{1}{n}\sum_{v\le u}\int_0^t w_{11}^{uu}(\boldsymbol{\delta})H_u(s,z)^{\otimes^2}Y_u(s,z)]d\Lambda_0(z). \tag{A9}$$

Using Lemma 1 and Condition I(i), (A9) converges in probability to

$$o_p(1)-\int_0^t h(s,z|\boldsymbol{\beta}_0)^{\otimes^2}y(s,z;w_{11}^{uv})d\lambda_0(z)+\int_0^t h(s,z|\boldsymbol{\beta}_0)y^{(1)}(s,z;w_{11}^{uv}|\boldsymbol{\delta}_0)d\lambda_0(z).$$

Using similar arguments, it can be shown that $W^{vu}_{\boldsymbol{\beta}}(s,t)$ converges in probability to

$$-\int_0^t h(s,z|\boldsymbol{\beta}_0)^{\otimes^2} y(s,z;w^{uv}_{21})d\lambda_0(z) + \int_0^t h(s,z|\boldsymbol{\beta}_0)y^{(1)}(s,z;w^{uv}_{21})d\lambda_0(z).$$

Adding up the in-probability limit of the derivatives with respect to $\boldsymbol{\beta}$ of all elements in (4.7), we get the $\mathbf{J}_{11}$ block of $\mathbf{J}$, which is a $(p \times p)$ matrix given by

$$\begin{aligned} J_{11}(s,t|\boldsymbol{\delta}_0) = \int_0^t & h(s,z|\boldsymbol{\beta}_0)[y^{(1)}(s,z;w^{uv}_{11}) + y^{(1)}(s,z;w^{uv}_{21}) + y^{(1)}(s,z;w^{uv}_{12}) \\ & + y^{(1)}(s,z;w^{uv}_{22})]d\lambda_0(z) - \int_0^t h(s,z|\boldsymbol{\beta}_0)^{\otimes^2}[y^{(1)}(s,z;w^{uv}_{11}) \\ & + y^{(1)}(s,z;w^{uv}_{21}) + y^{(1)}(s,z;w^{uv}_{12}) + y^{(1)}(s,z;w^{uv}_{22})]d\lambda_0(z). \end{aligned}$$

Obviously, blockwise, $\mathbf{J}_{12} = \mathbf{0}$, an $(r \times r)$ matrix of 0. The $\mathbf{J}_{22}$ block of $\mathbf{J}$ is obtained by taking the derivative of $\mathbf{W}^{uv}_{\boldsymbol{\delta}}(s,t)$ with respect to $\boldsymbol{\delta}$. Let $\mathbf{A}_{\delta_i\delta_j} = \frac{\partial^2 \mathbf{A}}{\partial\delta_i\partial\delta_j}$, and $\mathbf{A}^{-1}_{\delta_i\delta_j} = \frac{\partial^2 \mathbf{A}^{-1}}{\partial\delta_i\partial\delta_j}$. Let $\mathbf{A}^{-1}_{\delta_i}$ be the matrix of elementwise derivatives of $\mathbf{A}^{-1}$. Using trace and matrices properties coupled with some algebra, the (i,j)th element of $J_{22,n}$ per couple of regions (i,j) is given by

$$\mathbf{J}^{ij,uv}_{22,n} = \mathrm{tr}\{\mathbf{A}(s,t;\boldsymbol{\delta})\mathbf{A}_{\delta_i\delta_j} + \mathbf{A}^{-1}_{\delta_i}(s,t;\boldsymbol{\delta})\mathbf{A}_{\delta_j}(s,t;\boldsymbol{\delta})\}.$$

As a result, there exists a block matrix J_{22} such that $\frac{1}{n}\sum_{(u,v),u\leq v}\mathbf{J}^{ij,uv}_{22} \xrightarrow{p} \mathbf{J}^{ij}_{22}$. This completes the proof of the theorem.